§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と標準偏差とがありました．データの要素が

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

の n 個であるとき，その平均 $\bar{x} = \dfrac{1}{n}\sum\limits_{k=1}^{\infty} x_k$ に対して，データの分散 $\dfrac{1}{n}\sum\limits_{k=1}^{n}(x_k - \bar{x})^2$ は，

$$(x_1-\bar{x})^2, (x_2-\bar{x})^2, (x_3-\bar{x})^2, \dots, (x_n-\bar{x})^2$$

の平均です． 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 乗の平均です． 同じように，確率変数の分散とは，確率変数の値と平均値との差の 2 乗の平均値です．

定数 μ に対して，定理2.2より，確率変数 X の 2 次関数 $(X-\mu)^2$ は確率変数です．

定義　確率変数 X の平均値を μ とおく： $\mu = \mathrm{E}[X]$ ． X の分散 $\mathrm{V}[X]$ を次のように定義する：

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

更に，X の標準偏差 $\sigma[X]$ を次のように定義する：

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

例題 2.4.1　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． 成績の点数が 5 になる確率は 30% で，3 になる確率は 40% であり，1 になる確率は 20% であり，0 になる確率は 10% であるとする． 成績の点数を表す変数を X とおく． この変数 X は確率変数である． 確率変数 X の平均値 $\mathrm{E}[X]$ と分散 $\mathrm{V}[X]$ と標準偏差 $\sigma[X]$ とを求める．

確率変数 X の平均値 $\mathrm{E}[X]$ は，

$$\begin{aligned}\mathrm{E}[X] &= 5\cdot \mathrm{P}(\,X=5\,) + 3\cdot \mathrm{P}(\,X=3\,) + 1\cdot \mathrm{P}(\,X=1\,) + 0\cdot \mathrm{P}(\,X=0\,)\\ &= 5\cdot\frac{3}{10} + 3\cdot\frac{4}{10} + 1\cdot\frac{2}{10} + 0\cdot\frac{1}{10} = \frac{15+12+2}{10}\\ &= \frac{29}{10}\ .\end{aligned}$$

確率変数 X の分散 $\mathrm{V}[X]$ は，

$$\begin{aligned}\mathrm{V}[X] &= (5-2.9)^2\,\mathrm{P}(\,X=5\,) + (3-2.9)^2\,\mathrm{P}(\,X=3\,)\\ &\qquad + (1-2.9)^2\,\mathrm{P}(\,X=1\,) + (0-2.9)^2\,\mathrm{P}(\,X=0\,)\\ &= 4.41\cdot 0.3 + 0.01\cdot 0.4 + 3.61\cdot 0.2 + 8.41\cdot 0.1\\ &= 1.323 + 0.004 + 0.722 + 0.841\\ &= 2.89\ .\end{aligned}$$

X の標準偏差 $\sigma[X]$ は，

$$\sigma[X] = \sqrt{\mathrm{V}[X]} = \sqrt{2.89}\ .$$

終

問題 2.4.1　成績が 0 点，1 点，2 点，3 点，4 点，の 5 段階で点数化されるものとします． 成績の点数が 4 になる確率は 10% で，3 になる確率は 20% で，2 になる確率は 40% で，1 になる確率は 20% で，0 になる確率は 10% であるとします． 成績の点数を表を表す変数を X とおきます． この変数 X は確率変数です． 確率変数 X の平均値 $\mathrm{E}[X]$ と分散 $\mathrm{V}[X]$ と標準偏差 $\sigma[X]$ とを求めなさい．

連続型確率変数の分散を求める式は次のようになります． その証明は後の補遺に回します．

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

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

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

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

例題2.1.2で確かめたように，この関数 $f(x)$ は確率変数の確率密度関数になる． この関数 $f(x)$ を確率密度関数とする連続型確率変数 X の分散 $\mathrm{V}[X]$ を求める．

例題2.3.2において求めたように，確率変数 X の平均値は $\mathrm{E}[X]=3$ であった． X の分散 $\mathrm{V}[X]$ は，定理2.4.1より

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

$t=x-3$ とおく． $x=t+3$ ， $dx=dt$ ， $x=0$ のとき $t=-3$ ， $x=6$ のとき $t=3$ ．

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

故に $\mathrm{V}[X] = \dfrac{9}{5}$ ．

終

問題 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ます：

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

問題2.1.2で確かめたように，この関数 $f(x)$ は確率変数の確率密度関数になります． この関数 $f(x)$ を確率密度関数とする連続型確率変数 X の分散 $\mathrm{V}[X]$ を求めなさい．

以下の定理が成り立ちます． その証明は後の補遺に回します．

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

$$\mathrm{V}[X+c] = \mathrm{V}[X]\ , \qquad \sigma[X+c] = \sigma[X]\ .$$

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

$$\mathrm{V}[kX] = k^2\,\mathrm{V}[X]\ , \qquad \sigma[kX] = |k|\,\sigma[X]\ .$$

例題 2.4.3　確率変数 X について $\mathrm{E}[X]=8$ ， $\mathrm{V}[X]=5$ とする． 変数 Y を $Y=\dfrac{7-2X}{3}$ とおく． 確率変数 Y の分散 $\mathrm{V}[Y]$ と標準偏差 $\sigma[Y]$ とを求める．

確率変数 Y の分散 $\mathrm{V}[Y]$ は，

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

確率変数 Y の標準偏差 $\sigma[Y]$ は，

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

終

問題 2.4.3　確率変数 X について $\mathrm{E}[X]=7$ ， $\mathrm{V}[X]=3$ とします． 変数 Y を $Y=\dfrac{9-4X}{5}$ とおきます． 確率変数 Y の分散 $\mathrm{V}[Y]$ と標準偏差 $\sigma[Y]$ とを求めなさい．

データの分散について次のことが成り立ちました（定理1.3）：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ちます． その証明は後の補遺に回します．

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

$$\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2\ .$$

確率変数について**標準化**という考えが重要になります．

定義　確率変数 X の平均値 $\mathrm{E}[X]$ と標準偏差 $\sigma[X]$ とがあるとき，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：

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

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

$$\mathrm{E}\Big[\frac{X-\mu}{\sigma}\Big] = 0\ , \qquad \sigma\Big[\frac{X-\mu}{\sigma}\Big] = 1\ .$$

証明　$\mathrm{E}[X]$ と $\sigma[X]$ とは定数なので，定理2.3.2と定理2.3.1とより

$$\mathrm{E}\Big[\frac{X-\mu}{\sigma}\Big] = \frac{\mathrm{E}[X-\mu]}{\sigma} = \frac{\mathrm{E}[X]-\mu}{\sigma} = 0\ .$$

$\mathrm{E}[X]$ と $\sigma[X]$ とは定数なので，定理2.4.3と定理2.4.2とより

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

（証明終り）