

2 平均値と分散

2.1 平均値／期待値

2.1.1 有限データの場合

定義 2.1.1 有限データ X に含まれる数値の総和をデータのサイズで割った値を、そのデータの平均値 (mean value) と言い記号 $E[X]$ で表します。

数の羅列 10, 10, 8, 7, 10, 8, 8, 6, 10, 9, 8, 9

$$\begin{aligned} (\text{平均値}) &= \frac{10 + 10 + 8 + 7 + 10 + 8 + 8 + 6 + 10 + 9 + 8 + 9}{12} = \frac{103}{12} = 8.58333\dots \\ &= \frac{10 \cdot 4 + 9 \cdot 2 + 8 \cdot 4 + 7 \cdot 1 + 6 \cdot 1}{12} \\ &= 10 \times \frac{4}{12} + 9 \times \frac{2}{12} + 8 \times \frac{4}{12} + 7 \times \frac{1}{12} + 6 \times \frac{1}{12} \end{aligned}$$

$$(\text{平均値}) = \sum (\text{数値}) \times (\text{相対度数}) = \sum (\text{数値}) \times (\text{確率})$$

2.1.2 可算無限データの場合

定義 2.1.2 可算無限種類の数値からなるデータ X に対して、各数値にその数値の相対度数を掛けた値の総和が有限値として存在する場合、その総和を平均値と言い記号 $E[X]$ で表します。

n	1	2	3	...	n	...
$P[X = n]$	$\frac{1}{6}$	$\frac{5}{6} \frac{1}{6}$	$\left(\frac{5}{6}\right)^2 \frac{1}{6}$	...	$\left(\frac{5}{6}\right)^{n-1} \frac{1}{6}$	...

$$\frac{1}{6} \left\{ 1 + 2 \left(\frac{5}{6}\right) + 3 \left(\frac{5}{6}\right)^2 + \dots \right\} = 6$$

数	2	4	8	16	32	...	2^n	...
相対度数	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	...	$\frac{1}{2^n}$	...

$$2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + \dots = +\infty$$

2.1.3 非可算無限データの場合

定義 2.1.3 密度関数が $f(x)$ であるデータ／確率変数 X に対して次の積分が有限値として存在する時、これを X の平均値 (期待値) と言い記号 $E[X]$ で表します：

$$\underbrace{\int_{-\infty}^{\infty}}_{\text{全部足す}} \underbrace{x}_{\text{数値}} \underbrace{f(x)dx}_{\text{相対度数}}$$

問題 2.1.4 確率変数 X の分布密度関数が次の $h(x)$ で与えられているとき、確率 $P[-0.5 \leq X \leq 1.5]$ 、平均値 $E[X]$ を求めて下さい。

$$h(x) = \begin{cases} \frac{3}{4}(1-x^2) & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

問題 2.1.5 確率変数 X の分布密度関数が $h(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ であるとき、期待値 $E[X]$ を求めて下さい。

2.2 データの変換、その密度と期待値

データ X に含まれるそれぞれの数値 x に対して $F(x)$ と云う新たな数値を考え、この $F(x)$ を全て集めて1つの新たなデータ $F(X)$ として考える事があります。

2.2.1 1次関数

$$E[aX + b] = aE[X] + b$$

2.2.2 2次関数

X の密度を $h(x)$ とすると

$$E[(X - m)^2] = \int_{-\infty}^{\infty} (z - m)^2 h(z) dz$$

2.2.3 分散

定義 2.2.1 データ X が平均値 $E[X]$ をもち、更に $E[(X - E[X])^2]$ も存在する場合これをデータ X の分散 (variance) と言って記号 $Var[X]$ で表します。 X の密度関数が $h(x)$ であればこれは

$$Var[X] = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 h(x) dx$$

です。また、分散の正の平方根を標準偏差 (standard deviation) と言います。

$$Var[X] = E[X^2] - E[X]^2$$

事実 2.2.2 一般に、密度関数 $h(x)$ をもつデータ X と関数 $F(x)$ に対して積分 $\int_{-\infty}^{\infty} F(x)h(x)dx$ が有限値として存在するとき、この積分は $E[F(X)]$ に等しくなります。

2.3 問題演習

基本演習 2.1 次のデータ X について、 $E[X]$ 、 $E[X - 3]$ 、 $E[(X - 3)^2]$ を計算して下さい：

数値	7	6	5	4	3	2	1	0
相対度数	$\frac{2}{12}$	$\frac{1}{12}$	0	0	$\frac{3}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{1}{12}$

基本演習 2.2 X の密度関数が次の $f(x)$ で与えられているとき定数 a の値を求め、平均値 $E[X]$ 、分散 $Var[X]$ を求めて下さい。

$$f(x) = \begin{cases} a(2-x) & 0 \leq x \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

基本演習 2.3 区間 $[-1, 2]$ 上の一様分布に従う確率変数 X 、つまり、密度関数が次の $h(x)$ であるような確率変数 X ：

$$h(x) = \begin{cases} \frac{1}{3} & -1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

について、確率 $P[1 < X < 5]$ と期待値 $E[X]$ 、分散 $Var[X]$ を求めて下さい。

基本演習 2.4 有限データの場合にも $Var[X] = E[X^2] - E[X]^2$ を証明して下さい。

基本演習 2.5 密度 $f(x)$ をもったデータ X の平均値が m 、分散が v であるとき、派生データ $-2X + 3$ の平均値、分散を m, v で表して下さい。

発展演習 2.6 X の密度関数は $h(x)$ で分散が存在するとし、任意の $\alpha > 0$ を1つとります。分散の計算式の積分：

$$Var[X] = \int_{-\infty}^{\infty} (x - E[X])^2 h(x) dx$$

において、積分範囲を $|x - E[X]| \geq \alpha$ である x の範囲と $|x - E[X]| < \alpha$ である x の範囲に分けて後者を捨てることによって Chebyshev の不等式：

$$P[|X - E[X]| \geq \alpha] \leq \frac{1}{\alpha^2} Var[X]$$

が成立する事を示して下さい。またこの不等式からどんな事が読み取れますか。