

確率論 (情報通信工学科)

小川朋宏 (電気通信大学 大学院情報システム学研究科)

2010 年度 後学期 (beta version)

1 確率空間 (初学者向け)

1.1 標本空間と事象

1.1.1 用語の定義

- 標本空間 (sample space) : 対象となる偶然現象の結果全体の集合 (Ω とおく)
- 事象 (event) : $A \subset \Omega$ (結果について興味のある事柄)

例 1 (さいころを振る)

- 標本空間 : $\Omega = \{1, 2, 3, 4, 5, 6\}$
- 事象の例 : $A = \{1, 3, 5\}$ (出た目が奇数), $B = \{1, 2, 3\}$ (出た目が 3 以下)
- 事象の演算

$$A \cap B = \{1, 3\} \quad (\text{出た目が奇数かつ 3 以下})$$

$$A \cup B = \{1, 2, 3, 5\} \quad (\text{出た目が奇数または 3 以下})$$

$$A^c = \{2, 4, 6\} \quad (\text{出た目が奇数ではない, すなわち偶数})$$

- 確率の計算

$$\text{奇数の目が出る確率} = P(A) = P(\{1\}) + P(\{3\}) + P(\{5\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

1.1.2 確率とは

- 標本空間 Ω 全体を 1 としたときの, 事象 $A \subset \Omega$ の「割合」 $0 \leq P(A) \leq 1$
- 各事象について「割合」 $P(A)$ を与える「規則」

$$P : A \subset \Omega \mapsto P(A) \in [0, 1]$$

- 確率論 : $P(A)$ の満たすべき性質を公理化することで, 計算方法や性質を議論する .

1.1.3 復習：写像，べき集合

- 写像 $f: X \rightarrow Y$: 集合 X の各要素について，集合 Y の要素を一つ割り当てる「規則」

$$f: x \in X \mapsto y = f(x) \in Y$$

- べき集合：部分集合全体から成る集合

$$2^\Omega := \{A \mid A \subset \Omega\}$$

- 例： $\Omega = \{1, 2\}$ のとき， $2^\Omega = \{\phi, \{0\}, \{1\}, \{0, 1\}\}$ ¹
- Ω の部分集合 A を指定する (べき集合の要素 A を指定する)
 - $\Leftrightarrow$ 各要素 $\omega \in \Omega$ が A に含まれるか，含まれないかを指定する
 - $\Leftrightarrow$ 次の写像 $1_A: \Omega \rightarrow \{0, 1\}$ を指定する²

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

$\Leftrightarrow$ 以下のような表を与える

$$\begin{pmatrix} \Omega \text{ の要素} & \omega_1 & \omega_2 & \omega_3 & \omega_4 & \omega_5 & \cdots \\ 1_A \text{ の値} & 0 & 1 & 1 & 0 & 1 & \cdots \end{pmatrix}$$

- べき集合の要素数： $|2^\Omega| = 2^{|\Omega|}$
 ($\because \Omega$ の各要素 $\omega \in \Omega$ について， 1_A の値は 0 or 1 の，二つの選択肢がある)

1.1.4 復習：集合演算 ($\equiv$ 事象の演算)

Ω を集合とし， $A, B \subset \Omega$ を部分集合とする．

- 積集合 ($\equiv$ 積事象)： $A \cap B := \{\omega \in \Omega \mid \omega \in A \text{ かつ } \omega \in B\}$
- 和集合 ($\equiv$ 和事象)： $A \cup B := \{\omega \in \Omega \mid \omega \in A \text{ または } \omega \in B\}$
- 補集合 ($\equiv$ 余事象)： $A^c := \{\omega \in \Omega \mid \omega \notin A\}$
- $A_1, A_2, \dots, A_n \subset \Omega$ について，以下の集合も帰納的に定義される．

$$A_1 \cap A_2 \cap \cdots \cap A_n, \quad A_1 \cup A_2 \cup \cdots \cup A_n$$

- 包含関係： $A \subset B \stackrel{\text{def}}{\iff} \forall \omega \in \Omega (\omega \in A \Rightarrow \omega \in B)$
- 全事象 Ω ： 包含関係の定義より $\Omega \subset \Omega$
- 空事象 ϕ ： 包含関係の定義より $\phi \subset \forall A \subset \Omega$
 $\because A \subset \Omega$ を任意とする． $\forall \omega \in \Omega$ について， $\omega \notin \phi$ だから $(\omega \in \phi \Rightarrow \omega \in A)$ は真となる．
- 根元事象： $\{\omega\} \subset \Omega (\omega \in \Omega)$

¹ ϕ は空集合

² 1_A は，部分集合 $A \subset \Omega$ の特性関数 (indicator function) と呼ばれる

1.2 確率測度

定義 1 (互いに素) 事象 $A_1, A_2, \dots, A_n$ が互いに素 $\stackrel{\text{def}}{\iff} A_i \cap A_j = \phi \ (i \neq j)$

定義 2 (確率測度の公理, 初学者バージョン) Ω を有限集合とし, $\mathcal{F} = 2^\Omega$ とする. 写像

$$P : A \in \mathcal{F} \mapsto P(A) \in \mathbb{R}$$

が以下の条件を満たすとき, P を Ω 上の確率測度 (probability measure) という.

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

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

(3) $A, B \in \mathcal{F}$ が互いに素ならば

$$P(A \cup B) = P(A) + P(B)$$

このとき, $(\Omega, \mathcal{F}, P)$ を確率空間 (probability space) とよぶ.

補題 1 (確率測度の性質)

(1) $\forall A \subset \Omega, P(A) + P(A^c) = 1$

(2) $\forall A \subset \Omega, 0 \leq P(A) \leq 1$

(3) $P(\phi) = 0$

(4) $\forall A, \forall B \subset \Omega, A \subset B \Rightarrow P(A) \leq P(B)$

(5) (有限加法性) $A_1, A_2, \dots, A_n \subset \Omega$ が互いに素ならば

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

(6) $\forall A, \forall B \subset \Omega, P(A \cup B) = P(A) + P(B) - P(A \cap B)$

レポート 1 上記の性質 (1) ~ (6) を証明せよ (以下はヒント).

(1) $\Omega = A \cup A^c$ で, A と A^c は互いに素.

(2) 公理 (1) より $P(A) \geq 0$. 性質 (1) と公理 (1) より $P(A) \leq 1$ を示す.

(3) $\phi = \Omega^c$, 公理 (2), 性質 (1) を用いる.

(4) $B \setminus A := B \cap A^c$ と定義する. $B = A \cup (B \setminus A)$ で, A と $B \setminus A$ は互いに素.

(5) 公理 (3) より n についての帰納法で示す.

(6) $C = A \setminus B, D = B \setminus A, E = A \cap B$ とおくと, C, D, E は互いに素. 性質 (5) を用いる.

補題 2 Ω が有限集合のとき, 確率測度の公理 (1) ~ (3) は以下の条件と同値.

(1) $\forall \omega \in \Omega, P(\{\omega\}) \geq 0$

(2) $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$

(3) $\forall A \subset \Omega, P(A) = \sum_{\omega \in A} P(\{\omega\})$

証明: 明らか. 確認せよ.

□

注意 1 有限 (可算) 集合 Ω を標本空間として考える場合 ,

- 確率測度 $\equiv$ “ 根元事象に「0 以上で全部たすと 1」となる重みを与えたもの ”
- 事象 $A \in \Omega$ の確率 $=$ 含まれる根元事象の確率の和

例 2 (さいころ) 公平なさいころを 2 回投げる .

(1) 確率空間 $(\Omega, \mathcal{F}, P)$ はどのようなになるか ?

- 標本空間 : $\Omega = \{(i, j) \mid i = 1, 2, 3, 4, 5, 6, j = 1, 2, 3, 4, 5, 6\}$
- 事象全体 : $\mathcal{F} = 2^\Omega$
- 確率測度は $P(\{\omega\}) = \frac{1}{36} \ (\forall \omega \in \Omega)$ によって定まる .

(2) 1 回目と 2 回目の和が 5 になる事象 A を示し , その確率を求めよ .

- $A = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$
- $P(A) = \frac{4}{36} = \frac{1}{9}$

例 3 (公正なコインを n 回投げる) 表=1, 裏=0 と表記する .

(1) 確率空間 $(\Omega, \mathcal{F}, P)$ はどのようなになるか ?

- X と Y の直積集合 (復習) :

$$X \times Y := \{(x, y) \mid x \in X, y \in Y\} \quad (X \text{ の要素と } Y \text{ の要素を並べたベクトルの集合})$$

- $X^n := X \times X \times \cdots \times X$ (X の要素を n 個並べたベクトルの集合)
- 標本空間 : $\Omega = \{0, 1\}^n = \{\omega = (x_1, x_2, \dots, x_n) \mid x_i \in \{0, 1\} (i = 1, 2, \dots, n)\}$
- 事象全体 : $\mathcal{F} = 2^\Omega$
- 確率測度は $P(\{\omega\}) = \frac{1}{2^n} \ (\forall \omega \in \Omega)$ によって定まる .

(2) k 回目が表となる事象 A_k ($k = 1, 2, \dots, n$) の確率を求めよ .

- $A_k = \{\omega = (x_1, x_2, \dots, x_n) \in \Omega \mid x_k = 1\}$
- 事象 A_k の確率は

$$P(A_k) = \sum_{\omega \in A_k} P(\{\omega\}) = \sum_{\omega \in A_k} \frac{1}{2^n} = \frac{|A_k|}{2^n} = \frac{2^{n-1}}{2^n} = \frac{1}{2}$$

(3) k 回目と l 回目 ($k \neq l$) が表となる事象の確率を求めよ .

- k 回目と l 回目 ($k \neq l$) が表となる事象は

$$A_k \cap A_l = \{\omega = (x_1, x_2, \dots, x_n) \in \Omega \mid x_k = 1 \text{ かつ } x_l = 1\}$$

- 同様にして

$$P(A_k \cap A_l) = \sum_{\omega \in A_k \cap A_l} P(\{\omega\}) = \sum_{\omega \in A_k \cap A_l} \frac{1}{2^n} = \frac{|A_k \cap A_l|}{2^n} = \frac{2^{n-2}}{2^n} = \frac{1}{4}$$

1.3 条件付き確率

定義 3 $(\Omega, \mathcal{F}, P)$ を確率空間とする． $A \in \mathcal{F} (P(A) \neq 0)$ に対して，条件付き確率 (conditional probability) を以下で定義する．

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

注意 2

- $P(B|A)$ は事象 A の確率が 1 になるように「規格化」したもの．
- 事象 A が生じたもとで，事象 B が起きる確率．

補題 3 $A \in \mathcal{F} (P(A) \neq 0)$ を固定すると，写像

$$P(\cdot|A) : B \in \mathcal{F} \mapsto P(B|A) \in \mathbb{R}$$

は Ω 上の確率測度である．

証明： $P(\cdot|A)$ が確率測度の公理 (1) ~ (3) を満たすことを示す．

(1) 任意の $B \in \mathcal{F}$ について，定義式より $P(B|A) \geq 0$ である．

(2) 定義式より，

$$P(\Omega|A) = \frac{P(A \cap \Omega)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

(3) $B, C \in \mathcal{F}$ が互いに素であるとする，定義式より，

$$\begin{aligned} P(B \cup C|A) &= \frac{P(A \cap (B \cup C))}{P(A)} = \frac{P((A \cap B) \cup (A \cap C))}{P(A)} \quad (\because \text{分配法則}) \\ &= \frac{P(A \cap B)}{P(A)} + \frac{P(A \cap C)}{P(A)} = P(B|A) + P(C|A) \end{aligned}$$

ただし， $A \cap B$ と $A \cap C$ が互いに素であることと，公理 (3) を用いた． □

注意 3 (復習)

- 分配法則：

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

- ドモルガンの法則：

$$(A \cap B)^c = A^c \cup B^c, \quad (A \cup B)^c = A^c \cap B^c$$

例 4 子供が二人いる家族に会ったところ，男子がいることが分かった．男子がもう一人いる確率を求めよ．ただし，男子と女子が生まれる確率はそれぞれ $\frac{1}{2}$ とする．

- b=“boy”, g=“girl” とすると, 標本空間は

$$\Omega = \{bb, bg, gb, gg\}$$

- 確率測度: $P(\{\omega\}) = \frac{1}{4} \ (\omega \in \Omega)$
- 男子がいる事象: $A = \{bb, bg, gb\}$
- 男子がもう一人いる事象=男子が二人いる事象 $B = \{bb\}$
- 条件付き確率:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{3/4} = \frac{1}{3}$$

レポート 2 男子と女子が生まれる確率はそれぞれ $\frac{1}{2}$ とする. 子供が二人いる家族に会ったところ「第一子が男子」であることが分かった.

(1) 子供の男女構成についての標本空間 Ω を示せ (既出).

(2) 男子がもう一人いる確率を求めよ.

補題 4 (Chain rule)

- $P(A \cap B) = P(A)P(B|A)$ ($\because$ 条件付き確率の定義式)
- $P(A \cap B \cap C) = P(A \cap B)P(C|A \cap B) = P(A)P(B|A)P(C|A \cap B)$
- 同様にして, 帰納的に以下が示される.

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) \prod_{i=2}^n P(A_i | A_1 \cap A_2 \cap \cdots \cap A_{i-1})$$

1.4 ベイズの公式

定理 1 (全確率の公式) $A_1, A_2, \dots, A_n \in \mathcal{F}$ が互いに素で, $\Omega = \bigcup_{i=1}^n A_i$ のとき,

$$P(B) = \sum_{i=1}^n P(A_i \cap B) \quad (\forall B \in \mathcal{F})$$

証明: $B = \bigcup_{i=1}^n (A_i \cap B)$ で, $A_i \cap B$ は互いに素なので, 有限加法性 [確率測度の性質 (5)] を用いればよい. $\square$

定理 2 (ベイズの公式, Bayes' rule) $A_1, A_2, \dots, A_n \in \mathcal{F}$ が互いに素で, $\Omega = \bigcup_{i=1}^n A_i$ のとき, $\forall B \in \mathcal{F}$ に対して, $P(B) \neq 0$ ならば,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^n P(A_i)P(B|A_i)} \quad (i = 1, 2, \dots, n)$$

証明：

$$\begin{aligned}
 P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} \\
 &= \frac{P(A_i \cap B)}{\sum_{i=1}^n P(A_i \cap B)} \quad (\because \text{全確率の公式}) \\
 &= \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^n P(A_i)P(B|A_i)} \quad (\because \text{chain rule})
 \end{aligned}$$

□

レポート 3 ある製品を作る機械が 3 台あって、それらを A, B, C とする。 A, B, C はそれぞれ全体の 20%, 30%, 50% を生産する。また A, B, C の各機械から生産される製品のうち、5%, 4%, 2% の割合で不良品があることが経験的に知られている。

- (1) 製品全体の中から 1 個を取り出したとき、それが不良品である確率を求めよ。
- (2) 製品が不良品であることを知ったとき、それが A, B, C の各機械から生産されたものである確率を求めよ。

例 5 (クイズショーの問題, Monty Hall problem) カーテンで中が見えない部屋 A, B, C のいずれか一つに車が入っている。ゲストは車が入っている部屋を言い当てることができれば、その車が手に入る。司会者は、ゲストが部屋を一つ指定した後、指定された部屋以外で、車が入っていない部屋のカーテンを空け、「部屋を代えますか？」とゲストに尋ねた。ゲストは部屋を代えた方が良いであろうか？

解答：問題の対称性から、ゲストが指定した部屋は A であるとして一般性を失わない。この仮定のもとで、標本空間を「車が入っている部屋 x 」と「司会者が空けた部屋 y 」の組全体

$$\Omega = \{(x, y) \mid x = A, B, C, y = A, B, C\}$$

と設定してよい。ただし、ゲストが部屋 A を指定した場合、司会者が部屋 A を空けることはない。このような事象の確率は 0 とする。以下では C_A, C_B, C_C を、それぞれ、車が部屋 A, B, C にある事象であるとする。また、 O_A, O_B, O_C を、それぞれ、司会者が部屋 A, B, C を空ける事象であるとする。車が入っている部屋は完全にランダムであるとして、

$$P(C_A) = P(C_B) = P(C_C) = \frac{1}{3}$$

とする。司会者は車が部屋 B にあるときは、必ず部屋 C を空け、車が部屋 C にあるときは、必ず部屋 B を空けることになる。一方、車が部屋 A にあるとき、司会者は確率 p で部屋 B を空け、確率 $1 - p$ で部屋 C を空けると仮定する。これらを条件付き確率で表わすと、以下ようになる。

$$\begin{aligned}
 P(O_A|C_A) &= 0, & P(O_B|C_A) &= p, & P(O_C|C_A) &= 1 - p \\
 P(O_A|C_B) &= 0, & P(O_B|C_B) &= 0, & P(O_C|C_B) &= 1 \\
 P(O_A|C_C) &= 0, & P(O_B|C_C) &= 1, & P(O_C|C_C) &= 0
 \end{aligned}$$

C_A, C_B, C_C は互いに素で $\Omega = C_A \cup C_B \cup C_C$ であるから，ベイズの公式により³，

$$\begin{aligned}
 P(C_A|O_B) &= \frac{P(C_A \cap O_B)}{P(O_B)} \\
 &= \frac{P(C_A \cap O_B)}{P(C_A \cap O_B) + P(C_B \cap O_B) + P(C_C \cap O_B)} \quad (\because \text{全確率の公式}) \\
 &= \frac{P(C_A)P(O_B|C_A)}{P(C_A)P(O_B|C_A) + P(C_B)P(O_B|C_B) + P(C_C)P(O_B|C_C)} \quad (\because \text{chain rule}) \\
 &= \frac{1/3 \cdot p}{1/3 \cdot p + 1/3 \cdot 0 + 1/3 \cdot 1} \\
 &= \frac{p}{p+1}
 \end{aligned}$$

また， $P(C_B|O_B) = 0$ ， $P(C_A|O_B) + P(C_B|O_B) + P(C_C|O_B) = 1$ であるから，

$$P(C_C|O_B) = 1 - P(C_A|O_B) - P(C_B|O_B) = \frac{1}{p+1}$$

ここで $0 \leq p \leq 1$ に注意すれば，

$$P(C_A|O_B) = \frac{p}{p+1} \leq \frac{1}{p+1} = P(C_C|O_B) \quad (\text{等号成立} \Leftrightarrow p = 1)$$

が分かる．司会者が部屋 C を空けた場合については，同様に計算しても良いが，部屋 B と C の立場の対等性（対称性）を考えれば， $P(C_A|O_B)$ ， $P(C_C|O_B)$ において p の部分を $1-p$ に置き換えることで， $P(C_A|O_C)$ ， $P(C_B|O_C)$ が求まる．すなわち，

$$P(C_A|O_C) = \frac{(1-p)}{(1-p)+1}, \quad P(C_B|O_C) = \frac{1}{(1-p)+1},$$

また $0 \leq 1-p \leq 1$ より

$$P(C_A|O_C) = \frac{(1-p)}{(1-p)+1} \leq \frac{1}{(1-p)+1} = P(C_B|O_C) \quad (\text{等号成立} \Leftrightarrow p = 0)$$

が分かる．以上により，部屋を代えた方が良いことが分かる．特に $p = 1/2$ のときは次式となる．

$$P(C_A|O_B) = P(C_A|O_C) = \frac{1}{3}, \quad P(C_C|O_B) = P(C_B|O_C) = \frac{2}{3}$$

1.5 独立性

定義 4 (二つの事象の独立性) 事象 $A, B \in \mathcal{F}$ が独立 $\stackrel{\text{def}}{\Leftrightarrow} P(A \cap B) = P(A)P(B)$

注意 4 $P(A) \neq 0$ のとき，条件付き確率の定義から，以下が成立する．

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

独立性とは，条件付確率 $P(B|A)$ が条件部分によらず $P(B)$ に等しいことであると言っても良い．

注意 5 以下の命題は同値である

³ベイズの公式を中途半端に暗記すると間違えるので，このように導出できるようにしておきましょう．

$$(1) P(A \cap B) = P(A)P(B) \quad (A, B \text{ が独立})$$

$$(2) P(A) \neq 0 \Rightarrow P(B|A) = P(B)$$

証明：

- $(1) \Rightarrow (2)$: $P(A) \neq 0$ のとき，前注意より $P(B|A) = P(B)$ が成り立つので， (2) の命題は真である． $P(A) = 0$ のとき，条件部分が偽になるので (2) の命題は真である．
- $(2) \Rightarrow (1)$: $P(A) \neq 0$ のとき， (2) より $P(B|A) = P(B)$ が成り立つ．よって，前注意より (1) が成立する． $P(A) = 0$ のとき， $A \cap B \subset A$ であるから，

$$0 \leq P(A \cap B) \leq P(A) = 0$$

となって $P(A \cap B) = 0$ が示される．このとき (1) は両辺が 0 となって成立する．

□

例 6 (おみくじ) n 本の「おみくじ」のうち，当たりくじが m 本入っている．二人の人がくじを引くとき，最初の人引いたくじは戻さないとする．

(1) 最初に引く人と後で引く人のどちらが有利か？

A を最初に引く人が当たりくじを引く事象， B を二番目に引く人が当たりくじを引く事象とする．最初に引く人が当たる確率は

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

となる．二番目に引く人が当たる確率は，全確率の公式を用いることで，

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^c \cap B) \\ &= P(A)P(B|A) + P(A^c)P(B|A^c) \\ &= \frac{m}{n} \cdot \frac{m-1}{n-1} + \frac{n-m}{n} \cdot \frac{m}{n-1} \\ &= \frac{m\{(m-1) + (n-m)\}}{n(n-1)} \\ &= \frac{m}{n} \end{aligned}$$

となる．よって，くじを引く順番は関係ない．

(2) 最初の人引く事象と，二番目の人引く事象は独立か？

$$P(B|A) = \frac{m-1}{n-1} \neq \frac{m}{n} = P(B)$$

であるから（もちろん）独立ではない．

定義 5 (三つの事象の独立性) 事象 $A, B, C \in \mathcal{F}$ が独立であるとは，以下の式が同時に成立することである．

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

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

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

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

例 7 (二つの事象は独立だが, 三つの事象は独立でない例) 公正なコインを 2 回投げて,

$$x = \begin{cases} 1 & \text{最初のコインが表のとき} \\ 0 & \text{最初のコインが裏のとき} \end{cases}, \quad y = \begin{cases} 1 & \text{二番目のコインが表のとき} \\ 0 & \text{二番目のコインが裏のとき} \end{cases}$$

とおく. また $z = x + y \pmod{2}$ とする. このとき, (x, y, z) についての標本空間は

$$\Omega = \{(x, y, z) \mid x, y \in \{0, 1\}, z = x + y \pmod{2}\} = \{000, 011, 101, 110\}$$

で, 確率測度は次式で与えられる.

$$P(\{000\}) = P(\{011\}) = P(\{101\}) = P(\{110\}) = \frac{1}{4}$$

ここで, A を x が 1 となる事象 $A = \{\omega \in \Omega \mid x = 1\}$ であるとする. 同様に B, C を $B = \{\omega \in \Omega \mid y = 1\}$, $C = \{\omega \in \Omega \mid z = 1\}$ とおく. これらの事象の二つの組について,

$$P(A \cap B) = P(\{110\}) = \frac{1}{4}, \quad P(B \cap C) = P(\{011\}) = \frac{1}{4}, \quad P(C \cap A) = P(\{101\}) = \frac{1}{4}$$

であるから以下が成立する.

$$P(A \cap B) = P(A)P(B), \quad P(B \cap C) = P(B)P(C), \quad P(C \cap A) = P(C)P(A)$$

すなわち, 二つの事象の組はそれぞれ独立である. 一方, $A \cap B \cap C = \phi$ であるから,

$$P(A \cap B \cap C) = 0 \neq \frac{1}{8} = P(A)P(B)P(C)$$

となる. よって, 三つの事象 A, B, C は独立でない.

定義 6 (一般の独立性) 事象 $A_1, A_1, \dots, A_n \in \mathcal{F}$ が独立であるとは, これらのすべての組み合わせ $A_{i_1}, A_{i_2}, \dots, A_{i_m}$ ($1 \leq i_1 < i_2 < \dots < i_m \leq n, m = 2, 3, \dots, n$) について次式が成立することである.

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_m})$$

補題 5 事象 $A, B \in \mathcal{F}$ について以下は同値である.

(1) A と B は独立

(2) A と B^c は独立

証明: 最初に (1) $\Rightarrow$ (2) を示す. $P(A \cap B) = P(A)P(B)$ (A と B は独立) とする. 全確率の公式より $P(A \cap B) + P(A \cap B^c) = P(A)$ であるから,

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \\ &= P(A)\{1 - P(B)\} \\ &= P(A)P(B^c) \end{aligned}$$

よって (2) が示された. B の代わりに B^c として (1) $\Rightarrow$ (2) を適用すれば, (2) $\Rightarrow$ (1) が示される.

□

2 確率空間 (一般論に向けて)

前節では公理的設定のもとで、有限標本空間 Ω 上の確率空間 $(\Omega, \mathcal{F}, P)$ を導入した。なぜ、こんな面倒な定義をするのか？

(1) 標本空間の要素 $\omega \in \Omega$ に確率を割り当てる定義ではダメなのか？(補題 2, 注意 1 を参照)

(2) なぜ、わざわざ $\mathcal{F} = 2^\Omega$ とおくのか？

例 8 (ルーレット) ルーレットを回して、針が止まる位置について考える。簡単のため、ルーレットの円周は長さ 1 であるとする。また、真上を 0 として、時計回りに 0 から 1 未満の目盛が振られているとする⁴。

- 標本空間： $\Omega = [0, 1) = \{\omega \in \mathbb{R} \mid 0 \leq \omega < 1\}$
- ルーレットの針が止まる位置が「ランダム」のとき，

$$P([0, 0.5]) = 0.5, \quad P([0.8, 0.9]) = 0.1$$

のように、「長さの割合」や「面積の割合」に応じて確率を定めるのが自然であろう。

- このとき、針がちょうど 0.5 を指す確率は？
 $\varepsilon > 0$ として，

$$\begin{aligned} P([0.5 - \varepsilon, 0.5 + \varepsilon]) &= 2\varepsilon \\ \downarrow \quad (\varepsilon \rightarrow 0) \\ P(\{0.5\}) &= 0 \end{aligned}$$

- 同様に，すべての根元事象の確率は 0 になる。すなわち

$$\forall \omega \in \Omega, \quad P(\{\omega\}) = 0$$

上記 (1) の考え方は，うまく行かないことが分かる。

- 「事象 A の確率 = A の長さの総和」とするのが自然であろう。
- 例： $A = [0, 0.2] \cup (0.5, 0.6)$

$$P(A) = 0.2 + 0.1 = 0.3$$

- 例： $A = \bigcup_{k=1}^{\infty} \left(\sum_{l=1}^k \frac{1}{2^l}, \sum_{l=1}^{k+1} \frac{1}{2^l} \right)$

$$P(A) = \sum_{k=1}^{\infty} \frac{1}{2^{k+1}} = \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = \frac{1}{2}$$

⁴区間 $[0, 1)$ の実数すべてを目盛に振ることはできない (無限の精度で針の位置を測定することはできない)。任意の有限精度で針の位置が測れることを想定し、極限的に理想化した設定である。

- しかし「長さを測れない部分集合 $A \subset \Omega$ が存在」することが知られている⁵．よって 2^Ω の要素すべてを事象とすると不都合が生じる．そこで，

$$\mathcal{F} = \{A \subset \Omega \mid A \text{ は長さを測れる (可測)}\} \subsetneq 2^\Omega$$

として， $\mathcal{F}$ の要素のみを事象とよぶ．

- このとき， $\mathcal{F}$ は以下の性質を満たす．

- (1) $\Omega \in \mathcal{F}$
- (2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- (3) $A_i \in \mathcal{F} (i = 1, 2, \dots) \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

これらを用いると，以下が容易に示される．

- (4) $\phi \in \mathcal{F}$
- (5) $A_i \in \mathcal{F} (i = 1, 2, \dots) \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

一般の場合（有限標本空間も含む）「事象」は以下で定義される「 σ -代数」の要素であるとする．

定義 7 (σ -代数) 集合 Ω に対して， $\mathcal{F} \subset 2^\Omega$ が上記 (1)–(3) を満たすとき， $\mathcal{F}$ は Ω 上の σ -代数であるという．

注意 6

- $\mathcal{F} \subset 2^\Omega$ が σ -代数 $\stackrel{\text{def}}{\iff} \mathcal{F}$ が上記 (1)–(3) を満たす． $\iff \mathcal{F}$ が上記 (1)–(5) を満たす．
- 有限集合 Ω について， $\mathcal{F} = 2^\Omega$ はもちろん σ -代数である．
- σ -代数とは，部分集合の集まり $\mathcal{F} \subset 2^\Omega$ で

「全体 Ω を含み，集合演算 $\cup, \cap, ^c$ について閉じたもの」

である．

- $(\Omega, \mathcal{F})$ を可測空間とよぶ．

定義 8 (確率測度の公理) 集合 Ω と Ω 上の σ -代数 $\mathcal{F}$ が与えられているとする．写像 $P: \mathcal{F} \rightarrow \mathbb{R}$ が以下の性質を満たすとき， P を確率測度とよび， $(\Omega, \mathcal{F}, P)$ を確率空間 (probability space) とよぶ．

- (1) $\forall A \in \mathcal{F}, 0 \leq P(A) \leq 1$
- (2) $P(\Omega) = 1$
- (3) (σ -加法性) $A_i \in \mathcal{F} (i = 1, 2, \dots)$ が互いに素ならば

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

⁵測度論 (measure theory) や Lebesgue 積分論で示される．部分集合 $A \subset \Omega$ の「長さ」や「面積」が測れるとき， A は「可測」であると言い，そうでないとき A は「非可測」であると言う．

補題 6 (確率測度の性質)

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

(2) (有限加法性) $A_i \in \mathcal{F}$ ($i = 1, 2, \dots, n$) が互いに素ならば

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

(3) $\forall A \in \mathcal{F}, P(A) + P(A^c) = 1$

(4) $\forall A, \forall B \in \mathcal{F}, A \subset B \Rightarrow P(A) \leq P(B)$

(5) $\forall A, \forall B \in \mathcal{F}, P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(6) (劣加法性) 任意の事象列 $A_i \in \mathcal{F}$ ($i = 1, 2, \dots$) について

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

証明 :

(1) $A_i = \phi$ ($i = 1, 2, \dots$) とおくと A_i は互いに素である . $\phi = \bigcup_{i=1}^{\infty} A_i$ であるから , 公理 (3) の σ -加法性を用いると ,

$$P(\phi) = \sum_{i=1}^{\infty} P(\phi)$$

が成り立つ . 公理 (1) より $P(\phi)$ は有限であるから $P(\phi) = 0$ となる .

(2) A_i ($i = 1, 2, \dots, n$) が互いに素であるとする . $A_i = \phi$ ($i = n+1, n+2, \dots$) とおくと A_i ($i = 1, 2, \dots, n, n+1, \dots$) は互いに素である . 公理 (3) の σ -加法性と $P(\phi) = 0$ 用いると ,

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

(3)(4)(5) 確率測度の公理 (1), (2) と有限加法性は , 初学者バージョンの公理 (1), (2), (3) を含むので , 補題 1 と同様に証明できる (既に示した) .

(6) $B_1 = A_1, B_i = A_i \setminus \bigcup_{j=1}^{i-1} A_j$ ($i = 2, 3, \dots$) とおくと ,

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i, \quad B_i \subset A_i \quad (i = 1, 2, \dots)$$

で B_i ($i = 1, 2, \dots$) は互いに素である . よって , σ -加法性と (4) を用いることで ,

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

□

注意 7 全確率の公式 , chain rule , ベイズの公式が同様に証明できる .

注意 8 (集合族の和集合, 共通部分) Ω の部分集合族 $\{A_i\}_{i \in \mathbb{N}}$ について,

$$(1) \bigcap_{i=1}^{\infty} A_i := \{\omega \in \Omega \mid \forall i \in \mathbb{N}, \omega \in A_i\}$$

$$(2) \bigcup_{i=1}^n A_i := \{\omega \in \Omega \mid \exists i \in \mathbb{N}, \omega \in A_i\}$$

(3) 分配法則やドモルガンの法則が同様に成立する.

3 確率変数

3.1 確率変数の定義 (ポイント: 確率変数は関数)

例 9 公正なコインを 3 回投げる.

- 標本空間: $\Omega = \{0, 1\}^3$
- 確率測度: $P(\{\omega\}) = \frac{1}{8} \quad (\omega \in \Omega)$

$X =$ 「表の出た回数 $\times 100$ 円」もらえんとする. もらえる金額についての確率は?

ω	$X(\omega)$	確率
000	0 円	$P(\{\omega \in \Omega \mid X(\omega) = 0\}) = \frac{1}{8}$
001 010 100	100 円	$P(\{\omega \in \Omega \mid X(\omega) = 100\}) = \frac{3}{8}$
011 101 110	200 円	$P(\{\omega \in \Omega \mid X(\omega) = 200\}) = \frac{3}{8}$
111	300 円	$P(\{\omega \in \Omega \mid X(\omega) = 300\}) = \frac{1}{8}$

定義 9 (確率変数) $(\Omega, \mathcal{F}, P)$ を確率空間とするととき, 関数

$$X: \omega \in \Omega \mapsto X(\omega) \in \mathbb{R}$$

を (実 / 実数に値を取る / 実数値) 確率変数 (random variable) という.

注意 9

- 本当は以下の条件を要請する必要がある.

$$\forall x \in \mathbb{R}, \quad \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F} \quad (1)$$

この条件は X の値が区間 $(-\infty, x]$ に入る確率 $P(\{\omega \in \Omega \mid X(\omega) \leq x\})$ が計算できるための条件である.

- 一般には $X(\omega)$ の値が実数とは限らない. 集合 $\mathcal{X}$ に対して, 関数

$$X: \Omega \rightarrow \mathcal{X}$$

を $\mathcal{X}$ に値を取る確率変数 ($\mathcal{X}$ -valued random variable) という.

- $\mathcal{X} = \{a, b, c, \dots, z\}$ (アルファベット)
- $\mathcal{X} = \mathbb{R}^n$ (n 次元ベクトル)
- $\mathcal{X} = \{\text{晴}, \text{くもり}, \text{雨}\}$

よく用いられる確率変数として, 「離散型確率変数」と「連続型確率変数」がある. 以下では, これらを扱う.

3.2 像と逆像

定義 10 (像, 逆像) $f: \mathcal{X} \rightarrow \mathcal{Y}$ を写像とする.

- (1) 部分集合 $A \subset \mathcal{X}$ について, $f(A) := \{f(x) \mid x \in A\} \subset \mathcal{Y}$ を A の像という.
- (2) 部分集合 $B \subset \mathcal{Y}$ について, $f^{-1}(B) := \{x \in \mathcal{X} \mid f(x) \in B\} \subset \mathcal{X}$ を B の逆像という.

注意 10

- 写像 $f: \mathcal{X} \rightarrow \mathcal{Y}$ が全単射のとき, 「逆写像」 $f^{-1}: \mathcal{Y} \rightarrow \mathcal{X}$ が定義できる. まぎらわしい記号であるが, 「逆写像」と「逆像」を混同してはいけない.
- 確率変数についての条件 (1) は, 逆像を用いると以下のように書ける.

$$\forall x \in \mathbb{R}, \quad X^{-1}((-\infty, x]) \in \mathcal{F} \quad (2)$$

補題 7 (和集合・共通部分の像・逆像) $f: \mathcal{X} \rightarrow \mathcal{Y}$ を写像とする. 部分集合 $A, B \subset \mathcal{X}, C, D \subset \mathcal{Y}$ に対して以下が成り立つ.

- (1) $f(A \cap B) \subset f(A) \cap f(B)$
- (2) $f(A \cup B) = f(A) \cup f(B)$
- (3) $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$
- (4) $f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$

レポート 4

- (1) 補題 7 の (1) について, $f(A \cap B) = f(A) \cap f(B)$ は成立しない. 反例を与えよ.
- (2) 補題 7 の (3)(4) を証明せよ.

レポート 5 (復習: 独立性) 一つの壺に 5 個の赤玉と 7 個の白玉が入っている. まず一つの玉を取り出し, その玉を壺に戻さないでもう一つ取り出す. 「 A を最初に取り出した玉が赤色である事象」, 「 B を次に取り出した玉が赤色である事象」とする. 事象 A と B は独立か?

3.3 可算集合と非可算集合

定義 11 (単射, 全射, 全単射) $f: \mathcal{X} \rightarrow \mathcal{Y}$ を写像とする.

- (1) f が単射 (injection) $\stackrel{\text{def}}{\iff} \forall x_1, x_2 \in \mathcal{X}, f(x_1) = f(x_2) \Rightarrow x_1 = x_2$
- (2) f が全射 (surjection) $\stackrel{\text{def}}{\iff} \forall y \in \mathcal{Y}, \exists x \in \mathcal{X}, y = f(x)$
- (3) f が全単射 (bijection) $\stackrel{\text{def}}{\iff} f$ が単射かつ全射

定義 12 (可算無限)

- (1) 集合 $\mathcal{X}$ が可算 (countable) $\stackrel{\text{def}}{\iff} \exists g: \mathcal{X} \rightarrow \mathbb{N}$ (全単射)

(2) 集合 $\mathcal{X}$ が高々可算 (at most countable) $\stackrel{\text{def}}{\iff} \mathcal{X}$ が有限または可算

(3) 集合 $\mathcal{X}$ が非可算 $\stackrel{\text{def}}{\iff} \mathcal{X}$ は高々可算ではない

例 10 ($\mathbb{Z}, \mathbb{Q}$ は可算)

- $\mathbb{Z}$ (整数全体) は可算 :

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$$

$$\dots \quad 5 \quad 3 \quad 1 \quad 2 \quad 4 \quad \dots$$

- $\mathbb{N} \times \mathbb{N} = \{(m, n) \mid m, n \in \mathbb{N}\}$ は可算 :

	1	2	3	4
1	1	2	4	7
2	3	5	8	
3	6	9	$\ddots$	
4	10			

- $\mathbb{Q}_+$ (正の有理数全体) は可算 :

$$\mathbb{Q}_+ = \left\{ \frac{n}{m} \mid m, n \in \mathbb{N}, m, n > 0 \right\}$$

のように自然数の比で書ける . 上記のように斜めに自然数の番号を付けばよい . ただし , $\frac{n}{m}$ の中には , 約分されてそれまでと同じ有理数を与えることがあるので , その場合は飛ばして数える .

- $\mathbb{Q}$ (有理数全体) は可算 :

$$\mathbb{Q}_- = \left\{ -\frac{n}{m} \mid m, n \in \mathbb{N}, m, n > 0 \right\}$$

とおくと $\mathbb{Q}_-$ は可算である .

$$\mathbb{Q} = \mathbb{Q}_- \cup \{0\} \cup \mathbb{Q}_+$$

であるから , $\mathbb{Z}$ の場合と同様に自然数の番号を付けばよい .

定理 3 (カントールの対角線論法) 区間 $(0, 1)$ は非可算集合である .

証明 : 明らかに $(0, 1)$ は有限集合ではない . $(0, 1)$ が可算であるとして矛盾を導く .

(1) 実数は無限小数 (有理数の極限) で表わされる .

$$\frac{1}{3} = 0.3333333 \dots$$

$$\sqrt{5} = 2.2360679 \dots$$

$$\pi = 3.1415926 \dots$$

$$1 = 0.9999999 \dots$$

(2) 有理数は十進小数で二通りの表し方がある場合がある .

$$\begin{aligned} 1 &= 0.9999999 \cdots = 1.0000000 \cdots \\ 0.5 &= 0.4999999 \cdots = 0.5000000 \cdots \end{aligned}$$

このように 0 が無限に続く場合には , 最後の 0 でない数から 1 を引いて , 後に 9 を無限に並べておくことで , 一通りの無限小数で表わせる . ただし , 0 は $0.0000 \cdots$ とする .

(3) $(0, 1)$ が可算無限であると仮定すると , $(0, 1)$ は $\{x_n\}_{n=1}^{\infty}$ と書ける . よって , $(0, 1)$ の要素 x_n を十進小数で一意的に展開して , 以下のように並べることができる .

$$\begin{aligned} x_1 &= 0.x_{1,1}x_{1,2}x_{1,3}x_{1,4}x_{1,5} \cdots \\ x_2 &= 0.x_{2,1}x_{2,2}x_{2,3}x_{2,4}x_{2,5} \cdots \\ x_3 &= 0.x_{3,1}x_{3,2}x_{3,3}x_{3,4}x_{3,5} \cdots \\ x_4 &= 0.x_{4,1}x_{4,2}x_{4,3}x_{4,4}x_{4,5} \cdots \\ x_5 &= 0.x_{5,1}x_{5,2}x_{5,3}x_{5,4}x_{5,5} \cdots \\ &\vdots \end{aligned}$$

ここで ,

$$a_n = \begin{cases} 2, & x_{n,n} = 1 \\ 1, & x_{n,n} \neq 1 \end{cases}$$

として , $a = 0.a_1a_2a_3a_4a_5 \cdots \in (0, 1)$ とおく . すべての $n \in \mathbb{N}$ について , $a_n \neq x_{n,n}$ より $a \neq x_n$ である . これは , 上の表に $(0, 1)$ の要素がすべて現われていたことに反する .

□

3.4 累積分布関数

注意 11 (話しの流れ)

$$\text{確率変数} \begin{cases} \text{離散型確率変数 (確率の計算方法: 確率関数)} \\ \text{連続型確率変数 (確率の計算方法: 累積分布関数} \rightarrow \text{確率密度関数)} \\ \text{その他} \end{cases}$$

累積分布関数はすべての確率変数に対して定義される (離散型確率変数や連続型確率変数に限らない) .

定義 13 (累積分布関数) $(\Omega, \mathcal{F}, P)$ を確率空間とし , $X: \Omega \rightarrow \mathbb{R}$ を確率変数とする .

$$F_X(x) := P(\{\omega \in \Omega \mid X(\omega) \leq x\})$$

を X の (累積) 分布関数 (distribution function) という .

注意 12

- 右辺を $\Pr\{\omega \in \Omega \mid X(\omega) \leq x\}$ と書くこともある (Pr は probability の意味) . また , 省略して $P(X \leq x)$, $\Pr\{X \leq x\}$ のように書くことが多い . 慣れないうちは省略しない方が良い .
- $F_X(x)$ により確率変数 X の性質はすべて定まる .
- $F_X(x)$ の添字 X は関数名を区別するための添字 . 確率変数 X, Y, Z の累積分布関数を $F(x), G(y), H(z)$ のように区別しても良いが , 文字が足りなくなるので , $F_X(x), F_Y(y), F_Z(z)$ のように区別する . 文脈から明らかなきは , 添字は省略して $F(x), F(y), F(z)$ と書くこともある . ここでも以降では省略すること多い .

例 11 (ルーレット) ルーレットの例 (例 8) では針の位置についての確率空間 $(\Omega, \mathcal{F}, P)$ は以下で与えられた .

$$\begin{aligned}\Omega &= [0, 1) = \{\omega \in \mathbb{R} \mid 0 \leq \omega < 1\} \\ \mathcal{F} &= \{A \subset \Omega \mid A \text{ は長さを測れる (可測)}\} \subsetneq 2^\Omega \\ P(A) &= \frac{A \text{ の長さ}}{\text{全事象 } \Omega \text{ の長さ}}\end{aligned}$$

このとき , 恒等写像 $X(\omega) = \omega$ は確率変数で , 累積分布関数は以下で与えられる .

$$F_X(x) = P(\{\omega \in \Omega \mid \omega \leq x\}) = \begin{cases} 0 & (x < 0) \\ x & (0 \leq x < 1) \\ 1 & (x \geq 1) \end{cases}$$

例 12 公正なコインを 3 回投げる例 (例 9) で , $X(\omega) = \text{「表の出た回数} \times 100 \text{ 円」}$ もらえるとして , もらえる金額についての確率は以下で与えられた .

ω	$X(\omega)$	確率
000	0 円	$P(\{\omega \in \Omega \mid X(\omega) = 0\}) = \frac{1}{8}$
001 010 100	100 円	$P(\{\omega \in \Omega \mid X(\omega) = 100\}) = \frac{3}{8}$
011 101 110	200 円	$P(\{\omega \in \Omega \mid X(\omega) = 200\}) = \frac{3}{8}$
111	300 円	$P(\{\omega \in \Omega \mid X(\omega) = 300\}) = \frac{1}{8}$

このとき , 確率変数 X の累積分布関数は , 以下の階段状の関数である .

$$F_X(x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}) = \begin{cases} 0 & (x < 0) \\ \frac{1}{8} & (0 \leq x < 100) \\ \frac{4}{8} & (100 \leq x < 200) \\ \frac{7}{8} & (200 \leq x < 300) \\ 1 & (x \geq 300) \end{cases}$$

補題 8 (累積分布関数の性質)

- (1) (単調性) $a < b \Rightarrow F(a) \leq F(b)$
- (2) (右連続性) $\lim_{\varepsilon \rightarrow +0} F(x + \varepsilon) = F(x)$
- (3) $\lim_{x \rightarrow -\infty} F(x) = 0$
- (4) $\lim_{x \rightarrow +\infty} F(x) = 1$

証明には確率測度の連続性の議論が必要．現段階では省略する

3.5 離散型確率変数

定義 14 (離散型確率変数) $(\Omega, \mathcal{F}, P)$ を確率空間とし, $X: \Omega \rightarrow \mathbb{R}$ を確率変数とする． X の値域

$$\mathcal{X} := X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}$$

が高々可算のとき, X を離散型確率変数といい,

$$P_X(x) = P(\{\omega \in \Omega \mid X(\omega) = x\}) \quad (x \in \mathcal{X})$$

を X の確率関数 (probability function) という．

注意 13

- X が離散型確率変数とは, X の取りうる範囲 (値域) が離散的 (とびとび) であること．
- 右辺を $\Pr\{\omega \in \Omega \mid X(\omega) = x\}$ と書くこともある．また, 省略して $P(X = x)$, $\Pr\{X = x\}$ のように書くことが多い．慣れないうちは省略しない方がよい．
- 確率関数 $P_X(x)$ によって離散型確率関数 X の性質は定まるので, 確率空間 $(\Omega, \mathcal{F}, P)$ を表に出さないことが多い．
- 累積分布関数と同様に, $P_X(x)$ の添字 X は関数名を区別するための添字．
- 離散型確率変数 X の累積分布関数 $F_X(x)$ は階段状になる．

補題 9 (確率関数の性質)

- (1) $P_X(x) \geq 0 \quad (x \in \mathcal{X})$
- (2) $\sum_{x \in \mathcal{X}} P_X(x) = 1$

逆に, これらの条件 (0 以上, 足すと 1) を満たす関数 $P_X(x)$ が与えられると, 離散型確率変数 X が定義されたことになる．

3.6 連続型確率変数

定義 15 (連続型確率変数)

- $(\Omega, \mathcal{F}, P)$ を確率空間とし, $X: \Omega \rightarrow \mathbb{R}$ を確率変数とする. 累積分布関数 $F_X(x)$ が

$$F_X(x) = \int_{-\infty}^x f_X(x) dx$$

と書けるとき, X を連続型確率変数という.

- 微分と積分の関係により次式が成り立つ.

$$f_X(x) = \frac{dF_X(x)}{dx}$$

$f_X(x)$ を X の確率密度関数 (probability density function) という.

注意 14

- X が連続型確率変数とは, X の取りうる範囲 (値域) が連続的であること.
- 確率密度関数 $f_X(x)$ によって連続型確率関数 X の性質は定まるので, 確率空間 $(\Omega, \mathcal{F}, P)$ を表に出さないことが多い.
- これまでと同様, $f_X(x)$ の添字 X は関数名を区別するため. 省略されることもある.
- 連続型確率変数 X の累積分布関数 $F_X(x)$ は連続的になる.

補題 10 (確率密度関数の性質)

(1) $f_X(x) \geq 0$ ($f_X(x) \leq 1$ とは限らないことに注意)

$$(2) \int_{-\infty}^{\infty} f_X(x) dx = 1$$

逆に, これらの条件 (0 以上, 積分すると 1) を満たす関数 $f_X(x)$ が与えられると, 連続型確率変数 X が定義されたことになる.

注意 15 (確率密度関数の直感的意味)

- 確率変数 $X(\omega)$ が区間 $(a, b]$ に入る確率は以下で与えられる.

$$\begin{aligned} \Pr \{ \omega \in \Omega \mid a < X(\omega) \leq b \} &= \Pr \{ \omega \in \Omega \mid X(\omega) \leq b \} - \Pr \{ \omega \in \Omega \mid X(\omega) \leq a \} \\ &= \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx \\ &= \int_a^b f(x) dx \end{aligned}$$

- $a = x, a = x + \Delta x$ とおくと,

$$\begin{aligned} \Pr \{ \omega \in \Omega \mid x < X(\omega) \leq x + \Delta x \} &= \int_x^{x+\Delta x} f(x) dx \\ &\simeq f(x) \cdot \Delta x \quad (\Delta x \text{ が十分小さいとき}) \end{aligned}$$

$X(\omega)$ が微小な幅 Δx の区間 $(x, x + \Delta x]$ に入る確率は $f(x) \cdot \Delta x$ で与えられる.

● 連続型確率変数の場合

$$\Pr \{ \omega \in \Omega \mid X(\omega) = a \} = \int_a^a f(x) dx = 0$$

であるから，一点の確率には意味がない．微小でも良いので幅を考えることが重要．

例 13 (二つのルーレット)

問題：0 から 1 の目盛が振られたルーレットを二つ回す．それぞれの針が指す目盛 α, β についての確率は，理想化すると，以下の確率空間 $(\Omega, \mathcal{F}, P)$ で記述される．

$$\text{標本空間} : \Omega = [0, 1) \times [0, 1) = \{(\alpha, \beta) \mid 0 \leq \alpha < 1, 0 \leq \beta < 1\}$$

$$\sigma\text{-代数} : \mathcal{F} = \{A \subset \Omega \mid \text{“}A \text{ は面積が測れる”}\}$$

$$\text{確率測度} : P(A) = \frac{A \text{ の面積}}{\Omega \text{ の面積}}$$

このとき，それぞれの針が指す目盛 α, β の最大値を与える関数

$$X : (\alpha, \beta) \in \Omega \mapsto \max\{\alpha, \beta\} \in \mathbb{R}$$

は確率変数である．確率変数 X の確率密度関数を求めよ．

解答：最初に X の（累積）分布関数を求める．

$$\begin{aligned} F(x) &= P(\{(\alpha, \beta) \in \Omega \mid X(\alpha, \beta) \leq x\}) \\ &= P(\{(\alpha, \beta) \in \Omega \mid \max\{\alpha, \beta\} \leq x\}) \\ &= P(\{(\alpha, \beta) \in \Omega \mid \alpha \leq x \text{ かつ } \beta \leq x\}) \\ &(\because \max\{\alpha, \beta\} \leq x \Leftrightarrow \alpha \leq x \text{ かつ } \beta \leq x) \end{aligned}$$

面積を考えることで，確率分布関数が以下のように求められる．

$$F(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

これを微分することで確率密度関数が得られる．

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0 & x < 0 \\ 2x & 0 \leq x < 1 \\ 0 & x \geq 1 \end{cases}$$

レポート 6 例 13 と同じ二つのルーレットを考える．それぞれの針が指す目盛 α, β の和を与える確率変数

$$X : (\alpha, \beta) \in \Omega \mapsto \alpha + \beta \in \mathbb{R}$$

を考える． X の確率密度関数を求めよ．ヒント： $x < 0, 0 \leq x < 1, 1 \leq x < 2, x \geq 2$ で場合分け．

レポート 7 表の出る確率が p ，裏の出る確率が $1 - p$ のコインを 3 回投げる．標本空間 Ω は例 9 と同じである．確率変数

$$X : \Omega \rightarrow \text{表が出た回数}$$

を考える． X の確率関数を求めよ．解答は例 9 と同様の表を用いよ．

3.7 代表的な確率変数 (離散型)

3.7.1 二点分布 (binary distribution) $B(1, p)$ ($0 < p < 1$)

表の出る確率が p であるコインを 1 回投げる試行

$$\begin{aligned} \text{確率変数 } X &= \begin{cases} 1 & \text{表が出たとき (確率 } p) \\ 0 & \text{裏が出たとき (確率 } 1-p) \end{cases} \\ \text{確率関数 } P_X(1) &= p, \quad P_X(0) = 1-p \end{aligned}$$

コインを 1 回投げる試行をベルヌイ試行といい, X をベルヌイ確率変数とよぶ. 確率関数 P_X のことを二点分布とよぶ. 二点分布はベルヌイ分布 (Bernoulli distribution) ともよばれる.

3.7.2 二項分布 (binomial distribution) $B(n, p)$ ($n \in \mathbb{N}, 0 < p < 1$)

表の出る確率が p であるコインを n 回投げる時の表の回数

- 元になる確率空間 $(\Omega, \mathcal{F}, P)$

$$\begin{aligned} \Omega &= \{\omega = (\omega_1, \omega_2, \dots, \omega_n) \mid \omega_i = 0, 1 \ (i = 1, 2, \dots, n)\} \\ \mathcal{F} &= 2^\Omega \end{aligned}$$

$$P(\{\omega\}) = p^k(1-p)^{n-k} \quad \left(k = \sum_{i=1}^n \omega_i, \text{ 表が } k \text{ 回で裏が } n-k \text{ 回のとき} \right)$$

- 確率変数 $Y_n : \omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Omega \mapsto Y_n(\omega) = \sum_{i=1}^n \omega_i \in \mathbb{R}$
- 確率関数

$$\begin{aligned} P_{Y_n}(k) &= P(\{\omega \in \Omega \mid Y_n(\omega) = k\}) \\ &= (k \text{ 個の } 1 \text{ と } n-k \text{ 個の } 0 \text{ を並べる場合の数}) \times p^k(1-p)^{n-k} \\ &= (n \text{ 個の場所から } 1 \text{ の場所を } k \text{ 個選ぶ場合の数}) \times p^k(1-p)^{n-k} \\ &= {}_n C_k p^k(1-p)^{n-k} \end{aligned}$$

- 確率関数 $P_{Y_n}(k)$ のことを二項分布とよび, Y_n を二項分布に従う確率変数という.
- $X_1, X_2, \dots, X_n$ を独立なベルヌイ確率変数とするとき, $Y_n = \sum_{i=1}^n X_i$

注意 16 (復習: 順列・組合わせ)

- 順列 (permutation): 異なる n 個の物を, k 個並べる場合の数

$${}_n P_k = n(n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

- 組合わせ (combination) : 異なる n 個の物から , k 個取り出す場合の数

$${}_nC_k = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} = \frac{n!}{k!(n-k)!}$$

公式

$$\begin{aligned} {}_nC_k &= {}_nC_{n-k} \\ {}_{n+1}C_k &= {}_nC_k + {}_nC_{k-1} \end{aligned}$$

- 重複組合せ : n 種類の物から , 重複を許して r 個取り出す場合の数

$${}_nH_r = {}_{n+r-1}C_{n-1} = {}_{n+r-1}C_r$$

“取り出した r 個の物” と “種類を区別する $n-1$ 個の仕切り” を並べることを考える .

$$\circ \cdots \circ \mid \circ \cdots \circ \mid \cdots \mid \circ \cdots \circ$$

注意 17 (二項定理と二項分布)

$$\begin{aligned} (x+y)^n &= (x+y)(x+y)\cdots(x+y) = \sum_{k=0}^n x^k y^{n-k} \text{の項} \\ &= \sum_{k=0}^n (k \text{ 個の } x \text{ と } n-k \text{ 個の } y \text{ を並べる場合の数}) \times x^k y^{n-k} \\ &= \sum_{k=0}^n (n \text{ 個の場所から } x \text{ の場所を } k \text{ 個選ぶ場合の数}) \times x^k y^{n-k} \\ &= \sum_{k=0}^n {}_nC_k x^k y^{n-k} \quad \text{二項定理} \end{aligned}$$

ここで $x = p, y = 1 - p$ とおくと

$$\begin{aligned} \sum_{k=0}^n P_{Y_n}(k) &= \sum_{k=0}^n {}_nC_k p^k (1-p)^{n-k} \\ &= \{p + (1-p)\}^n \\ &= 1 \end{aligned}$$

$P_{Y_n}(k)$ が確率関数の資格を持つことが確認できた .

3.7.3 幾何分布 (geometrical distribution) $\text{Ge}(p)$ ($0 < p < 1$)

表の出る確率が p であるコインを表が出るまで投げたとき , 裏の出た回数 k

- 元になる確率空間 $(\Omega, \mathcal{F}, P)$ は二項分布と同じ
- 確率関数

$$P_X(k) = (1-p)^k p \quad (k = 0, 1, 2, \cdots)$$

- 確率関数の資格を確認

$$\begin{aligned}
 \sum_{k=0}^{\infty} P_X(k) &= \sum_{k=0}^{\infty} (1-p)^k p \\
 &= p \cdot \sum_{k=0}^{\infty} (1-p)^k \\
 &= p \cdot \frac{1}{1-(1-p)} \left[\because \text{無限等比級数の和} : \sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \quad (|x| < 1) \right] \\
 &= 1
 \end{aligned}$$

3.7.4 ポアソン分布 (Poisson distribution) $\text{Po}(\lambda)$ ($\lambda > 0$)

- 確率関数

$$P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (k = 0, 1, 2, \dots)$$

- 確率関数の資格確認

$$\begin{aligned}
 \sum_{k=0}^{\infty} P_X(k) &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \\
 &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
 &= e^{-\lambda} e^{\lambda} \left[\because \text{テーラー展開} : e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \right] \\
 &= 1
 \end{aligned}$$

ポアソン分布は「一定時間における事故件数」や「一定時間における放射線の放射回数」など（頻繁には起こらない）偶然現象の記述に用いられる．これは次の定理に基づく．

定理 4 (ポアソンの少数の法則) 二項分布 $B(n, p)$ は, $np = \lambda$ (一定) の関係を保って $n \rightarrow \infty$ とする時 (必然的に $p \rightarrow 0$ となる), ポアソン分布 $\text{Po}(\lambda)$ に近づく．すなわち, Y_n を二項分布 $B(n, p)$ に従う確率関数, X をポアソン分布 $\text{Po}(\lambda)$ に従う確率関数とすると,

$$P_{Y_n}(k) = {}_nC_k p^k (1-p)^{n-k} \rightarrow P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (np = \lambda, n \rightarrow \infty)$$

証明：

$$\begin{aligned}
 P_{Y_n}(k) &= {}_nC_k p^k (1-p)^{n-k} \\
 &= \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad (\because p = \lambda/n) \\
 &= \frac{n(n-1) \cdots (n-k+1)}{n^k} \cdot \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \\
 &\xrightarrow{n \rightarrow \infty} 1 \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot 1 \quad \left[\because k \text{ は定数, } \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x \right]
 \end{aligned}$$

□

例 14 (ポアソン近似) 1 年を 365 日とし, ある人の誕生日が今日である確率を $p = 1/365$ とする. n 人中 k 人の誕生日が今日である確率は

$$P_{Y_n}(k) = {}_nC_k \left(\frac{1}{365}\right)^k \left(1 - \frac{1}{365}\right)^{n-k}$$

n が十分大きいとき, $\lambda = np = n/365$ においてポアソン分布で近似すると,

$$P_{Y_n}(k) \simeq P_X(k) = e^{-n/365} \frac{(n/365)^k}{k!}$$

$n = 50$ のとき, 以下のように大変良い近似になっている (データ提供: IS 研究科 森田啓義 教授).

k	$P_{Y_n}(k)$	$P_X(k)$
0	0.87182	0.87198
1	0.11976	0.11945
2	0.00806	0.00818
3	0.00035	0.00037
4	0.00001	0.00001
$\vdots$		

3.8 代表的な確率変数 (連続型)

3.8.1 一様分布 (uniform distribution) $U(a, b)$ ($a < b$)

- 確率密度関数

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

a から b の区間で一様に発生する確率変数.

- 確率密度関数の資格

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} \cdot (b-a) = 1$$

3.8.2 指数分布 (exponential distribution) $Ex(\alpha)$ ($\alpha > 0$)

- 確率密度関数

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

- 確率密度関数の資格

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^{\infty} \alpha e^{-\alpha x} dx = [-e^{-\alpha x}]_0^{\infty} = 0 - (-1) = 1$$

3.8.3 正規分布 (normal distribution) $N(\mu, \sigma)$ ($\mu \in \mathbb{R}, \sigma > 0$)

- 確率密度関数

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- 確率密度関数の資格

補題 11 (ガウス積分)

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 \quad (3)$$

証明：最初に $\mu = 0, \sigma = 1$ の場合について

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1 \quad (4)$$

を示す．左辺の二乗が 1 になることを示せばよい．

$$\begin{aligned} \left\{ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right\}^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \cdot \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy \end{aligned} \quad (5)$$

ここで極座標への変換 $(x, y) \mapsto (r, \theta)$

$$x = r \cos \theta, \quad y = r \sin \theta \quad (r \geq 0, 0 \leq \theta < 2\pi)$$

を行う．このとき， $x^2 + y^2 = r(\cos^2 \theta + \sin^2 \theta) = r^2$ で，ヤコビ行列式 (ヤコビアン) は以下のように計算される．

$$\begin{aligned} \left| \frac{\partial(x, y)}{\partial(r, \theta)} \right| &= \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} \\ &= \cos \theta \cdot (r \cos \theta) - \sin \theta \cdot (-r \sin \theta) = r \cos^2 \theta + r \sin^2 \theta = r \end{aligned}$$

これより，微小体積要素は以下の変換を受ける．

$$dx dy = \left| \frac{\partial(x, y)}{\partial(r, \theta)} \right| dr d\theta = r dr d\theta$$

よって

$$(5) \text{ 式} = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\theta \quad (6)$$

変数変換 $t = \frac{r^2}{2}$ を行くと， $dt = r dr$ であるから，

$$(6) \text{ 式} = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-t} dt d\theta = \frac{1}{2\pi} \int_0^{2\pi} [-e^{-t}]_0^{\infty} d\theta = \frac{1}{2\pi} \int_0^{2\pi} 1 d\theta = \frac{1}{2\pi} \cdot 2\pi = 1$$

よって (4) が示された． μ, σ が一般の場合，変数変換 $y = \frac{x-\mu}{\sigma}$ を行う． $x = \sigma y + \mu$ より $dx = \sigma dy$ であるから，

$$(3) \text{ 式の左辺} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \sigma dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = 1$$

□

3.9 同時確率変数

3.9.1 同時確率変数の定義

定義 16 (同時確率変数) $(\Omega, \mathcal{F}, P)$ を確率空間とする。 Ω 上の二つの確率変数 $X : \Omega \rightarrow \mathbb{R}$ と $Y : \Omega \rightarrow \mathbb{R}$ が与えられたとき、確率変数の組 (X, Y) を同時確率変数という。これはベクトル値確率変数

$$\omega \in \Omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2$$

と見なすことができる。

注意 18 以下では、事象「 $X = x$ 」、事象「 $X = x$ かつ $Y = y$ 」、事象「 $X \leq x$ かつ $Y \leq y$ 」などの言い方をする。これらは以下の意味である。

- 事象「 $X = x$ 」: $\{\omega \in \Omega \mid X(\omega) = x\}$
- 事象「 $X = x$ かつ $Y = y$ 」: $\{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\}$
- 事象「 $X \leq x$ かつ $Y \leq y$ 」: $\{\omega \in \Omega \mid X(\omega) \leq x, Y(\omega) \leq y\}$

注意 19 (読み飛ばしても良い) 本当は関数 $X : \Omega \rightarrow \mathbb{R}, Y : \Omega \rightarrow \mathbb{R}$ が確率変数であるためには、(1) 式の条件：

$$\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}, \quad \{\omega \in \Omega \mid Y(\omega) \leq y\} \in \mathcal{F}$$

を満たす必要があった。これらの条件が満たされるとき、 σ -代数 $\mathcal{F}$ の性質から、

$$\{\omega \in \Omega \mid X(\omega) \leq x, Y(\omega) \leq y\} = \{\omega \in \Omega \mid X(\omega) \leq x\} \cap \{\omega \in \Omega \mid Y(\omega) \leq y\} \in \mathcal{F}$$

である。すなわち、 X と Y が確率変数ならば以下の同時累積分布関数が計算できる。

定義 17 (同時累積分布関数) 同時確率変数 (X, Y) について、事象「 $X \leq x$ かつ $Y \leq y$ 」の確率：

$$F_{XY}(x, y) := P(\{\omega \in \Omega \mid X(\omega) \leq x, Y(\omega) \leq y\})$$

を確率変数 X, Y の同時 (累積) 分布関数 (joint distribution function) という。

定義 18 (離散型同時確率変数, 同時確率関数) 同時確率変数 (X, Y) について、 X と Y の値域：

$$\mathcal{X} = X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}, \quad \mathcal{Y} = Y(\Omega) = \{Y(\omega) \mid \omega \in \Omega\}$$

が両方とも高々可算であるとき、事象「 $X = x$ かつ $Y = y$ 」の確率：

$$P_{XY}(x, y) := P(\{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\})$$

を確率変数 X, Y の同時確率関数 (joint probability function) という。

補題 12 (同時確率関数の性質)

$$P_{XY}(x, y) \geq 0, \quad \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y) = 1$$

定義 19 (連続型同時確率変数, 同時確率密度関数) X と Y はそれぞれ連続型確率変数であるとする. 同時確率変数 (X, Y) について, 同時累積分布関数 $F_{XY}(x, y)$ が

$$F_{XY}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{XY}(x, y) dx dy$$

と書ける場合, 微分と積分の関係により次式が成り立つ.

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$$

$f_{XY}(x, y)$ を X, Y の同時確率密度関数 (joint probability density function) という.

注意 20 (同時確率密度関数の意味) 一変数の場合と同様に, 以下の式が成り立つ.

$$\Pr \{ \omega \in \Omega \mid a < X(\omega) \leq b, c < Y(\omega) \leq d \} = \int_c^d \int_a^b f_{XY}(x, y) dx dy$$

Δx と Δy が十分小さいときは

$$\begin{aligned} \Pr \{ \omega \in \Omega \mid x < X(\omega) \leq x + \Delta x, y < Y(\omega) \leq y + \Delta y \} &= \int_x^{x+\Delta x} \int_y^{y+\Delta y} f_{XY}(x, y) dx dy \\ &\simeq f_{XY}(x, y) \Delta x \Delta y \end{aligned}$$

である. すなわち, 同時確率変数 (X, Y) が (x, y) の近傍の微小面積要素 (面積は $\Delta x \Delta y$) の中に値をとる確率は, 同時確率密度関数により $f_{XY}(x, y) \Delta x \Delta y$ で与えられる.

補題 13 (同時確率関数の性質)

$$f_{XY}(x, y) \geq 0, \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$$

注意 21 (三つ以上の同時確率変数も同様)

- 三つ以上の確率変数 $X_1, X_2, \dots, X_n$ に対しても, 同時累積分布関数 $F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$ や, 同時確率関数:

$$P_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = P(\{ \omega \in \Omega \mid X_1(\omega) = x_1, X_2(\omega) = x_2, \dots, X_n(\omega) = x_n \})$$

同時確率密度関数 $f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$ が同様に定義される. 省略して $F(x_1, x_2, \dots, x_n)$, $P(x_1, x_2, \dots, x_n)$, $f(x_1, x_2, \dots, x_n)$ のように書くことも多い.

3.9.2 同時確率関数と周辺確率関数

以降では主に離散型確率変数を扱う. 基本的な考え方は連続型確率変数でも同様であり, 和を積分に置き換えて考えればよい.

補題 14 (写像による定義域の分割) $f: \mathcal{A} \rightarrow \mathcal{B}$ を写像とし, $\text{Im } f = f(\mathcal{A})$ を f の値域とする. このとき以下が成り立つ.

- (1) 逆像の族 $\{f^{-1}(\{b\})\}_{b \in \text{Im } f}$ は互いに素

$$(2) \mathcal{A} = \bigcup_{b \in \text{Im } f} f^{-1}(\{b\})$$

すなわち，定義域 $\mathcal{A}$ は各 $b \in \text{Im } f$ の逆像により互いに素に分割される．

証明：

(1) $b_1, b_2 \in \text{Im } f, b_1 \neq b_2$ とすると補題 7 より，

$$f^{-1}(\{b_1\}) \cap f^{-1}(\{b_2\}) = f^{-1}(\{b_1\} \cap \{b_2\}) = f^{-1}(\emptyset) = \emptyset$$

となる．これは $f^{-1}(\{b_1\})$ と $f^{-1}(\{b_2\})$ が互いに素であることを示している．

(2) 同じく補題 7 より，

$$\bigcup_{b \in \text{Im } f} f^{-1}(\{b\}) = f^{-1}\left(\bigcup_{b \in \text{Im } f} \{b\}\right) = f^{-1}(\text{Im } f) = \mathcal{A}$$

□

確率変数 $X : \Omega \rightarrow \mathbb{R}$ と $Y : \Omega \rightarrow \mathbb{R}$ について，取り得る値の集合（値域）をそれぞれ

$$\mathcal{X} = \text{Im } X = X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}, \quad \mathcal{Y} = \text{Im } Y = Y(\Omega) = \{Y(\omega) \mid \omega \in \Omega\}$$

とおく．上の補題により， Ω の互いに素な分割が二通り作られる．

$$\Omega = \bigcup_{x \in \mathcal{X}} X^{-1}(\{x\}) = \bigcup_{x \in \mathcal{X}} \{\omega \in \Omega \mid X(\omega) = x\} \quad (7)$$

$$\Omega = \bigcup_{y \in \mathcal{Y}} Y^{-1}(\{y\}) = \bigcup_{y \in \mathcal{Y}} \{\omega \in \Omega \mid Y(\omega) = y\} \quad (8)$$

補題 15 (周辺分布の計算方法) 確率変数 X, Y の同時確率関数 $P_{XY}(x, y)$ と，確率関数 $P_X(x)$ ， $P_Y(y)$ について，次式が成り立つ．

$$P_X(x) = \sum_{y \in \mathcal{Y}} P_{XY}(x, y) \quad (9)$$

$$P_Y(y) = \sum_{x \in \mathcal{X}} P_{XY}(x, y) \quad (10)$$

証明：(8) と全確率の公式（定理 1）により

$$\begin{aligned} P_X(x) &= P(\{\omega \in \Omega \mid X(\omega) = x\}) = P(X^{-1}(\{x\})) = \sum_{y \in \mathcal{Y}} P(X^{-1}(\{x\}) \cap Y^{-1}(\{y\})) \\ &= \sum_{y \in \mathcal{Y}} P(\{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\}) = \sum_{y \in \mathcal{Y}} P_{XY}(x, y) \end{aligned}$$

が成り立つ．よって (9) が示された．(10) も同様である．

□

注意 22 同時確率関数 $P_{XY}(x, y)$ との対比で， $P_X(x)$ ， $P_Y(y)$ を周辺確率関数 (marginal probability function) という．それぞれ，同時分布 (joint distribution)，周辺分布 (marginal distribution) ということもある．

例 15 (同時確率関数と周辺確率関数) $\mathcal{X} = \{1, 2\}$, $\mathcal{Y} = \{1, 2, 3\}$ とする . 同時確率関数 $P_{XY}(x, y)$ が以下の表で与えられているとする . 和をとると 1 になることに注意 .

$x \setminus y$	1	2	3
1	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$
2	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$

このとき , 周辺確率関数 $P_X(x)$, $P_Y(y)$ は以下のように計算される .

$x \setminus y$	1	2	3	$P_X(x)$
1	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{9}{15}$
2	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{6}{15}$
$P_Y(y)$	$\frac{3}{15}$	$\frac{5}{15}$	$\frac{7}{15}$	

定義 20 (条件付き確率関数) $P_X(x) > 0$ のとき , 条件付き確率関数 (conditional probability function) が以下で定義される .

$$P_{Y|X}(y|x) := \frac{P_{XY}(x, y)}{P_X(x)}$$

注意 23

- $P_{Y|X}(y|x)$ は , 事象 $X = x$ が生じたもとで , 事象 $Y = y$ が起こる確率である . これは , 1.3 節で述べた事象の条件付き確率そのものである .
- 定義より次の chain rule が成り立つ .

$$P_{XY}(x, y) = P_X(x)P_{Y|X}(y|x)$$

- $P_{Y|X}(y|x)$ は各 x を固定すると確率関数である . すなわち ,

$$P_{Y|X}(y|x) \geq 0, \quad \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) = 1$$

例 16 (続き : 条件付き確率関数) 例 15 の同時確率関数 $P_{XY}(x, y)$ が与えられているとき , 条件付き確率関数 $P_{Y|X}(y|x)$ は以下のように計算される .

$\setminus y$	1	2	3
$P_{Y X}(y 1)$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{4}{9}$
$P_{Y X}(y 2)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$

レポート 8 例 15 の同時確率関数 $P_{XY}(x, y)$ が与えられているとき , 条件付き確率関数 $P_{X|Y}(x|y)$ を求めよ .

定義 21 (確率変数の独立性)

$$\begin{aligned} \text{確率変数 } X \text{ と } Y \text{ が独立} &\stackrel{\text{def}}{\iff} \forall x, \forall y, P_{XY}(x, y) = P_X(x)P_Y(y) \\ &\iff \text{すべての } x \text{ と } y \text{ について事象 } X = x \text{ と事象 } Y = y \text{ が独立} \end{aligned}$$

注意 24 (三つ以上の確率変数の独立性) 同様に三つ以上の確率変数の独立性が定義される .

確率変数 $X_1, X_2, \dots, X_n$ が独立

$$\stackrel{\text{def}}{\iff} \forall x_1, \forall x_2, \dots, \forall x_n, P_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = P_{X_1}(x_1) P_{X_2}(x_2) \cdots P_{X_n}(x_n)$$

$\iff$ すべての $x_1, x_2, \dots, x_n$ について事象 $X_1 = x_1$, 事象 $X_2 = x_2, \dots$, 事象 $X_n = x_n$ が独立

一般に三つ以上の確率変数 $X_1, X_2, \dots, X_n$ について, 二つの組 (X_i, X_j) ($1 \leq i < j \leq n$) が独立であっても, $X_1, X_2, \dots, X_n$ が独立になるとは限らないことに注意 .

3.10 確率変数の期待値と分散

定義 22 (期待値) 確率変数 X の期待値 (expectation) は次式で定義される .

$$E[X] =: \begin{cases} \sum_{x \in \mathcal{X}} P_X(x) x & (\text{離散型確率変数}) \\ \int_{-\infty}^{\infty} f_X(x) x \, dx & (\text{連続型確率変数}) \end{cases}$$

注意 25 (絶対収束) 確率変数 X の値域 $\mathcal{X} = \mathcal{X}(\Omega)$ が無限集合の場合, 期待値の定義で注意が必要である . 離散型確率変数では ($\mathcal{X}$ が加算無限集合の場合) ,

$$\sum_{x \in \mathcal{X}} |x| P_X(x) < \infty$$

が成立するとき絶対収束するという . このとき期待値は和の順序によらずに一意に収束する . 連続型確率変数では ($\mathcal{X}$ が非加算集合の場合) ,

$$\int_{-\infty}^{\infty} |x| f_X(x) \, dx < \infty$$

が成立するとき絶対収束するという . 期待値が定義できるためには, 絶対収束することが必要である .

定義 23 (分散) 確率変数 X の分散 (variance) は, $\mu = E[X]$ として, 次式で定義される .

$$V[X] := E[(X - \mu)^2]$$

定義 24 (標準偏差) 確率変数 X について, 分散の平方根 $\sigma_X = \sqrt{V[X]}$ を標準偏差 (standard deviation) とよぶ .

3.11 確率変数の関数

以下では期待値と分散の性質や計算方法を議論するため, 確率変数の関数について議論する .

定義 25 (確率変数の関数 = 確率変数) 確率変数 $X : \Omega \rightarrow \mathbb{R}$ と関数 $f : \mathbb{R} \rightarrow \mathbb{R}$ が与えられているとする . このとき合成関数 $Y : \Omega \rightarrow \mathbb{R}$ が

$$\omega \in \Omega \mapsto X(\omega) \in \mathbb{R} \xrightarrow{f} Y(\omega) = f(X(\omega)) \in \mathbb{R}$$

で定義され, Y は確率変数となる . このとき, $Y = f(X)$ と書いて, Y は確率変数 X の関数であるという .

注意 26

- X をランダムに値をとる実数だと見なしして確率変数と呼んでいた． X がランダムに値をとるとき，その結果として $Y = f(X)$ もランダムに値をとると見なす．
- 数学では一般的に，合成関数は $Y = f \circ X$ と書く．また， ω での値は $Y(\omega) = (f \circ X)(\omega) = f(X(\omega))$ のように表す．

注意 27 (同時確率変数の関数) 確率変数 X, Y と二変数関数 $f : (x, y) \in \mathbb{R}^2 \mapsto f(x, y) \in \mathbb{R}$ が与えられているとする．このとき合成関数 $Z : \Omega \rightarrow \mathbb{R}$ が

$$\omega \in \Omega \xrightarrow{(X,Y)} (X(\omega), Y(\omega)) \in \mathbb{R}^2 \xrightarrow{f} Z(\omega) = f(X(\omega), Y(\omega)) \in \mathbb{R}$$

で定義され， Z は確率変数となる．このとき， $Z = f(X, Y)$ と書いて， Z は同時確率変数 (X, Y) の関数であるという．

補題 16 (確率変数の関数：確率関数) 離散型確率変数 X と関数 $f : \mathbb{R} \rightarrow \mathbb{R}$ に対して， $Y = f(X)$ とする． X の値域を $\mathcal{X}$ ， Y の値域を $\mathcal{Y}$ とする．このとき Y の確率関数は次式で計算される．

$$P_Y(y) = \sum_{\substack{x \in \mathcal{X} \\ y=f(x)}} P_X(x) \quad (y \in \mathcal{Y})$$

証明：確率関数の定義より，

$$P_Y(y) = P(\{\omega \in \Omega \mid Y(\omega) = y\}) = P(Y^{-1}(\{y\}))$$

である．ここで逆像 $Y^{-1}(\{y\})$ について，

$$Y^{-1}(\{y\}) = X^{-1}(f^{-1}(\{y\}))$$

が成り立つ．自明に $\mathcal{X} \supset f^{-1}(\{y\}) = \bigcup_{x \in f^{-1}(\{y\})} \{x\}$ であること⁶を用いると，次式が成り立つ⁷．

$$Y^{-1}(\{y\}) = X^{-1}(f^{-1}(\{y\})) = X^{-1}\left(\bigcup_{x \in f^{-1}(\{y\})} \{x\}\right) = \bigcup_{x \in f^{-1}(\{y\})} X^{-1}(\{x\}) \quad (11)$$

ここで補題 7 を用いた．補題 14 の (1) より，部分集合の族 $\{X^{-1}(\{x\})\}_{x \in f^{-1}(\{y\})}$ は互いに素であるから，(11) は $Y^{-1}(\{y\})$ の互いに素な分割を与える．よって確率測度の公理により，

$$\begin{aligned} P_Y(y) &= P(Y^{-1}(\{y\})) = P\left(\bigcup_{x \in f^{-1}(\{y\})} X^{-1}(\{x\})\right) \\ &= \sum_{x \in f^{-1}(\{y\})} P(X^{-1}(\{x\})) \\ &= \sum_{x \in f^{-1}(\{y\})} P_X(x) = \sum_{\substack{x \in \mathcal{X} \\ y=f(x)}} P_X(x) \end{aligned}$$

□

⁶集合 A について $A = \bigcup_{a \in A} \{a\}$ という自明な式です．

⁷授業やったように，(11) 式の図を書いて復習して下さい．

同様に次の補題が成り立つ．

補題 17 (同時確率変数の関数：確率関数) 離散型確率変数 X, Y と二変数関数 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ に対して, $Z = f(X, Y)$ とする． X の値域を $\mathcal{X}$, Y の値域を $\mathcal{Y}$, Z の値域を $\mathcal{Z}$ とする．このとき Z の確率関数は次式で計算される．

$$P_Z(z) = \sum_{\substack{(x,y) \in \mathcal{X} \times \mathcal{Y} \\ z=f(x,y)}} P_{XY}(x,y) \quad (z \in \mathcal{Z})$$

証明：補題 16 で, X を (X, Y) に, Y を Z に置き換えれば良い⁸．

□

上記の補題より以下が成り立つ．

補題 18 (確率変数の和) X, Y を離散型確率変数とし $Z = X + Y$ とする．このとき Z の確率関数は次式で計算される．

$$P_Z(z) = \sum_{\substack{(x,y) \in \mathcal{X} \times \mathcal{Y} \\ z=x+y}} P_{XY}(x,y) \quad (z \in \mathcal{Z})$$

特に X と Y が独立なとき, 以下のたたみこみ (convolution) で計算される．

$$\begin{aligned} P_Z(z) &= \sum_{\substack{(x,y) \in \mathcal{X} \times \mathcal{Y} \\ z=x+y}} P_X(x)P_Y(y) \\ &= \sum_x P_X(x)P_Y(z-x) = \sum_y P_X(z-y)P_Y(y) \end{aligned}$$

レポート 9 X と Y が独立に幾何分布 $\text{Ge}(p)$ に従う確率変数であるとき, $Z = X + Y$ の確率関数を求めよ．

次に $Y = f(X)$ について期待値の計算方法を述べる．期待値の定義からすると, Y の期待値は確率関数 $P_Y(y)$ を用いて,

$$E[Y] = \sum_{y \in \mathcal{Y}} P_Y(y)y$$

で与えられる．しかし, 以下のように, Y の期待値を計算するだけならば Y の確率関数は必要がない．

補題 19 (確率変数の関数：期待値) $Y = f(X)$ のとき

$$E[Y] = \sum_{x \in \mathcal{X}} P_X(x)f(x)$$

証明：関数 $f: \mathbb{R} \rightarrow \mathbb{R}$ について, X の値域 $\mathcal{X}$ に f の定義域を制限した関数 $f: \mathcal{X} \rightarrow \mathcal{Y}$ を考える．補題 14 より, $y \in \mathcal{Y}$ の逆像 $f^{-1}(\{y\}) = \{x \in \mathcal{X} \mid f(x) = y\}$ は $\mathcal{X}$ の分割を与える．すなわち, $\mathcal{X}$ の部分集合の族 $\{f^{-1}(\{y\})\}_{y \in \mathcal{Y}}$ は互いに素で,

$$\mathcal{X} = \bigcup_{y \in \mathcal{Y}} f^{-1}(\{y\})$$

⁸補題 16 は X や Y がベクトルの場合でも, 関数 f を適切に設定すれば成立する．証明もほとんどそのままである．

が成り立つ⁹ . よって ,

$$\begin{aligned}\sum_{x \in \mathcal{X}} P_X(x) f(x) &= \sum_{y \in \mathcal{Y}} \sum_{x \in f^{-1}(\{y\})} P_X(x) f(x) = \sum_{y \in \mathcal{Y}} \sum_{x: y=f(x)} P_X(x) f(x) \\ &= \sum_{y \in \mathcal{Y}} y \sum_{x: y=f(x)} P_X(x) = \sum_{y \in \mathcal{Y}} y P_Y(y) = E[Y]\end{aligned}$$

□

同様に次の補題が成り立つ .

補題 20 (同時確率変数の関数 : 期待値) $Z = f(X, Y)$ のとき

$$E[Z] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y) f(x, y)$$

証明 : 補題 19 で , X を (X, Y) に , Y を Z に置き換えれば良い .

□

3.12 期待値と分散の性質

補題 21 (期待値の性質)

(1) 線形性 : $E[aX + bY] = aE[X] + bE[Y]$ (X と Y が独立であるとは限らない)

(2) X と Y が独立のとき , $E[XY] = E[X]E[Y]$

証明 :

(1) $aX + bY$ は (X, Y) の関数 $f(X, Y) = aX + bY$ と見なせるので , 上の補題により

$$\begin{aligned}E[aX + bY] &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y)(ax + by) \\ &= a \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y)x + b \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y)y \\ &= a \sum_{x \in \mathcal{X}} \left\{ \sum_{y \in \mathcal{Y}} P_{XY}(x, y) \right\} x + b \sum_{y \in \mathcal{Y}} \left\{ \sum_{x \in \mathcal{X}} P_{XY}(x, y) \right\} y \\ &= a \sum_{x \in \mathcal{X}} P_X(x)x + b \sum_{y \in \mathcal{Y}} P_Y(y)y \\ &= aE[X] + bE[Y]\end{aligned}$$

(2) X, Y が独立のとき , 独立性の定義より $P_{XY}(x, y) = P_X(x)P_Y(y)$ であるから ,

$$\begin{aligned}E[XY] &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y)xy \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_X(x)P_Y(y)xy \\ &= \left(\sum_{x \in \mathcal{X}} P_X(x)x \right) \left(\sum_{y \in \mathcal{Y}} P_Y(y)y \right) \\ &= E[X]E[Y]\end{aligned}$$

⁹授業やったように , 図を書いて復習して下さい .

□

定義 26 (共分散) 同時確率変数 X, Y に対して共分散 (covariance) を次式で定義する .

$$\text{Cov}(X, Y) := E[(X - \mu_X)(Y - \mu_Y)]$$

ただし , $\mu_X = E[X], \mu_Y = E[Y]$ とおいた .

補題 22 (分散の性質)

$$(1) V[X] \geq 0$$

$$(2) V[aX + b] = a^2 V[X] \quad (a, b \in \mathbb{R})$$

$$(3) V[X] = E[X^2] - E[X]^2$$

$$(4) V[X + Y] = V[X] + 2 \text{Cov}(X, Y) + V[Y]$$

$$(5) X \text{ と } Y \text{ が独立ならば } \text{Cov}(X, Y) = 0 \text{ となり } V[X + Y] = V[X] + V[Y]$$

証明 :

$$(1) \mu = E[X] \text{ とおくと ,}$$

$$V[X] = E[(X - \mu)^2] = \sum_{x \in \mathcal{X}} P_X(x)(x - \mu)^2 \geq 0$$

$$(2) E[aX + b] = aE[X] + b = a\mu + b \text{ であるから ,}$$

$$V[aX + b] = E \left[\{ (aX + b) - (a\mu + b) \}^2 \right] = E[a^2(X - \mu)^2] = a^2 E[(X - \mu)^2] = a^2 V[X]$$

$$(3)$$

$$\begin{aligned} V[X] &= E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu \cdot E[X] + \mu^2 = E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2 \end{aligned}$$

$$(4) \mu_X = E[X], \mu_Y = E[Y] \text{ とおくと , } E[X + Y] = \mu_X + \mu_Y \text{ であるから ,}$$

$$\begin{aligned} V[X + Y] &= E \left[\{ (X + Y) - (\mu_X + \mu_Y) \}^2 \right] \\ &= E \left[\{ (X - \mu_X) + (Y - \mu_Y) \}^2 \right] \\ &= E \left[(X - \mu_X)^2 + 2(X - \mu_X)(Y - \mu_Y) + (Y - \mu_Y)^2 \right] \\ &= E[(X - \mu_X)^2] + 2E[(X - \mu_X)(Y - \mu_Y)] + E[(Y - \mu_Y)^2] \\ &= V[X] + 2 \text{Cov}(X, Y) + V[Y] \end{aligned}$$

(5)

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y)(x - \mu_X)(y - \mu_Y) \\&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_X(x)P_Y(y)(x - \mu_X)(y - \mu_Y) \\&= \sum_{x \in \mathcal{X}} P_X(x)(x - \mu_X) \sum_{y \in \mathcal{Y}} P_Y(y)(y - \mu_Y) \\&= E[X - \mu_X] \cdot E[Y - \mu_Y] \\&= 0\end{aligned}$$

よって (4) より $V[X + Y] = V[X] + 2\text{Cov}(X, Y) + V[Y] = V[X] + V[Y]$.

□

注意 28 (読み飛ばしても良い) X と Y が独立ならば, $X' = f(X)$ と $Y' = g(Y)$ は独立である .
これは次のように確認できる .

$$\begin{aligned}P_{X'Y'}(x', y') &= \sum_{\substack{(x, y) \in \mathcal{X} \times \mathcal{Y} \\ (x', y') = (f(x), g(y))}} P_{XY}(x, y) \\&= \sum_{\substack{x \in \mathcal{X} \\ f(x) = x'}} \sum_{\substack{y \in \mathcal{Y} \\ f(y) = y'}} P_X(x)P_Y(y) \\&= \sum_{\substack{x \in \mathcal{X} \\ f(x) = x'}} P_X(x) \sum_{\substack{y \in \mathcal{Y} \\ f(y) = y'}} P_Y(y) \\&= P_{X'}(x')P_{Y'}(y')\end{aligned}$$

よって, 前補題 (5) は, 補題 21 の (2) を用いてもっと簡単に示される .

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[X - \mu_X] \cdot E[Y - \mu_Y] = 0$$

3.13 代表的な確率変数の期待値 (離散型)

3.13.1 二点分布 $B(1, p)$ ($0 < p < 1$)

二点分布 $B(1, p)$ に従う確率変数 X (ベルヌイ確率変数) について, 期待値と分散は定義から以下のように計算される .

$$\begin{aligned}E[X] &= p \times 1 + (1 - p) \times 0 = p \\V[X] &= p \times (1 - p)^2 + (1 - p) \times (0 - p)^2 = p(1 - p)\{(1 - p) + p\} = p(1 - p)\end{aligned}$$

3.13.2 二項分布 $B(n, p)$ ($n \in \mathbb{N}, 0 < p < 1$)

二項分布 $B(n, p)$ に従う確率変数 Y_n は、表の出る確率が p であるコインを n 回投げた時の表の回数 Y_n であった。 $X_1, X_2, \dots, X_n$ を独立なベルヌイ確率変数とすると、 $Y_n = \sum_{i=1}^n X_i$ であるから、

$$\begin{aligned} E[Y_n] &= E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n p = np \\ V[Y_n] &= V\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n V[X_i] = \sum_{i=1}^n p(1-p) = np(1-p) \end{aligned}$$

3.13.3 幾何分布 $Ge(p)$ ($0 < p < 1$)

幾何分布 $Ge(p)$ に従う確率変数 X について、期待値と分散は以下で与えられる。

$$E[X] = \frac{1-p}{p}, \quad V[X] = \frac{1-p}{p^2}$$

$E[X]$ は次のように計算できる。 $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ を x で微分すると、

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

ここで $k-1=l$ とおくと、

$$\frac{1}{(1-x)^2} = \sum_{l=0}^{\infty} (l+1)x^l = \sum_{l=0}^{\infty} lx^l + \sum_{l=0}^{\infty} x^l = \sum_{l=0}^{\infty} lx^l + \frac{1}{1-x}$$

よって、

$$\sum_{l=0}^{\infty} lx^l = \frac{1}{(1-x)^2} - \frac{1}{1-x}$$

これを $x = 1-p$ として用いることで、

$$\begin{aligned} E[X] &= \sum_{k=0}^{\infty} k \cdot p(1-p)^k = p \sum_{k=0}^{\infty} k(1-p)^k \\ &= p \left(\frac{1}{p^2} - \frac{1}{p} \right) = \left(\frac{1}{p} - 1 \right) = \frac{1-p}{p} \end{aligned}$$

分散 $V[X]$ も同様に計算できる。後のモーメント母関数の節で、別の計算方法を与える。

3.13.4 ポアソン分布 $Po(\lambda)$ ($\lambda > 0$)

レポート 10 (ポアソン分布の期待値と分散) 確率変数 X はポアソン分布 $Po(\lambda)$ に従うとする。

(1) 期待値 $E[X]$ を求めよ。

(2) $E[X(X-1)]$ を求めよ。

(3) 分散 $V[X]$ を求めよ .

解答 :

(1)

$$E[X] = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!} = \sum_{l=0}^{\infty} e^{-\lambda} \frac{\lambda^{l+1}}{l!} = \lambda \sum_{l=0}^{\infty} e^{-\lambda} \frac{\lambda^l}{l!} = \lambda$$

ただし , 途中で $k-1=l$ とおき , $\sum_{l=0}^{\infty} e^{-\lambda} \frac{\lambda^l}{l!} = 1$ を用いた .

(2)

$$E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) \cdot e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=2}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-2)!} = \sum_{l=0}^{\infty} e^{-\lambda} \frac{\lambda^{l+2}}{l!} = \lambda^2 \sum_{l=0}^{\infty} e^{-\lambda} \frac{\lambda^l}{l!} = \lambda^2$$

(3) $\lambda^2 = E[X(X-1)] = E[X^2 - X] = E[X^2] - E[X] = E[X^2] - \lambda$ より $E[X^2] = \lambda^2 + \lambda$ である .
よって ,

$$V[X] = E[X^2] - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

3.14 代表的な確率変数の期待値 (連続型)

3.14.1 一様分布 $U(a, b)$ ($a < b$)

一様分布 $U(a, b)$ に従う確率変数 X について , 期待値と分散は定義から以下のように計算される .

$$E[X] = \int_{-\infty}^{\infty} f(x)x \, dx = \int_a^b \frac{1}{b-a} \cdot x \, dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{a+b}{2}$$

$$\begin{aligned} V[X] &= \int_{-\infty}^{\infty} f(x)(x-\mu)^2 \, dx = \int_a^b \frac{1}{b-a} \cdot (x-\mu)^2 \, dx \\ &= \frac{1}{b-a} \left[\frac{(x-\mu)^3}{3} \right]_a^b = \frac{1}{3(b-a)} \left\{ \left(b - \frac{a+b}{2} \right)^3 - \left(a - \frac{a+b}{2} \right)^3 \right\} \\ &= \frac{1}{3(b-a)} \left\{ \frac{(b-a)^3}{8} - \frac{(a-b)^3}{8} \right\} = \frac{1}{3(b-a)} \cdot \frac{(b-a)^3}{4} = \frac{(b-a)^2}{12} \end{aligned}$$

3.14.2 指数分布 $Ex(\alpha)$ ($\alpha > 0$)

レポート 11 (指数分布の期待値と分散) 確率変数 X は指数分布 $Ex(\alpha)$ に従うとする .

(1) 期待値 $E[X]$ を求めよ .

(2) 分散 $V[X]$ を求めよ .

ヒント : 部分積分 . 答えは

$$E[X] = \frac{1}{\alpha}, \quad V[X] = \frac{1}{\alpha^2}$$

補題 23 (指数分布の無記憶性) X を指数分布に従う確率変数とすると, $x, t > 0$ に対して,

$$P(X > s + t) = P(X > s)P(X > t)$$

証明: $F(s)$ を X の累積分布関数とすると,

$$F(s) = \int_0^s \alpha e^{-\alpha x} \cdot x \, dx = [-e^{-\alpha x}]_0^s = 1 - e^{-\alpha s}$$

よって

$$P(X > s) = 1 - P(X \leq s) = 1 - F(s) = e^{-\alpha s}$$

この式を用いると,

$$P(X > s)P(X > t) = e^{-\alpha s}e^{-\alpha t} = e^{-\alpha(s+t)} = P(X > s + t)$$

□

例 17 ある蛍光灯の寿命 X は平均 $1/\alpha$ 年の指数分布に従うとする. $0 < s < t$ として, 蛍光灯を s ヶ月以上使用した条件のもとで, さらに t ヶ月以上もつ確率は?

$$\begin{aligned} P(X - s > t | X > s) &= P(X > s + t | X > s) = \frac{P(X > s + t, X > s)}{P(X > s)} \\ &= \frac{P(X > s + t)}{P(X > s)} = \frac{P(X > s)P(X > t)}{P(X > s)} = P(X > t) = e^{-\alpha t} \end{aligned}$$

$X > s$ の条件をつけても, 形が同じことがわかる. この性質を指数分布の無記憶性という.

3.14.3 正規分布 (normal distribution) $N(\mu, \sigma)$ ($\mu \in \mathbb{R}, \sigma > 0$)

正規分布 $N(\mu, \sigma)$ に従う確率変数 X について, $E[X] = \mu$, $V[X] = \sigma^2$ を以下の手順で示す.

(a) Y が $N(0, 1)$ (標準正規分布) に従うとき

$$E[Y] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{y^2}{2}} dy = 0$$

(b) 上式を y で微分することで, $V[Y] = 1$ となる.

(c) X が $N(\mu, \sigma)$ に従うとき, $Y = \frac{X - \mu}{\sigma}$ は標準正規分布 $N(0, 1)$ に従う (スケール変換, 規格化).

(d) 平均と分散の性質を用いて $E[X] = \mu$, $V[X] = \sigma^2$ を示す.

解答:

(a) 奇関数だから.

(b) (a) の左辺を微分すると,

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\partial}{\partial y} \left(y e^{-\frac{y^2}{2}} \right) dy &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}} dy \\ &= 1 - V[Y] \end{aligned}$$

となる. これは右辺の定数 0 の微分に等しい. すなわち $1 - V[Y] = 0$. よって $V[Y] = 1$ となる.

(c)

$$\begin{aligned} F_Y(y) &= \Pr\{Y \leq y\} = \Pr\left\{\frac{X - \mu}{\sigma} \leq y\right\} = \Pr\{X \leq \sigma y + \mu\} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\sigma y + \mu} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad [\because \text{分布関数の定義}] \end{aligned}$$

ここで $t = \frac{x-\mu}{\sigma}$ において積分の変数変換を行うと,

$$F_Y(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{t^2}{2}} dt$$

となる. これは Y が $N(0, 1)$ に従うことを示す.

(d) $X = \sigma Y + \mu$ であるから,

$$\begin{aligned} E[X] &= E[\sigma Y + \mu] = \sigma E[Y] + \mu = \mu \quad [\because E[Y] = 0] \\ V[X] &= V[\sigma Y + \mu] = \sigma^2 V[Y] = \sigma^2 \quad [\because V[Y] = 1] \end{aligned}$$

4 大数の法則

4.1 確率に関する不等式

補題 24 (Cauchy-Schwarz の不等式)

$$E[XY]^2 \leq E[X^2]E[Y^2]$$

証明: $Z = tX + Y$ とおくと,

$$Z^2 = (tX + Y)^2 = t^2 X^2 + 2tXY + Y^2 \geq 0$$

である. ここで期待値をとると,

$$E[Z^2] = E[X^2]t^2 + 2E[XY]t + E[Y^2] \geq 0$$

上式は任意の $t \in \mathbb{R}$ について成立するので, t の二次式として判別式が負になる.

$$E[XY]^2 - E[X^2]E[Y^2] \leq 0$$

よって補題が証明された. □

系 1 (共分散と分散の関係)

$$\text{Cov}(X, Y) \leq \sqrt{V[X]V[Y]}$$

証明: $X' = X - E[X]$, $Y' = Y - E[Y]$ として, X' , Y' に Cauchy-Schwarz の不等式を適用すると,

$$\text{Cov}(X, Y) = E[X'Y'] \leq \sqrt{E[X'^2]E[Y'^2]} = \sqrt{V[X]V[Y]}$$

□

補題 25 (Markov の不等式) Z を非負の値をとる確率変数とするととき ,

$$\forall a > 0, \quad \Pr \{Z > a\} \leq \frac{E[Z]}{a}$$

証明 : 離散型確率変数の場合は以下のように示される .

$$E[Z] = \sum_z z P_Z(z) \geq \sum_{z: z>a} z P_Z(z) \geq \sum_{z: z>a} a P_Z(z) = a \sum_{z: z>a} P_Z(z) = a \Pr \{Z > a\}$$

連続型確率変数の場合は , 和を積分に置き換えることで同様に示される . □

補題 26 (Chebyshev の不等式) 期待値と分散を持つ任意の確率変数 X に対して ,

$$\forall t > 0, \quad \Pr \{|X - E[X]| > t\} \leq \frac{V[X]}{t^2}$$

証明 : $|X - E[X]| > t \iff (X - E[X])^2 > t^2$ であるから ,

$$\begin{aligned} \Pr \{|X - E[X]| > t\} &= \Pr \{(X - E[X])^2 > t^2\} \\ &\leq \frac{E[(X - E[X])^2]}{t^2} \quad [\because \text{Markov の不等式}] \\ &= \frac{V[X]}{t^2} \quad [\because \text{分散の定義}] \end{aligned}$$

ただし , Markov の不等式を $Z = (X - E[X])^2$, $a = t^2$ として適用した . □

4.2 大数の法則

定義 27 (算術平均) 確率変数の列 $X_1, X_2, \dots, X_n, \dots$ が与えられたとき ,

$$Y_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (n = 1, 2, \dots)$$

を算術平均という .

注意 29

- 算術平均と期待値を混同してはいけない .
- 算術平均は確率変数であることに注意 .

定義 28 (独立同一分布, independently and identically distributed) n 個の確率変数

$$X_1, X_2, \dots, X_n$$

が独立で , これらの周辺確率関数が同一なとき , すなわち , 同時確率関数が

$$P_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = P(x_1)P(x_2) \dots P(x_n)$$

と書けるとき , $X_1, X_2, \dots, X_n$ は独立同一分布に従うという . これを

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P$$

と書くこともある . 連続型の場合も同様に , 確率密度関数を用いて定義される .

定理 5 (大数の弱法則, weak law of large numbers) n 個の確率変数 $X_1, X_2, \dots, X_n$ が独立同一分布に従うとき, $E[X_1^2] < \infty$ であれば, 以下が成り立つ.

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \Pr \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - E[X_1] \right| > \varepsilon \right\} = 0 \quad (12)$$

注意 30 (12) は以下と同値である.

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \Pr \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - E[X_1] \right| \leq \varepsilon \right\} = \lim_{n \rightarrow \infty} \left[1 - \Pr \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - E[X_1] \right| > \varepsilon \right\} \right] = 1$$

どんな小さな $\varepsilon > 0$ についても, 算術平均と期待値の差が ε 以内である確率は 1 に近づく. このとき, Y_n は $E[X_1]$ に “確率収束する” という.

証明: $X_1, X_2, \dots, X_n$ の期待値は等しいので, $\mu = E[X_1] = E[X_2] = \dots = E[X_n]$ とおく. 同様に分散についても $\sigma^2 = V[X_1] = V[X_2] = \dots = V[X_n]$ とおく. ここで算術平均を $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$ とおくと,

$$\begin{aligned} E[Y_n] &= E \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu \\ V[Y_n] &= V \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n V[X_i] = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

である. よって, Chebyshev の不等式より,

$$\Pr \{ |Y_n - \mu| > \varepsilon \} \leq \frac{V[Y_n]}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

□

注意 31 条件 $E[X_1^2] < \infty$ は, 期待値や分散が有限の値となることを保証している. 実際, Cauchy-Schwarz の不等式を用いると,

$$E[|X_1|]^2 = E[1 \cdot |X_1|]^2 \leq E[1^2] E[X_1^2] < \infty$$

であるから $E[|X_1|] < \infty$. すなわち, 期待値 $E[X_1]$ は有限の値に絶対収束する. このとき,

$$V[X_1]^2 = E[X_1^2] - E[X_1]^2 < \infty$$

となって分散も有限の値になる.

定理 6 (大数の強法則) 確率変数 $X_1, X_2, \dots, X_n, \dots$ が独立同一分布に従うとき, $E[X_1^4] < \infty$ であれば, 以下が成り立つ.

$$\Pr \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = E[X_1] \right\} = 1 \quad (13)$$

注意 32 確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数は標本空間 Ω 上の関数であった。(13) を省略しないで書くと以下の通りである。

$$P \left(\left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i(\omega) = E[X_1] \right\} \right) = 1$$

このとき, Y_n は $E[X_1]$ に“概収束する”という。大数の強法則を, 証明まで正確に理解するには, 事象の極限操作について学ぶ必要がある。

注意 33 (大数の法則と中心極限定理) 独立同一分布に従う確率変数列 $X_1, X_2, \dots, X_n$ の算術平均 $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$ について, $\mu = E[X_1]$, $\sigma^2 = V[X_1]$ とおくと, $E[Y_n] = \mu$, $V[Y_n] = \sigma^2/n$ であることに注意。

- 大数の法則: Y_n は期待値に確率収束 (概収束) する。
- 中心極限定理: Y_n を平均 0 分散 1 に規格化した確率変数

$$\frac{Y_n - \mu}{\sqrt{\sigma^2/n}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$$

は正規分布 $N(0, 1)$ に近づく。

5 モーメント母関数

定義 29 (モーメント母関数) 実確率変数 X に対して

$$M_X(t) = E[e^{tX}]$$

を X のモーメント母関数 (moment generating function) または積率母関数という。

注意 34 t は期待値が収束する適当な範囲で考える。 $M_X(0) = 1$ だから $t = 0$ の周りでは収束する。

定義 30 (n 次モーメント) 確率変数 X に対して $E[X^n]$ を X の n 次モーメントという。

補題 27 (モーメント母関数と n 次モーメントの関係)

$$E[X^n] = M_X^{(n)}(0) := \left. \frac{d^n}{dt^n} M_X(t) \right|_{t=0}$$

証明: 指数関数 e^x の $x = 0$ での Taylor 展開

$$e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \dots$$

より

$$\begin{aligned} M_X(t) &= E[e^{tX}] = E \left[1 + tX + \frac{(tX)^2}{2!} + \dots + \frac{(tX)^n}{n!} + \dots \right] \\ &= 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \dots + \frac{t^n}{n!}E[X^n] + \dots \end{aligned}$$

一方, $M_X(t)$ の $t = 0$ での Taylor 展開は

$$M_X(t) = 1 + M_X'(0)t + \frac{M_X^{(2)}(0)}{2!}t^2 + \dots + \frac{M_X^{(n)}(0)}{n!}t^n + \dots$$

これらを比較すれば補題が示される。

□

系 2 (期待値と分散の計算方法)

$$E[X] = M'_X(0), \quad V[X] = E[X^2] - E[X]^2 = M''_X(0) - M'_X(0)^2$$

注意 35 (特性関数) モーメント母関数 $M_X(t) = E[e^{tX}]$ において, t を it に置き換えた関数は特性関数 (characteristic function) と呼ばれる. 連続型の場合 $f_X(x)$ を確率密度関数として,

$$\phi_X(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx$$

である. ここで

$$\phi_X(-t) = \int_{-\infty}^{\infty} e^{-itx} f_X(x) dx$$

は $f_X(x)$ の Fourier 変換に他ならない. 関数とその Fourier 変換は 1 対 1 に対応していたので, 確率密度関数と特性関数は 1 対 1 に対応している.

例 18 (二点分布) $P_X(1) = p, P_X(0) = 1 - p$ のとき

$$M_X(t) = e^{t \cdot 1} p + e^{t \cdot 0} (1 - p) = 1 + p(e^t - 1)$$

$$M'_X(t) = pe^t$$

$$M''_X(t) = pe^t$$

であるから,

$$E[X] = M'_X(0) = p$$

$$V[X] = M''_X(0) - M'_X(0)^2 = p - p^2 = p(1 - p)$$

レポート 12 (二項分布) Y_n を二項分布 $B(n, p)$ に従う確率変数とする.

(1) Y_n のモーメント母関数 $M_{Y_n}(t)$ を求めよ.

(2) $M'_{Y_n}(t), M''_{Y_n}(t)$ を求めよ.

(3) (2) を利用して Y_n の期待値と分散を求めよ.

解答:

(1) Y_n の確率関数は

$$P_{Y_n}(k) = \binom{n}{k} p^k (1 - p)^{(n-k)} \quad (k = 0, 1, \dots, n)$$

であるから, 二項定理を用いると,

$$\begin{aligned} M_{Y_n}(t) &= E[e^{tY_n}] = \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1 - p)^{(n-k)} = \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1 - p)^{(n-k)} \\ &= \{pe^t + (1 - p)\}^n = \{1 + p(e^t - 1)\}^n \end{aligned}$$

(2)

$$M'_{Y_n}(t) = n \{1 + p(e^t - 1)\}^{n-1} \cdot pe^t$$

$$M''_{Y_n}(t) = n(n-1) \{1 + p(e^t - 1)\}^{n-2} \cdot (pe^t)^2 + n \{1 + p(e^t - 1)\}^{n-1} \cdot pe^t$$

(3)

$$E[X] = M'_X(0) = np$$

$$V[X] = M''_X(0) - M'_X(0)^2 = \{n(n-1)p^2 + np\} - (np)^2 = -np^2 + np = np(1-p)$$

例 19 (幾何分布) X を幾何分布 $\text{Ge}(p)$ に従う確率変数とする. $q = 1 - p$ とおくと X の確率関数は $P_X(k) = pq^k$ ($k = 0, 1, 2, \dots$) である.

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} P_X(k) = \sum_{k=0}^{\infty} e^{tk} pq^k = p \sum_{k=0}^{\infty} (qe^t)^k = \frac{p}{1 - qe^t}$$

$$M'_X(t) = -\frac{p}{(1 - qe^t)^2} \cdot (-qe^t) = \frac{pqe^t}{(1 - qe^t)^2}$$

$$M''_X(t) = -\frac{2pqe^t}{(1 - qe^t)^3} \cdot (-qe^t) + \frac{pqe^t}{(1 - qe^t)^2} = \frac{2pq^2e^{2t}}{(1 - qe^t)^3} + \frac{pqe^t}{(1 - qe^t)^2}$$

であるから,

$$E[X] = M'_X(0) = \frac{pq}{(1 - q)^2} = \frac{1 - p}{p}$$

$$V[X] = M''_X(0) - M'_X(0)^2 = \frac{2pq^2}{(1 - q)^3} + \frac{pq}{(1 - q)^2} - \left(\frac{1 - p}{p}\right)^2$$

$$= \frac{2q^2}{p^2} + \frac{pq}{p^2} - \frac{q^2}{p^2} = \frac{q^2}{p^2} + \frac{pq}{p^2} = \frac{(q + p)q}{p^2} = \frac{1 - p}{p^2}$$

例 20 (ポアソン分布) X をポアソン分布 $\text{Po}(\lambda)$ に従う確率変数とする. X の確率関数は $P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ ($k = 0, 1, 2, \dots$) である.

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

$$M'_X(t) = \lambda e^t e^{\lambda(e^t - 1)}$$

$$M''_X(t) = \lambda e^t e^{\lambda(e^t - 1)} + (\lambda e^t)^2 e^{\lambda(e^t - 1)}$$

であるから,

$$E[X] = M'_X(0) = \lambda$$

$$V[X] = M''_X(0) - M'_X(0)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$$

例 21 (指数分布) X を指数分布 $\text{Ex}(\alpha)$ ($\alpha > 0$) に従う確率変数とする. X の確率密度関数は

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

であるから ,

$$\begin{aligned} M_X(t) &= E[e^{tX}] = \int_0^\infty e^{tx} \cdot \alpha e^{-\alpha x} dx = \int_0^\infty \alpha e^{(t-\alpha)x} dx \\ &= \left[\frac{\alpha}{t-\alpha} e^{(t-\alpha)x} \right]_0^\infty = \begin{cases} -\frac{\alpha}{t-\alpha} & (t-\alpha < 0) \\ \infty & (t-\alpha \geq 0) \end{cases} \end{aligned}$$

$t < \alpha$ のとき ,

$$M'_X(t) = \frac{\alpha}{(t-\alpha)^2}, \quad M''_X(t) = -\frac{2\alpha}{(t-\alpha)^3}$$

となる . よって ,

$$\begin{aligned} E[X] &= M'_X(0) = \frac{1}{\alpha} \\ V[X] &= M''_X(0) - M'_X(0)^2 = \frac{2\alpha}{\alpha^3} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2} \end{aligned}$$

補題 28 確率変数 X のモーメント母関数を $M_X(t)$ とするとき , $Y = aX + b$ ($a, b \in \mathbb{R}$) のモーメント母関数は $M_Y(t) = e^{bt} M_X(at)$ となる .

証明 : $M_Y(t) = E[e^{tY}] = E[e^{t(aX+b)}] = E[e^{(at)X} e^{bt}] = e^{bt} M_X(at)$ □

例 22 (正規分布) 正規分布 $N(\mu, \sigma)$ に従う確率変数 X のモーメント母関数を求める . $Y = \frac{X-\mu}{\sigma}$ とおくと Y は正規分布 $N(0, 1)$ に従う . Y のモーメント母関数を求める .

$$M_Y(t) = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} \cdot e^{\frac{t^2}{2}} dx = e^{\frac{t^2}{2}}$$

$X = \sigma Y + \mu$ であるから ,

$$M_X(t) = E[e^{t(\sigma Y + \mu)}] = e^{\mu t} M_Y(\sigma t) = e^{\mu t} e^{\frac{(\sigma t)^2}{2}} = e^{\frac{\sigma^2}{2} t^2 + \mu t}$$

これを用いて X の平均と分散を求める .

$$\begin{aligned} M'_X(t) &= (\sigma^2 t + \mu) e^{\frac{\sigma^2}{2} t^2 + \mu t} \\ M''_X(t) &= \sigma^2 e^{\frac{\sigma^2}{2} t^2 + \mu t} + (\sigma^2 t + \mu)^2 e^{\frac{\sigma^2}{2} t^2 + \mu t} \end{aligned}$$

これより ,

$$\begin{aligned} E[X] &= M'_X(0) = \mu \\ V[X] &= M''_X(0) - M'_X(0)^2 = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2 \end{aligned}$$

定理 7 (独立な確率変数の和のモーメント母関数) 独立な確率変数 X, Y の和 $Z = X + Y$ のモーメント母関数は ,

$$M_Z(t) = M_X(t) M_Y(t)$$

で与えられる .

証明： 離散型確率変数について示す．独立な確率変数の和 $Z = X + Y$ の確率関数は

$$P_Z(z) = \sum_{(x,y): z=x+y} P_X(x)P_Y(y)$$

であるから，

$$\begin{aligned} M_Z(t) &= E[e^{tZ}] = \sum_z e^{tz} \sum_{(x,y): z=x+y} P_X(x)P_Y(y) \\ &= \sum_z \sum_{(x,y): z=x+y} e^{tz} P_X(x)P_Y(y) = \sum_z \sum_{(x,y): z=x+y} e^{t(x+y)} P_X(x)P_Y(y) \\ &= \sum_x \sum_y e^{tx} e^{ty} P_X(x)P_Y(y) = \left(\sum_x e^{tx} P_X(x) \right) \left(\sum_y e^{ty} P_Y(y) \right) = M_X(t)M_Y(t) \end{aligned}$$

連続型確率変数でも同様に示される． □

系 3 $X_1, X_2, \dots, X_n$ が独立であるとき， $Y_n = X_1 + X_2 + \dots + X_n$ のモーメント母関数は

$$M_{Y_n}(t) = M_{X_1}(t)M_{X_2}(t) \cdots M_{X_n}(t)$$

で与えられる．

証明： $X_1, X_2, \dots, X_n$ が独立ならば， $X_1 + X_2 + \dots + X_{n-1}$ と X_n は独立である．よって前定理を用いると，

$$M_{X_1+X_2+\dots+X_n}(t) = M_{X_1+X_2+\dots+X_{n-1}}(t)M_{X_n}(t)$$

が成り立つ．帰納法を用いれば系が示される． □

例 23 (二項分布) Y_n を二項分布 $B(n, p)$ に従う確率変数とすると，二点分布 $B(1, p)$ に従う独立な確率変数 $X_1, X_2, \dots, X_n$ を用いて， $Y_n = X_1 + X_2 + \dots + X_n$ と表される．二点分布のモーメント母関数は

$$M_{X_1}(t) = M_{X_2}(t) = \dots = M_{X_n}(t) = 1 + p(e^t - 1)$$

であったから，

$$M_{Y_n}(t) = M_{X_1}(t)M_{X_2}(t) \cdots M_{X_n}(t) = \{1 + p(e^t - 1)\}^n$$

6 事象の極限 (入門)

定義 31 (単調列) 確率空間 $(\Omega, P, \mathcal{F})$ 上の事象列 $\{A_n\}_{n=1}^{\infty}$ ($A_n \in \mathcal{F}$) について

$$A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$$

が成り立つとき， $\{A_n\}_{n=1}^{\infty}$ は単調増加列であるという．また，

$$A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$$

が成り立つとき， $\{A_n\}_{n=1}^{\infty}$ は単調減少列であるという．単調増加列または単調減少列であるとき， $\{A_n\}_{n=1}^{\infty}$ は単調列であるという．

定義 32 (単調列の極限)

$$\lim_{n \rightarrow \infty} A_n := \bigcup_{n=1}^{\infty} A_n \quad (\text{単調増加列}), \quad \lim_{n \rightarrow \infty} A_n := \bigcap_{n=1}^{\infty} A_n \quad (\text{単調減少列})$$

注意 36 (復習)

$$\bigcup_{n=1}^{\infty} A_n = \{\omega \in \Omega \mid \exists n \in \mathbb{N}, \omega \in A_n\}, \quad \bigcap_{n=1}^{\infty} A_n = \{\omega \in \Omega \mid \forall n \in \mathbb{N}, \omega \in A_n\}$$

例 24 (区間の極限)

• 単調増加列の極限

$$\bigcup_{n=1}^{\infty} \left[a + \frac{1}{n}, b \right] = (a, b], \quad \bigcup_{n=1}^{\infty} \left[a, b - \frac{1}{n} \right] = [a, b), \quad \bigcup_{n=1}^{\infty} \left[a + \frac{1}{n}, b - \frac{1}{n} \right] = (a, b)$$

• 単調減少列の極限

$$\bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b \right) = [a, b), \quad \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n} \right) = (a, b], \quad \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n} \right) = [a, b],$$

定理 8 (確率測度の連続性) 事象列 $\{A_n\}_{n=1}^{\infty}$ が単調列のとき

$$P\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$$

証明: $\{A_n\}_{n=1}^{\infty}$ が単調増加列のとき,

$$B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus (A_1 \cup A_2), \dots, B_n = A_n \setminus \left(\bigcup_{k=1}^{n-1} A_k\right), \dots$$

とおくと, $\{B_n\}_{n=1}^{\infty}$ は互いに素で,

$$\bigcup_{k=1}^n B_k = A_n = \bigcup_{k=1}^n A_k \quad (14)$$

が成り立つ. よって,

$$\bigcup_{k=1}^{\infty} B_k = \lim_{n \rightarrow \infty} A_n = \bigcup_{k=1}^{\infty} A_k$$

であるから,

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} A_n\right) &= P\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} P(B_k) \quad [\because \sigma\text{-加法性}] \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n P(B_k) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=1}^n B_k\right) \quad [\because \text{有限加法性}] \\ &= \lim_{n \rightarrow \infty} P(A_n) \quad [\because (14)] \end{aligned}$$

$\{A_n\}_{n=1}^{\infty}$ が単調減少列のときも同様に示される.

□

注意 37 (確率変数の条件について) (読み飛ばしても良い) $(\Omega, \mathcal{F}, P)$ を確率空間とするとき, 関数

$$X : \omega \in \Omega \mapsto X(\omega) \in \mathbb{R}$$

が確率変数であるためには, 本当は以下の条件を要請する必要があった.

$$\forall x \in \mathbb{R}, \quad \{\omega \in \Omega \mid X(\omega) \leq x\} = X^{-1}((-\infty, x]) \in \mathcal{F} \quad (15)$$

例えば, 集合族 $(-\infty, x]$ ($x \in \mathbb{R}$) から

$$(a, b] = (-\infty, a]^c \cap (-\infty, b], \quad [a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right], \quad (a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right]$$

のように区間が生成されるので, (15) が満たされていれば,

$$X^{-1}((a, b]) = X^{-1}((-\infty, a]^c \cap (-\infty, b]) = X^{-1}((-\infty, a])^c \cap X^{-1}((-\infty, b]) \in \mathcal{F}$$

$$X^{-1}([a, b]) = X^{-1}\left(\bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right]\right) = \bigcap_{n=1}^{\infty} X^{-1}\left(\left(a - \frac{1}{n}, b\right]\right) \in \mathcal{F}$$

$$X^{-1}((a, b)) = X^{-1}\left(\bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right]\right) = \bigcup_{n=1}^{\infty} X^{-1}\left(\left(a, b - \frac{1}{n}\right]\right) \in \mathcal{F}$$

となって, これらの区間に $X(\omega)$ が含まれる確率が計算できる. 同様に

$A = \text{“集合族 } (-\infty, x] \text{ } (x \in \mathbb{R}) \text{ から } \cap, \cup, ^c \text{ の操作で生成される } \mathbb{R} \text{ の部分集合”}$

とすると, (15) ならば $X^{-1}(A) \in \mathcal{F}$ である. 条件 (15) は, このような $A \subset \mathbb{R}$ について事象 $X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$ の確率が計算できるための要請である.

参考文献

- [1] 伏見正則, 確率と確率過程, 朝倉書店, 2004 (講談社, 1987).
- [2] 尾関和彦, 情報技術のための離散数学入門, 共立出版, 2004.
- [3] 佐藤坦, はじめての確率論: 測度から確率へ, 共立出版, 1994.
- [4] 熊谷隆, 確率論, 共立出版, 2003.
- [5] 森田啓義, 「確率論」講義資料, 2008.