

## 第157講 確率変数の分散・標準偏差

### 学習目標

この講では、確率変数の値が平均からどれだけ離れているかを数値化する方法を学ぶ。  
教材終了後に、次のことができるようになる。

- ・ 確率変数の平均からのずれを二乗して平均する意味を説明できる。
- ・ 有限個の値をとる確率変数について、分散と標準偏差を定義できる。
- ・  $V(X) = E(X^2) - \{E(X)\}^2$  を用いて計算できる。
- ・ 分散と標準偏差の単位や、分散が0になる場合の意味を判断できる。

### 平均からのずれを測る理由

確率変数の値 $x_i$ と平均 $\mu$ の差

$x_i - \mu$ を考える。

これは平均より上にあるか下にあるかを表す符号付きのずれであり、値そのものの距離を表しているわけではない。

平均からの距離を考える場合は、符号をなくした $|x_i - \mu|$ を見る必要がある。

しかし、ずれを二乗すれば符号の打消しを避け、平均からの隔たりを非負の値として扱うことができるため、分散ではずれを二乗して扱う。

また、符号付きのずれをそのまま期待値にすると、有限分布では次のように必ず0になる。

$$\begin{aligned} E(X - \mu) &= \sum_{i=1}^n (x_i - \mu)p_i \\ &= \sum_{i=1}^n x_i p_i - \mu \sum_{i=1}^n p_i \\ &= \mu - \mu \times 1 \\ &= 0 \end{aligned}$$

これは、平均より上のずれと下のずれが打ち消し合うためである。

そこで、分散ではずれを二乗する。

同じ正の確率で起こる値を比較する場合、平均から遠い値ほど二乗したずれが大きくなり、分散への寄与も大きくなる。

ただし、分散への寄与の大きさは確率も掛け合わせて決まるため、距離だけで大小を判断してはいけない。

### 分散と標準偏差

$n$ を正整数とする。

確率変数 $X$ が異なる $n$ 個の値 $x_1, x_2, \dots, x_n$ をとり、それぞれの確率を $p_1, p_2, \dots, p_n$ とする。

このとき、各確率は $p_i \geq 0$  ( $i = 1, 2, \dots, n$ )を満たし、

$$\sum_{i=1}^n p_i = 1$$

である。

平均を $\mu = E(X)$ とすると、分散 $V(X)$ を

$$V(X) = \sum_{i=1}^n (x_i - \mu)^2 p_i$$

で定める。

各項について、 $(x_i - \mu)^2 p_i \geq 0$ であるため、分散は必ず

$V(X) \geq 0$ を満たす。

標準偏差 $\sigma$ は、分散の非負の平方根として

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

で定める。

$V(X) > 0$ の場合、方程式 $s^2 = V(X)$ の解は正と負の2つ存在するが、 $V(X) = 0$ の場合の解は0だけである。いずれの場合も、標準偏差は散らばりの大きさを表す量であるため、非負の値を採用する。

## 分散の別の計算方法

分散の定義

$$V(X) = \sum_{i=1}^n (x_i - \mu)^2 p_i$$

は意味を理解するには重要であるが、計算では展開した形が便利である。

二乗を展開すると、

$$\begin{aligned} V(X) &= \sum_{i=1}^n (x_i^2 - 2\mu x_i + \mu^2) p_i \\ &= \sum_{i=1}^n x_i^2 p_i - 2\mu \sum_{i=1}^n x_i p_i + \mu^2 \sum_{i=1}^n p_i \end{aligned}$$

となる。

ここで、

$$E(X^2) = \sum_{i=1}^n x_i^2 p_i$$

であり、

$$E(X) = \sum_{i=1}^n x_i p_i = \mu$$

である。また、

$$\sum_{i=1}^n p_i = 1$$

なので、

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

となる。

ここで注意すべきことは、 $E(X^2)$ は「値を二乗してから期待値をとる」ものであり、 $\{E(X)\}^2$ は「期待値を求めてから二乗する」ものである。一般には同じではない。

### 分散の大きさと単位

元の値の単位がcmであるとき、ずれ $x_i - \mu$ もcmである。

分散ではずれを二乗するため、単位は $\text{cm}^2$ になる。

一方、標準偏差は分散の平方根をとるため、単位はcmとなり、元の値と同じ単位になる。

また、分散が0の場合を考える。

$$V(X) = \sum_{i=1}^n (x_i - \mu)^2 p_i$$

は、すべての項が非負の和である。

したがって、この和が0になるためには、正の確率が付いているすべての値について、

$$(x_i - \mu)^2 p_i = 0$$

でなければならない。

$p_i > 0$ ならば、 $(x_i - \mu)^2 = 0$ なので、 $x_i = \mu$ となる。

逆に、正の確率が付くすべての値が $\mu$ であれば、すべての項が0になり、 $V(X) = 0$ となる。

なお、確率0の候補値は、値が $\mu$ と異なっても分散への寄与が

$$(x_i - \mu)^2 \times 0 = 0 \text{ となるため、分散0の条件を考えるときには影響しない。}$$

#### 例題 1

確率変数 $X$ が次の表のように値をとるとき、平均 $\mu$ 、分散 $V(X)$ 、標準偏差 $\sigma$ を求めよ。

| $x_i$        | 0             | 1             | 2             |
|--------------|---------------|---------------|---------------|
| $P(X = x_i)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

#### 例題 1 の解答・解説

平均を求める。

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

次に、分散を定義から求める。

$$\begin{aligned} V(X) &= (0 - 1)^2 \times \frac{1}{4} + (1 - 1)^2 \times \frac{1}{2} + (2 - 1)^2 \times \frac{1}{4} \\ &= \frac{1}{4} + 0 + \frac{1}{4} \\ &= \frac{1}{2} \end{aligned}$$

したがって、

$$\sigma = \sqrt{V(X)} = \sqrt{\frac{1}{2}} = \frac{\sqrt{2}}{2}$$

である。

**例題 2**

確率変数  $X$  が次の表のように値をとるとき、平均  $\mu$ 、分散  $V(X)$ 、標準偏差  $\sigma$  を求めよ。

|              |               |               |
|--------------|---------------|---------------|
| $x_i$        | -1            | 1             |
| $P(X = x_i)$ | $\frac{3}{4}$ | $\frac{1}{4}$ |

**例題 2 の解答・解説**

平均を求める。

$$\begin{aligned}
 \mu &= E(X) \\
 &= (-1) \times \frac{3}{4} + 1 \times \frac{1}{4} \\
 &= -\frac{1}{2}
 \end{aligned}$$

分散は平均からのずれの二乗の期待値である。

$$\begin{aligned}
 V(X) &= \left(-1 + \frac{1}{2}\right)^2 \times \frac{3}{4} + \left(1 + \frac{1}{2}\right)^2 \times \frac{1}{4} \\
 &= \left(-\frac{1}{2}\right)^2 \times \frac{3}{4} + \left(\frac{3}{2}\right)^2 \times \frac{1}{4} \\
 &= \frac{1}{4} \times \frac{3}{4} + \frac{9}{4} \times \frac{1}{4} \\
 &= \frac{3}{16} + \frac{9}{16} \\
 &= \frac{3}{4}
 \end{aligned}$$

したがって、

$$\sigma = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}$$

である。

**例題 3**

次の 2 つの確率分布を比較し、それぞれの平均、分散、標準偏差を求めよ。

分布  $X$  :

|          |               |               |
|----------|---------------|---------------|
| <b>値</b> | -1            | 1             |
| 確率       | $\frac{1}{2}$ | $\frac{1}{2}$ |

分布  $Y$  :

|          |               |               |
|----------|---------------|---------------|
| <b>値</b> | -3            | 3             |
| 確率       | $\frac{1}{2}$ | $\frac{1}{2}$ |

### 例題 3 の解答・解説

分布  $X$  について、

$$\begin{aligned} E(X) &= (-1) \times \frac{1}{2} + 1 \times \frac{1}{2} \\ &= 0 \end{aligned}$$

である。

平均からのずれは  $-1$  と  $1$  なので、

$$V(X) = (-1)^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2} = 1$$

したがって、 $\sigma_X = \sqrt{1} = 1$  である。

次に分布  $Y$  について、

$$\begin{aligned} E(Y) &= (-3) \times \frac{1}{2} + 3 \times \frac{1}{2} \\ &= 0 \end{aligned}$$

である。

平均からのずれは  $-3$  と  $3$  なので、

$$V(Y) = (-3)^2 \times \frac{1}{2} + 3^2 \times \frac{1}{2} = 9$$

したがって、 $\sigma_Y = \sqrt{9} = 3$  である。

2 つの分布はどちらも平均は  $0$  であるが、分散と標準偏差が異なる。

この例では、 $Y$  の値は  $X$  よりも平均  $0$  から遠い位置にあるため、ずれの二乗の平均である分散も大きくなる。

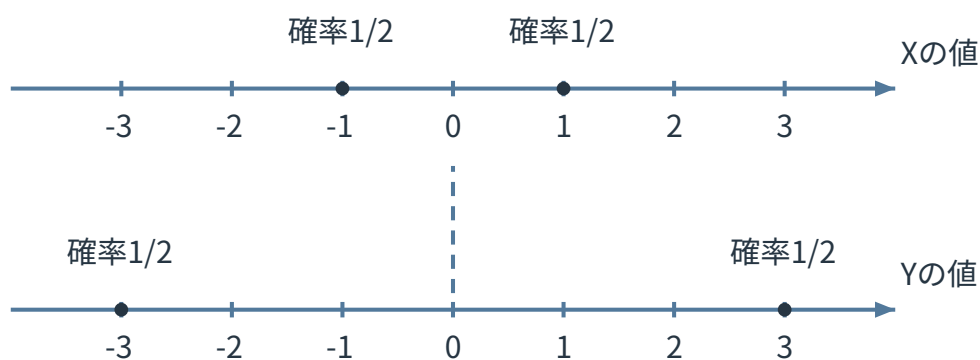


図 1  $X$  は  $-1$  と  $1$ 、 $Y$  は  $-3$  と  $3$  をそれぞれ確率  $1/2$  でとり、期待値はいずれも  $0$  である。上下の横軸は同じ尺度で、 $Y$  の値は  $X$  の値より中心  $0$  から離れている。縦の配置は分布を分けるためだけのもので、値の高さや確率の軸を表さない。

## 演習問題

### 練習問題 1

確率変数  $X$  が次の表のように値をとるとき、分散  $V(X)$  と標準偏差  $\sigma$  を求めよ。

|              |               |               |               |
|--------------|---------------|---------------|---------------|
| $x_i$        | 0             | 2             | 4             |
| $P(X = x_i)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

分散を定義から計算せよ。

### 練習問題 1 の解答・解説

平均を求める。

$$\begin{aligned}\mu &= 0 \times \frac{1}{4} + 2 \times \frac{1}{2} + 4 \times \frac{1}{4} \\ &= 2\end{aligned}$$

平均からのずれを二乗して期待値をとる。

$$\begin{aligned}V(X) &= (0 - 2)^2 \times \frac{1}{4} + (2 - 2)^2 \times \frac{1}{2} + (4 - 2)^2 \times \frac{1}{4} \\ &= 4 \times \frac{1}{4} + 0 + 4 \times \frac{1}{4} \\ &= 2\end{aligned}$$

したがって、 $\sigma = \sqrt{2}$  である。

平均からの距離そのものではなく、ずれを二乗してから平均している点が分散の特徴である。

### 練習問題 2

確率変数  $X$  が次の表のように値をとるとき、 $E(X^2)$  を用いて分散  $V(X)$  を求めよ。

|              |               |               |               |
|--------------|---------------|---------------|---------------|
| $x_i$        | -2            | 0             | 3             |
| $P(X = x_i)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

ただし、 $V(X) = E(X^2) - \{E(X)\}^2$  を用いてよい。

## 練習問題 2 の解答・解説

まず平均を求める。

$$\begin{aligned} E(X) &= (-2) \times \frac{1}{4} + 0 \times \frac{1}{2} + 3 \times \frac{1}{4} \\ &= \frac{1}{4} \end{aligned}$$

次に  $E(X^2)$  を求める。

$$\begin{aligned} E(X^2) &= (-2)^2 \times \frac{1}{4} + 0^2 \times \frac{1}{2} + 3^2 \times \frac{1}{4} \\ &= 1 + \frac{9}{4} \\ &= \frac{13}{4} \end{aligned}$$

よって、

$$\begin{aligned} V(X) &= E(X^2) - \{E(X)\}^2 \\ &= \frac{13}{4} - \left(\frac{1}{4}\right)^2 \\ &= \frac{13}{4} - \frac{1}{16} \\ &= \frac{51}{16} \end{aligned}$$

である。

$E(X^2)$  は値を二乗してから期待値をとる値であり、 $\{E(X)\}^2$  は平均を求めてから二乗した値なので、順序が異なる。

## 練習問題 3

確率変数  $X$  が次の表のように値をとる。

|              |               |               |
|--------------|---------------|---------------|
| $x_i$        | 1             | 5             |
| $P(X = x_i)$ | $\frac{1}{4}$ | $\frac{3}{4}$ |

ある生徒は平均を

$$\frac{1+5}{2} = 3$$

として計算した。

この計算が誤りである理由を説明し、正しい平均  $\mu$  と分散  $V(X)$  を求めよ。

### 練習問題 3 の解答・解説

誤りの原因は、2つの値の単純平均をとっていることである。  
確率変数の平均は、それぞれの値を確率で重み付けした期待値で求める。

$$\begin{aligned}\mu &= 1 \times \frac{1}{4} + 5 \times \frac{3}{4} \\ &= 4\end{aligned}$$

したがって、平均は3ではなく4である。  
分散は、

$$\begin{aligned}V(X) &= (1-4)^2 \times \frac{1}{4} + (5-4)^2 \times \frac{3}{4} \\ &= 9 \times \frac{1}{4} + 1 \times \frac{3}{4} \\ &= 3\end{aligned}$$

である。  
確率が異なる場合、各値を同じ重さで扱ってはいけない。

#### 練習問題 4

確率変数  $X$  が次の条件を満たす。

|              |   |   |    |
|--------------|---|---|----|
| $x_i$        | 2 | 5 | 10 |
| $P(X = x_i)$ | 0 | 1 | 0  |

平均  $\mu$  と分散  $V(X)$  を求めよ。  
また、分散が0になる理由を説明せよ。

### 練習問題 4 の解答・解説

確率変数  $X$  は、値5を確率1でとる。  
したがって、 $\mu = E(X) = 5$  である。  
分散を定義から求める。

$$\begin{aligned}V(X) &= (2-5)^2 \times 0 + (5-5)^2 \times 1 + (10-5)^2 \times 0 \\ &= 0 + 0 + 0 \\ &= 0\end{aligned}$$

である。  
確率0の候補値2, 10は、分散の式では確率を掛けるため、それらのずれが分散へ与える寄与は0になる。  
また、確率1で5をとるため、平均も5となり、その値からのずれは必ず0である。  
したがって、分散は0になる。  
逆に、分散が0である場合も、正の確率が付く値はすべて平均と一致する。

#### 練習問題 5

確率変数  $X$  が次の表のように値をとる。

|              |               |               |               |
|--------------|---------------|---------------|---------------|
| $x_i$        | -2            | 0             | 2             |
| $P(X = x_i)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

平均からのずれの期待値  $E(X - \mu)$  を求めよ。

また、この値が0になることと、分散が0になることの違いを説明し、 $V(X)$  を求めよ。

### 練習問題 5 の解答・解説

まず平均を求める。

$$\begin{aligned}\mu &= (-2) \times \frac{1}{4} + 0 \times \frac{1}{2} + 2 \times \frac{1}{4} \\ &= 0\end{aligned}$$

したがって、

$$\begin{aligned}E(X - \mu) &= E(X) \\ &= 0\end{aligned}$$

である。

しかし、これは分散が0であることを意味しない。

ずれをそのまま平均すると、正のずれと負のずれが打ち消し合う。

分散では二乗するため、

$$\begin{aligned}V(X) &= (-2)^2 \times \frac{1}{4} + 0^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} \\ &= 2\end{aligned}$$

となる。

ずれの期待値が0になることと、値が平均から離れていないことは別である。

### 練習問題 6

確率変数  $X$  が次の表のように値をとる。

|              |               |               |
|--------------|---------------|---------------|
| $x_i$        | 0             | 3             |
| $P(X = x_i)$ | $\frac{1}{2}$ | $\frac{1}{2}$ |

次の 2 つの値を求め、違いを説明せよ。

$E(X^2)$

と

$\{E(X)\}^2$

## 練習問題 6 の解答・解説

平均は、

$$E(X) = 0 \times \frac{1}{2} + 3 \times \frac{1}{2} = \frac{3}{2}$$

である。

まず、

$$\begin{aligned} E(X^2) &= 0^2 \times \frac{1}{2} + 3^2 \times \frac{1}{2} \\ &= \frac{9}{2} \end{aligned}$$

である。

一方、

$$\begin{aligned} \{E(X)\}^2 &= \left(\frac{3}{2}\right)^2 \\ &= \frac{9}{4} \end{aligned}$$

となる。

2 つは異なる値になる。

$E(X^2)$  は「先に二乗してから平均」、 $\{E(X)\}^2$  は「先に平均してから二乗」であり、同じ操作ではない。

## 練習問題 7

確率変数  $X$  が長さを表す値をとり、その単位は cm である。

分散  $V(X) = 16$  であるとき、標準偏差  $\sigma$  の値と単位を答えよ。

また、分散と標準偏差の単位が異なる理由を説明せよ。

## 練習問題 7 の解答・解説

分散  $V(X) = 16$  であるから、標準偏差は分散の非負の平方根をとる。

$\sigma = \sqrt{16} = 4$  である。

元の値の単位が cm なので、標準偏差の単位も同じになる。

したがって、標準偏差は 4 cm である。

分散はずれを二乗して平均した量であるため、単位も二乗されて  $\text{cm}^2$  となる。

その平方根をとった標準偏差は、cm となり、元の値と同じ単位で表される。

## 練習問題 8

次の 2 つの確率分布  $X, Y$  を比較する。

分布  $X$  :

| 値  | -2            | 2             |
|----|---------------|---------------|
| 確率 | $\frac{1}{2}$ | $\frac{1}{2}$ |

分布Y:

| 値  | -4            | 4             |
|----|---------------|---------------|
| 確率 | $\frac{1}{2}$ | $\frac{1}{2}$ |

それぞれの平均、分散、標準偏差を求め、平均が同じでも分布の広がりがない理由を説明せよ。

### 練習問題 8 の解答・解説

分布Xについて、

$$\mu_X = (-2) \times \frac{1}{2} + 2 \times \frac{1}{2} = 0$$

である。

分散は、

$$V(X) = (-2)^2 \times \frac{1}{2} + 2^2 \times \frac{1}{2} = 4$$

したがって、 $\sigma_X = 2$ である。

分布Yについても、

$$\mu_Y = (-4) \times \frac{1}{2} + 4 \times \frac{1}{2} = 0$$

である。

しかし、

$$V(Y) = (-4)^2 \times \frac{1}{2} + 4^2 \times \frac{1}{2} = 16$$

なので、 $\sigma_Y = 4$ となる。

どちらも平均は0であるが、平均からの距離が異なる。

平均は中心の位置しか表さず、分布の広がりを表すには分散や標準偏差を見る必要がある。