# 確率分布と確率密度関数

## 確率変数と確率分布

### 確率変数 (random variable)

- 試行の結果、ある値をとる確率が決まる変数を、「**確率変数**」という
- サイコロを1回投げる場合を考える
  - サイコロの出た目の数 {1, 2, 3, 4, 5, 6} を $X$ (確率変数) とする
    - 確率変数は大文字で書く
  - $X=1$ (つまり1の目がでる) の事象の確率は、次のように表すことができる

$$\mathrm{P}(X=1)=\frac{1}{6}$$

  - 同じように、1以外の目が出る確率は、次のように表せる

$$\mathrm{P}(X=2)=\cdots=\mathrm{P}(X=6)=\frac{1}{6}$$

  - なお、$X=1$ という事象は、$\{X=1\}$ とも表せる

### 確率分布 (probability distribution)

- サイコロを1回投げたときにでた目の数を確率変数 $X$ を使うと、その確率は次のようになる

$$\mathrm{P}(X=1)=\cdots=\mathrm{P}(X=6)=\frac{1}{6}$$

- 確率変数 $X$ のとる値と、それに対応する確率を表にまとめたもの、つまり確率変数 $X$ に対応する確率の分布を、「**確率分布**」という

| $X$ | 1 | 2 | 3 | 4 | 5 | 6 | 計 |
|---|---|---|---|---|---|---|---|
| **確率** | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | 1 |

- 一般に、確率変数 $X$ が、次のような n 個の値をとるとき、

$$x_1,x_2,\cdots,x_n$$

- その確率が次のようになるのであれば、

$$\mathrm{P}(X=x_k)=p_k\,(k=1,2,\cdots,n)$$

- 次のことが成り立つ

$$\begin{cases} p_1\geq 0, p_2\geq 0,\cdots,p_n\geq 0 \\ p_1+p_2+\cdots+p_n=1 \end{cases}$$

| $X$ | $x_1$ | $x_2$ | $\cdots$ | $x_n$ | 計 |
|---|---|---|---|---|---|
| **確率** | $p_1$ | $p_2$ | $\cdots$ | $p_n$ | 1 |

### 確率分布の例

- サイコロを1回投げたときにでた目の数が奇数か偶数かを考える
  - 奇数がでたときの確率変数を $Y=0$ 、偶数 (=奇数でない) がでたときの確率変数を $Y=1$ とする
  - 確率変数 $Y$ の確率分布は、次のようになる

| $Y$ | 0 | 1 | 計 |
|---|---|---|---|
| **確率** | $\frac{3}{6}$ | $\frac{3}{6}$ | $1$ |

- コインを1回投げたときに表が出るか裏が出るかを考える
  - 表が出る回数を、確率変数 $X$ で表すと、その確率分布は次のようになる

| $X$ | 0 | 1 | 計 |
|---|---|---|---|
| **確率** | $\frac{1}{2}$ | $\frac{1}{2}$ | $1$ |

# 確率密度関数

## 確率変数の種類

- これまで考えてきたのは、確率変数が離散的な（飛び飛びの値を取る）場合である
  - このような確率変数を、離散型確率変数（または離散変数）という
- 確率変数が連続的な値を場合もある（身長、体重、年齢など）
  - このような確率変数を、連続型確率変数（または連続変数）という

## 確率密度関数と累積分布関数

- 「確率変数 $X$ のとる値が $x$ 以下である」という事象とその確率を

$$\{X \leq x\},\ \ \mathrm{P}(X \leq x) = F(x)$$

と表し、関数 $F(x)$ を「**累積分布関数（または分布関数）**」という
  - つまり累積分布関数は、確率変数 $x$ が最小値から指定された値までをとる間の確率（最小値に対応する確率から指定された値までに対応する確率をすべて足したもの）を与える
- 連続型の確率変数の場合は、その分布関数も連続的になる（グラフは曲線になる）
- 連続した確率変数 $x$ がある区間 $a \leq x \leq b$ の値をとる確率は、関数の曲線と $x$ 軸の囲む図形の面積になる
- このような、面積が確率を与えるような関数を「**確率密度関数**」という

$$\mathrm{P}(a \leq x \leq b) = \int_a^b f(x)dx$$

# 確率分布の分類

確率変数が離散的か連続的かで、次のように確率分布を分類することができます。テキストで取り上げられている、おもな分布を挙げておきます。

- 離散型確率変数（コイン、サイコロ、トランプなど）
  - 二項分布（60ページ）
    - ポアソン分布（61ページ）
    - 幾何分布（63ページ）
  - 累積一様分布（59ページ）
- 連続型確率変数（大きさ、重さ、長さなど）
  - 正規分布（53ページ）
    - 標準正規分布（55ページ）