第 3 章

確率分布

確率分布には，確率変数が離散的な場合と連続的な場合がある．ここではもっぱら離散的な場合に限定して解説し，連続的な場合については次章以降で導入する．

3.1 確率変数と確率関数

2 個のコインを投げて，その裏表がどうなるかを考えよう．2 個のコインの区別はしないものとし，表を A, 裏を B として，起こりうる事象 E_1, E_2, E_3 を次のように決める．

E_1 : 2 枚とも A
E_2 : 1 枚が A, もう 1 枚が B
E_3 : 2 枚とも B

の 3 通りであり，それぞれの確率は $1/4, 2/4, 1/4$，すなわち

$$
\begin{aligned}
&P(E_1) = 1/4 \\
&P(E_2) = 2/4 \\
&P(E_3) = 1/4
\end{aligned}
\tag{3.1}
$$

と書くことができる．現実にコインを投げて試行を重ねると，その結果は，この確率に比例した割合に徐々に近づいていくであろう．

ここで上の表し方を一歩進めて，事象 E_i の代わりに何らかの数値を使うことにすると，取り扱いが便利になる．たとえば 1 回の試行で B が現れる回数を X とすれば，

$$
\begin{aligned}
&P(X = 0) = 1/4 \\
&P(X = 1) = 2/4 \\
&P(X = 2) = 1/4
\end{aligned}
\tag{3.2}
$$

と書けるであろう．

このように，事象を数値で代表させるようにした場合の変数を**確率変数 (stochastic variable)** といい，大文字の X, Y などでしばしば表される．X は，具体的には数値 $x_1, x_2, \ldots$ という値をとる．このケースのように確率変数が飛び飛びの値をとる場合は，とくに**離散的確率変数 (discrete probability variable)** という．いくつか例を挙げてみよう．

- コインの表が出たら 1，裏が出たら 0 と確率変数を決めておくと，可能な値は 0 と 1
- サイコロを 1 回投げる試行では，確率変数はサイコロの目の数で，可能な値は 1 から 6 までの整数
- ランダムに選んできた 20 人の集団の中で特定の血液型をもつ人数を考えるとき，確率変数はその人数，可能な値は 0 から 20 までの整数
- 3000 人を対象の内閣支持率の調査では，確率変数は支持する人の数，可能な値は 0 から 3000 までの整数
- 宝くじでは，確率変数は当選金額，可能な値は末等から一等までの金額.

$X = x_1, x_2, \ldots$ に対する確率は，**確率関数 (probabilty function)** または**確率密度 (probability density)** と呼ばれる．確率関数はしばしば $f(x_i)$ のように表される．

$$P(X = x_i) = f(x_i)$$

この表記では，上の式 (3.2) の関係は

$$\begin{aligned} f(0) &= 1/4 \\ f(1) &= 2/4 \\ f(2) &= 1/4 \end{aligned} \tag{3.3}$$

となる[*1].

また，度数分布における累積度数に相当する関数は，しばしば $F(X)$ で表され，**分布関数 (distribution function)** と呼ばれる[*2]．ここの例では，

$$\begin{aligned} F(0) &= 1/4 \\ F(1) &= 3/4 \\ F(2) &= 4/4 \end{aligned} \tag{3.4}$$

[*1] ここでは離散的な確率変数を取り上げており，その場合については，$P(X = x)$ は $f(x)$ と同じものであるから，ここの定義は単なる言い換えに過ぎない．しかし後にみるように，連続的な確率変数の場合には，変数がある有限の範囲にある確率が定義されるので，ある値 x に対して $P(X = x) = f(x)$ という関係は成り立たない.

[*2] ここではいくつかの教科書を参照して，確率関数と分布関数という呼び方を紹介しているが，実際には確率関数という呼称はそれほど一般的ではなく，確率分布ということが多いようだ.

となる．

例題 3–1 2 個のサイコロを振る．このときの確率変数としては何を用いるのが適当か．またその値はどのような範囲をとるか．

2 個のサイコロの目の数の和を確率変数とするのが自然である．そのとき確率変数の値は 2 から 12 までの範囲をとる．

なお，この答えは唯一ではない．たとえば目の数の積とか，差でもかまわない．ただしそれぞれに応じて変数の範囲と確率分布は異なる．

3.2 離散的な確率関数の例 — 離散型一様分布

離散的な確率関数で最も基本的なものは，**離散型一様分布 (discrete uniform distribution)** と呼ばれるものである．

コインの投げ上げやサイコロを転がしたとき，起こり得る事象はどれも等しい確率をもつということにしよう．コイン投げであれば表と裏がどちらも 1/2 の確率で，サイコロであればどの目も 1/6 の確率で出るということにしてしまう．

サイコロの場合，目の数を確率変数とすると，

$$\begin{aligned} f(1) &= 1/6 \\ f(2) &= 1/6 \\ &\cdots \\ f(6) &= 1/6 \end{aligned} \tag{3.5}$$

となる．このようにどれも等しい確率で実現するような確率関数を**離散型一様分布**という．一般的には，根元事象の数を n として，確率関数が

$$f(x) = \begin{cases} 1/n & (x = 1, 2, \ldots, n) \\ 0 & (\text{それ以外}) \end{cases}$$

で表されるのが離散型一様分布である．

3.3 離散的な確率変数の性質

離散的な確率変数が，

$$X = x_1, x_2, \ldots, x_n$$

のように n 個の x_i の集合ですべての場合を尽くしているとしよう．すると，確率関数 $f(x)$, 分布関数 $F(x)$ について次の式が成り立つ．

$$\sum_{i=1}^{n} f(x_i) = f(x_1) + \ldots + f(x_n) = 1 \tag{3.6}$$

$$F(x_i) = \sum_{k=1}^{i} f(x_k) \tag{3.7}$$

$$F(-\infty) = 0 \tag{3.8}$$

$$F(\infty) = 1 \tag{3.9}$$

$F(x)$ は，確率（正かゼロの値しかとらない）の足し合わせで定義されているので，減少しない関数であることにも注意しておこう．

例題 3–2　コインを 3 回投げる試行を考える．表が出たコインの数を確率変数 X とする．X のとり得る値と，それらに対する確率を求めて，確率関数を書き下ろしなさい．

$X = 0, 1, 2, 3$ であり，それらに対する確率は，$1/8, 3/8, 3/8, 1/8$ であるから，確率関数は

$$\begin{aligned} f(0) &= 1/8 \\ f(1) &= 3/8 \\ f(2) &= 3/8 \\ f(3) &= 1/8 \end{aligned}$$

となり，また，分布関数は

$$\begin{aligned} F(0) &= 1/8 \\ F(1) &= 4/8 \\ F(2) &= 7/8 \\ F(3) &= 1 \end{aligned}$$

と書ける．

3.4　離散的確率変数の期待値と分散

多数の試行の後で，確率変数 X の平均がどの値に収束するかというのが，**平均 (mean, average)** または**期待値 (expectation value)** [*3]である．期待値は $E[X]$ または μ で表され，次のように与えられる．

$$
\begin{aligned}
E[X] = \mu &= \sum_{i=1}^{n} x_i f(x_i) \\
&= x_1 f(x_1) + \ldots + x_n f(x_n)
\end{aligned}
\tag{3.10}
$$

この形は 1 章 p.17 の式 (1.9) とよく似ていることに注意しよう．

また，確率変数 X の**分散**は $V[X]$ または σ^2 で表され，次の式で与えられる．

$$
V[X] = \sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) \tag{3.11}
$$

この形を見ると，分散というのは偏差の 2 乗の期待値であるとも言える．なお 1.1.4 節の式 (1.6) にならって分散の式を変形でき，

$$
V[X] = E[X^2] - E[X]^2 \tag{3.12}
$$

ここでも**分散 ＝ 2 乗の平均 − 平均の 2 乗** となる．

例題 3–3　ある賭けで，確率 0.1, 0.05, 0.01 でそれぞれ 100 円, 200 円, 1000 円儲けることができ，また確率 0.001 で 2000 円損をするものとする．これは離散的確率分布の例であるが，確率変数はどのような値をとるか．また，儲ける金額の期待値を求めよ．この賭けによってあなたは儲けるだろうか．

2000 円の損は −2000 円の儲けと考える．また 0 円の儲けの確率は $1 - 0.1 - 0.05 - 0.01 - 0.001 = 0.839$ である．従って儲けの額は $-2000, 0, 100, 200, 1000$ であり，これらが確率変数の値 $x_1, x_2, \ldots, x_5$ に相当する．確率関数 $f(x)$ は $f(x_1) = 0.001, f(x_2) = 0.839, f(x_3) = 0.1, f(x_4) = 0.05, f(x_5) = 0.01$ である．これから期待値を求めるには，

[*3] 平均と期待値は概念上はまったく異なるものである．すなわち本来，平均というのはすでに存在する複数のデータから算出されるものであり，期待値というのはまだ実現していない事象がどうなるかについて，多数回の試行を行った結果を想定して，その平均を考えるのである．しかし，数学的には，同じように扱っても問題はない．

式 (3.10) を使えばよい．

$$\begin{aligned} E[X] &= f(x_1)x_1 + f(x_2)x_2 + \cdots \\ &= 0.001 \times (-2000) + 0.839 \times 0 + 0.1 \times 100 + \cdots 0.01 \times 1000 \\ &= 28 \end{aligned}$$

結局期待値は 28 円となるから，十分に多数回やっていれば儲かることになる．

例題 3–4　前節のコインを 3 回投げる試行の期待値と分散を求めよ．また，実際にコインを 3 回投げる試行を数十回行ってみて，結果を理論値と比較してみよ．

期待値は，

$$\mu = \frac{1}{8} \cdot 0 + \frac{3}{8} \cdot 1 + \frac{3}{8} \cdot 2 + \frac{1}{8} \cdot 3 = \frac{3}{2} = 1.5$$

すなわち，平均して 1.5 に収束する．次に分散を求めるには，

$$E[X^2] = \frac{1}{8} \cdot 0^2 + \frac{3}{8} \cdot 1^2 + \frac{3}{8} \cdot 2^2 + \frac{1}{8} \cdot 3^2 = 3$$

を求めておいて，

$$E[X^2] - \{E[X]\}^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4} = 0.75$$

となる．

例題 3–5　事象の数が n であるような離散型一様分布の平均値と分散を求めよ．

$$\mu = 1 \cdot \frac{1}{n} + 2 \cdot \frac{1}{n} + \ldots + n \cdot \frac{1}{n} = \frac{n(n+1)}{2}\frac{1}{n} = \frac{n+1}{2}$$

$$\sigma^2 = E[X^2] - \{E([X]\}^2 = 1^2 \cdot \frac{1}{n} + 2^2 \cdot \frac{1}{n} + \ldots + n^2 \cdot \frac{1}{n} - \frac{(n+1)^2}{2^2} = \frac{n^2-1}{12}$$

ここで高校数学の数列の問題で扱われる次の式を使った。

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

3.5　確率変数の関数の期待値と分散

確率変数の関数で表される量があったとする．元の確率変数の平均と分散がわかっているとして，その関数の平均と分散を知りたいことがある[*4]．また，2 つの確率変数が存在する場合に，それらの和などがどのように振舞うかを知りたいこともある．それらについて，基本的な関係式を提出しておく．

- 二つの確率変数の和の期待値は，それぞれの期待値の和になる．

$$E[X+Y]=E[X]+E[Y] \tag{3.13}$$

- 二つの確率変数の和および差の分散は，**それらが独立な時にだけ**，それぞれの分散の和になる．

$$V[X\pm Y]=V[X]+V[Y] \tag{3.14}$$

- 分散の基本的な性質[*5]

$$V[X]=E[X^2]-(E[X])^2 \tag{3.15}$$

- 確率変数の一次式の期待値

$$E[aX+b]=aE[X]+b \tag{3.16}$$

- 確率変数の一次式の分散には，定数項は現れない．

$$V[aX+b]=a^2V[X] \tag{3.17}$$

- 二つの確率変数の積の期待値は，**それらが独立な時にだけ**，積の形になる．

$$E[XY]=E[X]E[Y] \tag{3.18}$$

例題 3–6　駅伝の二つの区間を A, B の両選手がリレーするものとする．最初の区間を A 選手が走った実績は，平均 27.6 分，標準偏差が 1.5 分であり，次の区間を B 選手が

[*4] 抽象的な表現で分かりにくいが，次のようなケースを考えればよい．今，宝くじで当たる賞金の確率分布が知られているとしよう（これは宝くじの主催者なら分かることである）．税務署としては，賞金額（これは確率変数である）によって，税額を変えて徴収するわけであるから，税額が賞金額にどのように依存しているかという関数関係（これは税法で決まっている）を使って，徴収できる金の期待値をあらかじめ予測することに関心がある．つまり，確率変数（＝賞金額）の関数（＝税金）の期待値（平均してどれだけ取れるか）と分散（期待値からどのていどばらつくか）を考えるのが，ここの目的である．

[*5] この形に類似した関係式は，すでに何度か現れている．

走った実績は，平均 16.2 分，標準偏差が 1.2 分であるとする．二人の合計タイムの平均と標準偏差はどうなるか．また，その計算に必要とされる仮定はなにか．

式 (3.13) から，平均については単に足せばよい．標準偏差については，まず 2 乗して分散にしてから足し合わせて，その平方根をとる．つまり $\sqrt{1.5^2 + 1.2^2} = 1.92$ 分となる．これが成立するのは A, B 両選手の走りが互いに独立であり，相互に影響し合うことがないものとしたときにのみ成立する．標準偏差には単純な足し算が成立しないことに注意しよう．