

§ 2.4 確率変数の分散と標準偏差

1.3節で述べたように、データの散布度として分散があった。 データに n 個の要素

$$x_1, x_2, , x_3, \dots, x_n$$

があるとき、その平均 $\bar{x} = \frac{1}{n} \sum_{k=1}^{\infty} x_k$ に対して、データの分散 $\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$ は、
 $(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, (x_3 - \bar{x})^2, \dots, (x_n - \bar{x})^2$

の平均である。つまり、データの分散とは、データの各々の要素と平均との差の2乗の平均である。 同様に、確率変数の分散とは、確率変数の値と平均値との差の2乗の平均値である。

定数 μ に対して、定理 2.2 により、確率変数 X の2次関数 $(X - \mu)^2$ は確率変数である。

[定義] 確率変数 X の平均値を μ とおく： $\mu = E[X]$ 。 X の**分散** (variance) $V[X]$ を次のように定義する：

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

更に、 X の**標準偏差** (standerd deviation) $\sigma[X]$ を次のように定義する：

$$\sigma[X] = \sqrt{V[X]} .$$

[例] 成績が5点、3点、1点、0点、の4段階で点数化されるものとする。成績の点数は、5である確率は30%であり、3である確率は40%であり、1である確率は20%であり、0である確率は10%であるとする。成績の点数を表す変数 X は離散型確率変数であり、 X の平均値は $E[X] = \frac{29}{10}$ 。 X の分散 $V[X]$ と標準偏差 $\sigma[X]$ とを求める。

$$\begin{aligned} V[X] &= \left(5 - \frac{29}{10}\right)^2 P(X=5) + \left(3 - \frac{29}{10}\right)^2 P(X=3) \\ &\quad + \left(1 - \frac{29}{10}\right)^2 P(X=1) + \left(0 - \frac{29}{10}\right)^2 P(X=0) \\ &= \left(\frac{21}{10}\right)^2 \cdot \frac{3}{10} + \left(\frac{1}{10}\right)^2 \cdot \frac{4}{10} + \left(-\frac{19}{10}\right)^2 \cdot \frac{2}{10} + \left(-\frac{29}{10}\right)^2 \cdot \frac{1}{10} \\ &= \frac{441}{100} \cdot \frac{3}{10} + \frac{1}{100} \cdot \frac{4}{10} + \frac{361}{100} \cdot \frac{2}{10} + \frac{841}{100} \cdot \frac{1}{10} \\ &= \frac{289}{100} . \end{aligned}$$

$$\sigma[X] = \sqrt{V[X]} = \sqrt{\frac{289}{100}} = \frac{17}{10} . \quad \text{[終]}$$

[問2.4.1] 成績が4点、3点、2点、1点、0点、の5段階で点数化されるものとする。成績の点数は、4である確率は10%であり、3である確率は20%であり、2である確率は40%であり、1である確率は20%であり、0である確率は10%であるとする。成績の点数を表す変数 X は離散型確率変数であり、 X の平均値は $E[X] = 2$ 。 X の分散 $V[X]$ と標準偏差 $\sigma[X]$ とを求めよ。

定理 2.3.3 により、連続型確率変数の分散を求める式は次のようになる。

[定理 2.4.1] 関数 $f(x)$ が確率変数 X の確率密度関数であるとき、 X の平均値を μ とおくと

$$V[X] = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx .$$

[例] 実数全体を定義域とする関数 $f(x)$ を次のように定める：

$$f(x) = \begin{cases} \frac{x(6-x)}{36} & (0 \leq x \leq 6 \text{ のとき}) \\ 0 & (x < 0 \text{ または } x > 6 \text{ のとき}) \end{cases} .$$

各実数 x について $f(x) \geq 0$ で、 $\int_{-\infty}^{\infty} f(x) = 1$ なので、関数 $f(x)$ はある連続型確率変数 X の確率密度関数である。確率変数 X の平均値は $E[X] = \int_0^6 \frac{x(6-x)}{36} dx = 3$ 。確率変数 X の分散 $V[X]$ と標準偏差 $\sigma[X]$ とを求める。

$$V[X] = E[(X - 3)^2] = \int_{-\infty}^{\infty} (x - 3)^2 f(x) dx = \int_0^6 (x - 3)^2 \frac{x(6-x)}{36} dx .$$

変数 y を $y = x - 3$ とおく。 $x = y + 3$ 。 $dx = dy$ 。 $x = 0$ のとき $y = -3$ 。 $x = 6$ のとき $y = 3$ 。

$$\begin{aligned} \int_0^6 (x - 3)^2 \frac{x(6-x)}{36} dx &= \int_{-3}^3 y^2 \frac{(y+3)(3-y)}{36} dy = \frac{1}{36} \int_{-3}^3 (9y^2 - y^4) dy \\ &= \frac{1}{36} \left[3y^3 - \frac{1}{5} y^5 \right]_{-3}^3 = \frac{1}{4 \cdot 3^2} \left(3^4 - \frac{1}{5} \cdot 3^5 + 3^4 \frac{1}{5} \cdot 3^5 \right) \\ &= \frac{1}{4} \left(2 \cdot 3^2 - \frac{2}{5} \cdot 3^3 \right) = \frac{9}{2} \left(1 - \frac{3}{5} \right) \\ &= \frac{9}{5} . \end{aligned}$$

$$\text{故に } V[X] = \frac{9}{5} . \text{ また } \sigma[X] = \sqrt{\frac{9}{5}} = \frac{3}{\sqrt{5}} . \quad \text{[終]}$$

[問2.4.2] 実数全体を定義域とする関数 $f(x)$ を問題 2.1.2 の関数と同じものと定める：

$$f(x) = \begin{cases} \frac{3}{32} (2x + 3 - x^2) & (-1 \leq x \leq 3 \text{ のとき}) \\ 0 & (x < -1 \text{ または } x > 3 \text{ のとき}) \end{cases} .$$

問 2.1.2 で調べたように、関数 $f(x)$ は確率変数の確率密度関数である。問 2.3.2 で計算したように、関数 $f(x)$ を確率密度関数とする連続型確率変数 X の平均値は $E[X] = 1$ 。確率変数 X の分散 $V[X]$ と標準偏差 $\sigma[X]$ とを求めよ。

以下の定理が成り立つ。

[定理 2.4.2] 確率変数 X の分散 $V[X]$ があるならば、定数 c に対して、確率変数 $X + c$ の分散 $V[X + c]$ 及び標準偏差 $\sigma[X + c]$ は

$$V[X + c] = V[X] , \quad \sigma[X + c] = \sigma[X] .$$

[定理 2.4.3] 確率変数 X の分散 $V[X]$ があるならば、定数 k に対して、確率変数 kX の分散 $V[kX]$ 及び標準偏差 $\sigma[kX]$ は

$$V[kX] = k^2 V[X] , \quad \sigma[kX] = |k| \sigma[X] .$$

[例] 確率変数 X について $E[X] = 8$, $V[X] = 5$ とする。変数 Y を $Y = \frac{7 - 2X}{3}$ とおく。確率変数 Y の分散 $V[Y]$ と標準偏差 $\sigma[Y]$ とを求める。

$$\begin{aligned} V[Y] &= V\left[\frac{7 - 2X}{3}\right] = \left(\frac{1}{3}\right)^2 V[7 - 2X] = \frac{1}{9} V[-2X] = \frac{1}{9} (-2)^2 V[X] = \frac{4}{9} \cdot 5 \\ &= \frac{20}{9} . \end{aligned}$$

$$\sigma[Y] = \sqrt{V[Y]} = \sqrt{\frac{20}{9}} = \frac{\sqrt{20}}{3} . \quad \text{[終]}$$

[問2.4.3] 確率変数 X について $E[X] = 7$, $V[X] = 3$ とする。変数 Y を $Y = \frac{9 - 4X}{5}$ とおく。確率変数 Y の分散 $V[Y]$ と標準偏差 $\sigma[Y]$ とを求めよ。

1.3節で述べたように、データの分散はデータの2乗の平均から平均の2乗を引いたものである。これと同様のことが確率変数の分散についても成り立つ。

[定理 2.4.4] 確率変数 X の平均値 $E[X]$ と確率変数 X^2 の平均値 $E[X^2]$ とがあるならば、 X の分散 $V[X]$ は

$$V[X] = E[X^2] - E[X]^2 .$$

確率変数について**標準化**という考えが重要になる。

[定義] 確率変数 X の平均値 $E[X]$ と標準偏差 $\sigma[X]$ とがあり $\sigma[X] > 0$ であるとき、 X を標準化した確率変数とは次の確率変数のことである：

$$\frac{X - E[X]}{\sigma[X]} .$$

[定理 2.4.5] 確率変数 X の平均値 $\mu = E[X]$ と標準偏差 $\sigma = \sigma[X]$ とがあり $\sigma > 0$ であるとき、 X を標準化した確率変数 $\frac{X - \mu}{\sigma}$ の平均値は0で標準偏差は1である：

$$E\left[\frac{X - \mu}{\sigma}\right] = 0 , \quad \sigma\left[\frac{X - \mu}{\sigma}\right] = 1 .$$

[証明] $E[X]$ と $\sigma[X]$ とは定数なので、定理 2.3.2 と定理 2.3.1 とにより

$$E\left[\frac{X - \mu}{\sigma}\right] = \frac{E[X - \mu]}{\sigma} = \frac{E[X] - \mu}{\sigma} = 0 .$$

$E[X]$ と $\sigma[X]$ とは定数なので、定理 2.4.3 と定理 2.4.2 とにより、

$$\sigma\left[\frac{X - \mu}{\sigma}\right] = \frac{\sigma[X - \mu]}{\sigma} = \frac{\sigma[X]}{\sigma} = 1 .$$

証明終了

標準化とは平均が0で標準偏差が1である分布に換算することである。