2026年度　データサイエンス数学基礎/統計学　（担当：松澤　寛）　レジメ

No.7　確率変数と離散型確率分布

確率変数

2枚のコインを投げて表の出る枚数を X とする．

- X のとりうる値は $X = 0, 1, 2$
- $\boldsymbol{X}$ **の値は試行の結果により定まる**（偶然に支配される）が，**それぞれの値をとる確率は定まっている**．
- $X = k\ (k = 0, 1, 2)$ である確率を $P(X = k)$ とかく：

$$P(X=0) = \frac{1}{4},\quad P(X=1) = \frac{1}{2},\quad P(X=2) = \frac{1}{4}$$

試行の結果で値が定まり，各値をとる確率が定まっている変数を**確率変数**という．

確率変数のとりうる値とその値をとる確率の対応を**確率分布**，それを表にしたものを**確率分布表**という．

例1 上の例の X の確率分布（表）

$$X = \begin{cases} \dfrac{1}{4} & (k = 0, 2) \\ \dfrac{1}{2} & (k = 1) \end{cases},$$

k	0	1	2	計
$P(X=k)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

X のとりうる値を $x_1, x_2, \cdots, x_n$ とし，

$$p_i = P(X = x_i)\ (i = 1, 2, \cdots, n)$$

とすると次が成り立つ

$$\boldsymbol{0 \leq p_i \leq 1\ (i = 1, 2, \cdots, n)} \text{ かつ } \boldsymbol{\sum_{i=1}^{n} p_i = 1}$$

逆に，上をみたす数の組 $p_1, p_2, \cdots, p_n$ が与えられたとき，$p_i = P(X = x_i)$ は確率分布となる．確率分布は試行ではなく n 個の数，$x_1, \cdots, x_n$ とそれに対応する上をみたす数 $p_1, p_2, \cdots, p_n$ が本質的である．

確率変数の平均（期待値）

- 試行の回数 N が大きければ，$X = x_i$ となる回数はおよそ $p_i N$ 回（そもそもの確率の考え方）
- X のとる値の平均はおよそ

$$\frac{1}{N}(x_1 p_1 N + x_2 p_2 N + \cdots + x_n p_n N) = \sum_{i=1}^{n} x_i p_i$$

- この右辺を**確率変数 $\boldsymbol{X}$ の平均**または**期待値**といい，$\boldsymbol{E[X]}$ で表す：

$$\boldsymbol{E[X] = \sum_{i=1}^{n} x_i p_i}$$

右辺は試行が何かであることに関係なく，x_i「たち」と p_i「たち」だけで決まる．よって**確率分布の平均**ともいう．

例 2 最初の例の X について

$$
\begin{aligned}
E[X] &= 0 \times P(X=0) + 1 \times P(X=1) + 2 \times P(X=2) \\
&= 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1
\end{aligned}
$$

分散・標準偏差

- 確率変数 X の平均を μ とするとき，$(X-\mu)^2$ の平均を X の**分散**といい，$\boldsymbol{V[X]}$ で表す．
- $\sqrt{V[X]}$ を X の**標準偏差**といい $\boldsymbol{\sigma}$ で表す．

$$
\boldsymbol{V[X] = E[(X-\mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2 p_i, \quad \sigma = \sqrt{V[X]}}
$$

分散の公式

$$\boldsymbol{V[X] = E[X^2] - (E[X])^2 = E[X^2] - \mu^2}$$

平均・分散の性質

a, b を定数とするとき

- $E[aX+b] = aE[X]+b$
- $V[aX+b] = a^2V[X]$

二項分布 $\boldsymbol{B(n,p)}$

- 1 つのサイコロを投げるとき,「1 の目が出る」か「1 の目が出ないか」のいずれかの事象が起こる．
- 1 つの試行の結果，2 つの事象 A, B のいずれかが起こり，繰り返し行っても各回の事象でそれぞれが起こる確率が変わらなとき，この試行を**ベルヌーイ試行**という．
- ベルヌーイ試行を n 回繰り返すとき，事象 A が起こる回数を X とする．
- X は $\mathbf{0}$, $\mathbf{1}$, $\boldsymbol{\cdots}$, $\boldsymbol{n}$ の値をとる確率変数である．
- 1 回の試行で事象 A が起こる確率を $\boldsymbol{p}$ とすると，$X=k$ $(k=0,1,\cdots,n)$ となる確率は反復試行の確率より

$$
\boldsymbol{P(X=k) = {}_n\mathrm{C}_k p^k q^{n-k} \quad (0<p<1,\ q=1-p)}
$$

- 一般に確率変数 X のとりうる値が $0,1,\cdots,n$ で，各 $k=0,1,\cdots,n$ について $P(X=k)$ が上の式で与えられるとき，X の確率分布を**二項分布**といい，$\boldsymbol{B(n,p)}$ で表す．また，X は二項分布 $B(n,p)$ に**従う**といい，$\boldsymbol{X \sim B(n,p)}$ と表す．

注 二項定理より

$$
\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n} {}_n\mathrm{C}_k p^k q^{n-k} = (p+q)^n = 1^n = 1
$$

例 3 1 つのサイコロを 6 回投げるとき，1 の目が出る回数を X とする．

- 1回の試行で事象「1の目が出る」が起こる確率は $p = \dfrac{1}{6}$, $q = 1 - p = \dfrac{5}{6}$

- $P(X = k) = {}_6\mathrm{C}_k \left(\dfrac{1}{6}\right)^k \left(\dfrac{5}{6}\right)^{6-k}$

- $X \sim B\left(6, \dfrac{1}{6}\right)$

例題 1

赤玉3個と白玉6個が入っている袋の中から，1個ずつ3個復元抽出するとき，赤玉の出る回数を X とする．このとき，X はどのような確率分布に従うか．また，その確率分布表を作れ．

解

- 1回の試行で赤玉の出る確率は $\dfrac{3}{9} = \dfrac{1}{3}$

- 各回の試行は独立であるから

$$P(X = k) = {}_3\mathrm{C}_k \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{3-k} \quad (k = 0, 1, 2, 3)$$

つまり X は二項分布 $B\left(3, \dfrac{1}{3}\right)$ に従う．

$$P(X = 0) = {}_3\mathrm{C}_0 \left(\frac{2}{3}\right)^3 = \frac{8}{27}, \quad P(X = 1) = {}_3\mathrm{C}_1 \left(\frac{1}{3}\right) \left(\frac{1}{3}\right)^2 = \frac{12}{27}$$
$$P(X = 2) = {}_3\mathrm{C}_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right) = \frac{6}{27}, \quad P(X = 3) = {}_3\mathrm{C}_3 \left(\frac{1}{3}\right)^3 = \frac{1}{27}$$

よって確率分布表は

k	0	1	2	3	計
$P(X = k)$	$\frac{8}{27}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{1}{27}$	1

□

二項分布の平均と分散

二項分布 $B(n, p)$ について

平均は $\boldsymbol{np}$,　分散は $\boldsymbol{npq}$　$(q = 1 - p)$

つまり $X \sim B(n, p)$ のとき

$$E[X] = np, \quad V[X] = npq$$