4 確率変数と確率分布

4.1 確率変数の定義

例 **4.1.** **(コイン投げ)** コインを 1 回投げ、表か裏かを観測するという試行を考える。結果を X で表し、

$$X = \begin{cases} 0 & (\text{裏}) \\ 1 & (\text{表}) \end{cases}$$

と定義すると、X の取り得る値の集合は $\{0,\ 1\}$ であり、

$$P(X = x) = 1/2\ (x = 0, 1). \tag{4.1}$$

例 **4.2.** **(コイン投げ 2)** コインを n 回投げ、表の出る回数を観測するという試行を考える。$\boldsymbol{X = n}$ **回中で表が出る回数** と定義すると、X の取り得る値の集合は $= \{0,\ 1,\ \cdots,\ n\}$ であり、

$$P(X = x) = {}_nC_x/2^n \quad (x = 0, 1, \cdots, n) \tag{4.2}$$

となる。表が出る確率が p の場合は,

$$P(X = x) = {}_nC_x p^x (1-p)^{n-x} \quad (x = 0, 1, \cdots, n) \tag{4.3}$$

となる（次回)。

定義 (上例のように、) 変数 X が次の条件

* X の取り得る値の集合（以後、**値域**と呼ぶ）が分かっている

* 各値に確率が付与されている

を満たすとき、X を**確率変数 (random variable, r.v.)** と言う。データを「確率変数の実現値」と定義する。確率変数には**離散型**と**連続型**がある。

- **離散型確率変数:** 確率変数 X の値域が、次のような形の集合

  $$\{x_1, x_2, \cdots, x_N\}\ (\text{要素の数が無限個でもよい})$$

  であるとき、X を**離散型確率変数**と言う (X は飛び飛びの値しか取らない)。X の確率変数としての性質は、

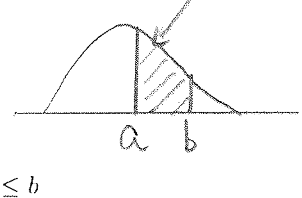
  $$P(X = x_k) = f(x_k) \quad (k = 1, 2, \cdots, N)$$

  によって定まる。これを X の**確率分布**と言う。

  $P(a \leq X \leq b) = \int_a^b f(x)\,dx$

- **連続型確率変数:** 確率変数 X の値域が、

  $$(-\infty, \infty)、[0, \infty)\ \text{などの区間} \tag{4.4}$$

  であり、かつある非負の関数 $f(x)$ が存在して

  $$P(a \leq X \leq b) = \int_a^b f(x)dx \quad \text{for all } a \leq b \tag{4.5}$$

  が成立するとき、X を**連続型確率変数**と言う。X の確率変数としての性質は、関数 $f(x)$ によって定まる。$f(x)$ を**確率密度関数** (probability density function, pdf) と言う。

例 4.3. (連続型分布の練習) あるガソリンスタンドは毎週月曜日の朝にガソリンの補給を受ける。このスタンドの週当りの需要量を X(千 ℓ) とする。過去の経験から、X の確率密度関数は

$$f(x) = \begin{cases} \frac{3}{125}(5-x)^2 & (0 \le x \le 5) \\ 0 & (\text{その他}) \end{cases} \tag{4.6}$$

であることが分かっている。

(a) ある週の需要量が 2000ℓ 未満である確率を求めよ。(b) タンクの容量が 4000ℓ であるとする。このスタンドがガソリンの需要を満たせない確率を求めよ。(c) 平均と分散を求めよ。

4.2 コイン投げ

例 4.2 の続きを議論する。(4.2) 式の確率変数 X について次の事実が成り立つ。

定理 4.1. $E(X) = n/2$、 $V(X) = n/2^2$ である。

(証明) 次式を計算すればよいことを納得しよう。まずはそれだけでも十分である。

$$E(X) = \sum_{x=0}^{n} x \frac{{}_nC_x}{2^n} = \frac{1}{2^n}\sum_{x=0}^{n} x \, {}_nC_x \tag{4.7}$$

$$= \frac{1}{2^n} \times \left\{ 0 \times {}_nC_0 + 1 \times {}_nC_1 + 2 \times {}_nC_2 + \cdots + n \times {}_nC_n \right\} \tag{4.8}$$

証明には、２項定理を用いる：

定理 4.2. (２項定理)

(1) A、B を任意の実数、m を正の整数とすると、次式が成立する。

$$(A+B)^m = {}_mC_0 B^m + {}_mC_1 AB^{m-1} + {}_mC_2 A^2 B^{m-2} + \cdots + {}_mC_m A^m = \sum_{k=0}^{m} {}_mC_k A^k B^{m-k}$$

(2) 上式で、$A = B = 1$ として、$2^m = \sum_{k=0}^{m} {}_mC_k = {}_mC_0 + {}_mC_1 + {}_mC_2 + \cdots + {}_mC_m$ が成り立つ。//

２項定理を用いて、$E(X) = n/2$ を示してみよう。

$$E(X) = \frac{1}{2^n}\sum_{x=0}^{n} x \, {}_nC_x \tag{4.9}$$

$$= \frac{1}{2^n}\sum_{x=0}^{n} x \frac{n!}{x!\,(n-x)!} \quad (\text{２項係数の定義}) \tag{4.10}$$

$$= \frac{1}{2^n}\sum_{x=1}^{n} x \frac{n!}{x!\,(n-x)!} \quad (x = 0 \text{ の項は明らかに } 0) \tag{4.11}$$

$$= \frac{1}{2^n}\sum_{x=1}^{n} \frac{n!}{(x-1)!\,(n-x)!} \quad \left(\frac{x}{x!} = \frac{1}{(x-1)!}\right) \tag{4.12}$$

$$= \frac{1}{2^n}\sum_{x=1}^{n} \frac{n \times (n-1)!}{(x-1)!\,[(n-1)-(x-1)]!} \quad (n! = n \times (n-1)!) \tag{4.13}$$

$$= \frac{n}{2^n}\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,[(n-1)-(x-1)]!} \quad (n \text{ を外へ出した}) \tag{4.14}$$

ここで、$m=n-1$ と $k=x-1$ という置き換えを行うと

$$= \frac{n}{2^n}\sum_{k=0}^{m}\frac{m!}{k!(m-k)!} = \frac{n}{2^n}\sum_{k=0}^{m} {}_mC_k \qquad (m=n-1,\ k=x-1 \text{ と置き換え}) \tag{4.15}$$

$$= \frac{n}{2^n}\times 2^m \quad (2\text{ 項定理の }(2)) \tag{4.16}$$

$$= \frac{n2^{n-1}}{2^n} = \frac{n}{2} \tag{4.17}$$

分散については次回の 2 項分布のところで学ぶ。

$$\begin{aligned} V(X) &= \sum_{x=0}^{n}\left(x-\frac{n}{2}\right)^2 \frac{{}_nC_x}{2^n} = \frac{1}{2^n}\sum_{x=0}^{n}\left(x-\frac{n}{2}\right)^2 {}_nC_x \\ &= \frac{1}{2^n}\left\{\left(0-\frac{n}{2}\right)^2 \times_n C_0 + \left(1-\frac{n}{2}\right)^2 \times_n C_1 + \cdots + \left(n-\frac{n}{2}\right)^2 \times_n C_n\right\} \end{aligned}$$

一見して、分散の計算が厄介であることが分かるであろう。次の公式

$$V(X) = E(X^2) - \{E(X)\}^2, \quad V(X) = E\{X(X-1)\} + E(X) - \{E(X)\}^2$$

を用いた方が容易である (次回)。

4.3 期待値と分散の練習

例 4.4. (宝くじ) 次表のような宝くじを考える。発行枚数は 10000 枚とする。

当選金 (円)	0	100	1000	10000
枚数	9000	800	150	50

この宝くじを 1 枚購入したときの当選金を X とおけば、抽選前の X は次の分布を持つ離散型確率変数とみなせる。

x_k	0	100	1000	10000
$P(X=x_k)$	0.9	0.08	0.015	0.005

X の期待値 $\mathbf{E}(X)$ は定義により、

$$\mathbf{E}(X) = 0\times 0.9 + 100\times 0.08 + 1000\times 0.015 + 10000\times 0.005 = 73(\text{円}) \tag{4.18}$$

となる。期待値の定義に基づいて、$\mathbf{E}[0.8X+2]$ を求める。$g(X)=0.8X+2$ とおくと、定義より

$$\mathbf{E}[0.8X+2] = \mathbf{E}[g(X)] \tag{4.19}$$

$$= g(0)\times P(X=0) + g(100)\times P(X=100) + g(1000)\times P(X=1000) + g(10000)\times P(X=10000)$$

$$= (0.8\times 0+2)\times 0.9 + (0.8\times 100+2)\times 0.08$$

$$+ \quad (0.8\times 1000+2)\times 0.015 + (0.8\times 10000+2)\times 0.005 \tag{4.20}$$

$$= \quad 60.4 \tag{4.21}$$

となる. 期待値の線形性を用いることにより、期待値 $\mathbf{E}[0.8X+2]$ の計算はずっと簡単であり、

$$\mathbf{E}[0.8X+2] = 0.8\mathbf{E}(X)+2 = 0.8\times 73+2 = 60.4. \tag{4.22}$$

と求めればよい。分散を計算する。$\mathbf{E}(X)=73$ であったから、定義により、

$$\begin{aligned}\mathbf{V}(X) &= (0-73)^2\times 0.9+(100-73)^2\times 0.08+(1000-73)^2\times 0.015\\ &\quad +(10000-73)^2\times 0.005 = 510471.0\end{aligned} \tag{4.23}$$

となる。$\mathbf{D}(X)=\sqrt{510471.0}=714.5$ である。□

例 4.5. (期待値の練習) ある講義の履修者数 X の確率分布が次の通りであったとする。

x_k	180	190	200	210
$P(X=x_k)$	1/8	1/4	1/2	1/8

教科書が余れば1部当り200円の損失があり、不足すれば当座のコピー代として学生一人当たり400円の費用がかかる。教師は用意する教科書の数を190、200、210の中から選ばなければならない。どれを選ぶべきか。

例 4.6. **(公式の練習)** ある駅の売店における一日当りの新聞の販売部数 X の平均は $E(X)=100$ 部、標準偏差は $D(X)=10$ 部である。1部当りの利益が20円、固定的にかかる費用が1日当たり500円とするとき、1日当りの利益の期待値と標準偏差を求めよ。

4.4 Chebyshev の不等式と応用例

定理 4.3. (Chebyshev の不等式) 確率変数 X の平均を μ、分散を σ^2 で表す：

$$E(X)=\mu,\quad V(X)=\sigma^2.$$

このとき、任意の $k\geq 0$ に対して次の不等式が成り立つ。

$$P\Big(\mu-k\sigma\leq X\leq \mu+k\sigma\Big)\geq 1-\frac{1}{k^2}. \tag{4.24}$$

特に、$k=2,3$ として、

$$P\Big(\mu-2\sigma\leq X\leq \mu+2\sigma\Big)\geq \frac{3}{4}=0.75 \tag{4.25}$$

$$P\Big(\mu-3\sigma\leq X\leq \mu+3\sigma\Big)\geq \frac{8}{9}\approx 0.889 \tag{4.26}$$

が成立する。

例 4.7. (コイン投げ) 表が出る確率が0.5であるようなコインを10000回投げる。表の出る回数を X とおけば、X の確率分布は

$$P(X=x)=f(x)=\frac{{}_{10000}C_x}{2^{10000}}\quad (x=0,1,\cdots,10000)$$

であり、

$$\mu=E(X)=10000/2=5000,\ \sigma^2=V(X)=10000/4=2500,\ \sigma=50$$

である。Chebyshev の不等式を用いて、$P(4800\leq X\leq 5200)$ を求めてみよう。

$$\begin{aligned}P\Big(4800\leq X\leq 5200\Big) &= P\Big(5000-4\times 50\leq X\leq 5000+4\times 50\Big)\\ &= P\Big(\mu-4\sigma\leq X\leq \mu+4\sigma\Big)\\ &\geq 1-\frac{1}{4^2}=\frac{15}{16}\approx 0.9375\end{aligned} \tag{4.27}$$

実際にはこの確率は0.99994であるから (後日)、Chebyshev の不等式はかなり粗いものであることが分かる。しかし、この不等式はどのような確率分布に対しても成立するという長所がある。//