確率変数の期待値と分散

期待値（平均値）

期待値とは

- 確率変数 X の確率分布が次のようなとき、

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

- 確率変数 X の平均値、または期待値は、次のように表せる

$$\mu=E(X)=x_1p_1+x_2p_2+\cdots+x_np_n$$

- 期待値とは、1回の試行の結果として期待される値の大きさを表す

期待値の計算例（1）

- 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、その確率分布は次のようになる

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

- 確率変数 X の期待値（平均値）は、 $n=6$ なので、

$$x_1=1,\ x_2=2,\ \cdots,\ x_6=6$$

$$p_1=\frac{1}{6},\ p_2=\frac{1}{6},\ \cdots,\ p_6=\frac{1}{6}$$

- したがって、次のようになる

$$\begin{aligned}E(X)&=1\times\frac{1}{6}+2\times\frac{1}{6}+\cdots+6\times\frac{1}{6}\\&=\frac{21}{6}=3.5\end{aligned}$$

- つまり、サイコロを何回も投げたときに、でた目の平均をとると 3.5 になることを示している

期待値の計算例（2）

- サイコロを5回連続で投げたときに1の目が出る回数を確率変数 X とすると、その確率分布は次のようになる

X	0	1	2	3	4	5
確率	0.4019	0.4019	0.1608	0.0322	0.0032	0.0001

- 確率変数 X の期待値（平均値）は、 $n=6$ なので、

$$x_1=0,\ x_2=1,\ \cdots,\ x_6=5$$

$$p_1=0.4019,\ p_2=0.4019,\ \cdots,\ p_6=0.0001$$

- したがって、次のようになる

$$\begin{aligned}E(X)&=0\times0.4019+1\times0.4019+2\times0.1608+3\times0.0322+4\times0.0032+5\times0.0001\\&\simeq0.83\end{aligned}$$

- つまり、サイコロを5回連続投げて1の目が出るのは1回あるかないか程度であることを示している

期待値の計算例（3）

- 宝くじの期待値を求めることもできる。宝くじの場合は「当せん金×当せん確率」の合計が期待値となる。
- 例えば、平成21年年末ジャンボ宝くじは、1ユニット（1000万枚）あたり、次のような当せん本数になっている。なお、当せん確率は「当せん本数÷1000万×100」から求めている。

等級	当せん金	当せん本数	当せん確率
1等	200,000,000円	1本	0.00001%
1等前後賞	50,000,000円	2本	0.00002%
1等組違い賞	100,000円	99本	0.00099%
2等	100,000,000円	2本	0.00002%
3等	5,000,000円	10本	0.0001%
4等	100,000円	600本	0.006%
5等	10,000円	10,000本	0.1%
6等	3,000円	100,000本	1%
7等	300円	1,000,000本	10%
元気に2010年賞	1,000,000円	100本	0.001%

- 宝くじがいくら当たるかの期待値を調べるには、「当せん金×当せん確率」の合計を求めるので、

$$\begin{aligned} E(X) &= 200{,}000{,}000 \times 0.00001\% + 50{,}000{,}000 \times 0.00002\% + \cdots + 300 \times 10\% + 1{,}000{,}000 \times 0.001\% \\ &= 141.99 \end{aligned}$$

- つまり、宝くじ1枚（300円）を買うと、1枚につき141.99円の還元が期待できる、ということを示している。

期待値と算術平均との関係

- n 個のデータ $x_1, x_2, \cdots, x_n$ の平均値は、次のように表せる

$$\begin{aligned} \bar{x} &= \frac{x_1 + x_2 + \cdots + x_n}{n} \\ &= x_1 \times \frac{1}{n} + x_2 \times \frac{1}{n} + \cdots + x_n \times \frac{1}{n} \end{aligned}$$

- ここで確率について、$p_1 = \frac{1}{n}, p_2 = \frac{1}{n}, \cdots, p_n = \frac{1}{n}$ とおく、つまり各々の確率が等しいと考えると、

$$\begin{aligned} \bar{x} &= x_1 p_1 + x_2 p_2 + \cdots + x_n p_n \\ &= E(X) \end{aligned}$$

- すなわち、各々の確率が等しくても等しくなくても、平均値（期待値）を求めることができる

分散と標準偏差

確率変数の分散と標準偏差

- 確率変数 X の確率分布が次のようなとき、

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

- 確率変数 X の分散は次のように表す

$$\sigma^2 = V(X) = (x_1-\mu)^2 p_1 + (x_2-\mu)^2 p_2 + \cdots + (x_n-\mu)^2 p_n$$
$$= \sum_{i=1}^{n} (x_i-\mu) p_i$$

- μ は期待値 $E(X)$ 簡単に表したもの
- 分散の正の平方根を、確率変数 X の標準偏差といい、次のように表す

$$\sigma = \sqrt{V(X)}$$

確率変数の分散と標準偏差の特徴

- 分散や標準偏差が小さいほど、確率変数の値は平均に集中し、ばらつきが小さい
- 分散や標準偏差が大きいほど、確率変数の値は平均から離れ、ばらつきが大きい
- 分散は変数の単位の2乗を表す（例えば変数の単位がcmなら、分散の単位cm^2）ため、元の単位と同じ標準偏差を用いて平均からのばらつきを表す

確率変数の分散と標準偏差の計算

- 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、その確率分布は次のようになる

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

- したがって、分散は次のようにして求められる

$$\sigma^2 = V(X) = (1-3.5)^2 \times \frac{1}{6} + (2-3.5)^2 \times \frac{1}{6} + \cdots + (6-3.5)^2 \times \frac{1}{6}$$
$$= \frac{35}{12} \simeq 2.92$$

- また、標準偏差は次のようになる

$$\sigma = \sqrt{V(X)}$$
$$= \sqrt{\frac{35}{12}} \simeq 1.71$$

- つまり、サイコロを何回も投げたとき、そのでた目の平均が 3.5 ± 1.71 （1.79〜5.21）の範囲になる確率が高いことを示している