第 5 章　確率変数と確率分布

1　確率の概念

確率の起源　17 世紀のフランス．ブルボン王朝期，ルイ 14 世の時代．貴族ド・メレのサイコロ投げに関する疑問に端を発する．

(Q1) 1 個のサイコロを 4 回投げて，少なくとも 1 回 1 の目が出る．

(Q2) 2 個のサイコロを同時に 24 回投げて，少なくとも 1 回 1 のペアが出る．

ド・メレの誤った計算：(Q1) については，1 回投げて 1 の目が出る可能性は 1/6 で，4 回投げるから 4×1/6=2/3 となる．(Q2) については，1 回投げて 1 のペアが出る可能性が 1/36 で，24 回投げるから，この場合にも 24×1/36=2/3 となってしまう．

パスカルとフェルマーの回答：

$$\text{(Q1)}\quad 1-\left(\frac{5}{6}\right)^4=0.518 \qquad\qquad \text{(Q2)}\quad 1-\left(\frac{35}{36}\right)^{24}=0.491$$

ギャンブラーの法則：起きる可能性が p のゲームを n 回試したとき，少なくとも 1 回成功する可能性は $1-(1-p)^n$ となる．このとき，

$$1-(1-p)^n>\frac{1}{2}$$

をみたす n が存在する．(Q1) のゲームでは，$n=4$ の場合に可能性がはじめて 1/2 を越える．(Q2) では，$n=25$ ではじめて 1/2 を越える．

等可能性原理 起こりうる結果が有限個で，それぞれが起きる可能性が同等に確からしい場合，問題としている事象の確率は，その場合の数をすべての場合の数で割ればよい，とする原理である．たとえば，サイコロ投げで偶数の目が出る確率は，偶数の目が出る場合の数 3 を，すべての場合の数 6 で割った値，すなわち，1/2 となる．

（例外は，将棋の駒を投げたときに底面となる面の確率を求めるような場合）

経験的確率 数多くの実験を繰り返すことにより求められる確率

主観的確率 主観的に付与される確率

公理的確率 起こりうる結果が有限個でない場合に，確率が具体的に計算できることや先験的に存在することを前提とせずに，公理として定義される確率

全事象（標本空間）	ある試行において起こりうる結果すべてからなる集まり
基本事象	全事象の個々の要素
事象	基本事象の集まり
積事象 $A\cap B$	事象 A と B が同時に起きる事象
和事象 $A\cup B$	事象 A と B の少なくとも一方が起きる事象
余事象 $\overline{A}$	事象 A が起こらないという事象
空事象 $\varnothing$	起こりえない事象

サイコロ投げの例

全事象：　$\Omega=(1,2,3,4,5,6)$

基本事象：　$\omega_1=(1)$，…，$\omega_6=(6)$

事象：　A=(1，2)， B=(2，4，6)，C=(3，4，5，6)，…

A は 2 以下の目が出る事象，B は偶数の目が出る事象，

C は 3 以上の目が出る事象

$$A\cap B=(2),\qquad A\cup B=(1,2,4,6),\qquad \overline{A}=C,\qquad A\cap C=\phi$$

ここで，事象 A と C は同時には起こりえず，積事象は空事象となっている．このとき，A と C は**互いに排反**である，という．

確率の公理　次の 3 つの条件（C1)，(C2)，(C3) をみたす関数 P を確率という．

(C1) 任意の事象 A に対して $0\leqq P(A)\leqq 1$

(C2) 全事象 Ω に対して $P(\Omega)=1$，空事象 φ に対して $P(\phi)=0$

(C3) A と B が互いに排反ならば $P(A\cup B)=P(A)+P(B)$

確率の公理から，確率がもつべきさまざまな性質，例えば次のような性質を導くことができる．

(a) $P(\overline{A})=1-P(A)$

(b) 事象 A が起きれば必ず事象 B が起きる，すなわち，$A\subset B$ ならば $P(A)\leqq P(B)$

(c) 任意の事象 A，B に対して $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

(d) A_1，A_2，… A_n が互いに排反ならば

$$P(A_1\cup A_2\cup\cdots\cup A_n)=P(A_1)+P(A_2)+\cdots+P(A_n)$$

2　確率変数と確率分布

$\boxed{\text{確率変数}}$ どの値が実現するかについては断定できないが，とりうる値の全体がわかっており，それぞれの値の実現可能性が確率で表現できるような数値変数である．数学的には，全事象で定義され，実数値をとる関数

$$X:\ \omega\in\Omega\ \rightarrow\ X(\omega)\in(-\infty,\ \infty)$$

のことをいう．**離散的確率変数**と**連続的確率変数**に分類される．

なぜ確率変数を考えるか？　事象は一般に事柄であり，事柄を扱うのは数学的な操作性に欠けるので，数値に変換して操作性を高めるため．

$\boxed{\text{確率分布}}$

離散的確率変数の場合 離散的確率変数 X が x という値をとる確率を $P(X=x)$ で表すとき，各 x と確率 $P(X=x)$ との対応を**確率分布**という．確率分布は，母集団におけるすべてのメンバーの分布と考えることができるので，**母集団分布**とも呼ばれる．

確率分布の性質：離散的確率変数の場合

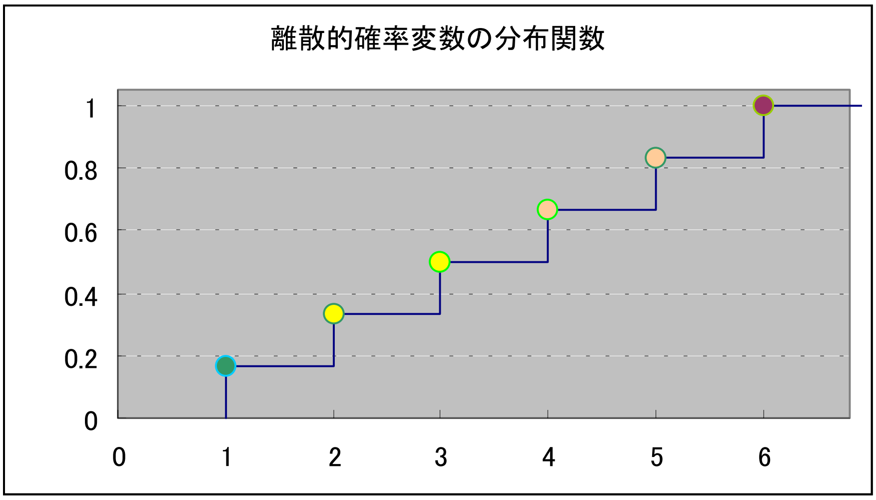
(a) すべての x に対して $P(X=x)\geq 0$

(b) $\sum_x P(X=x)=1$

確率分布が定義されれば，さまざまな確率を計算することができる．特に，任意の実数 x に対して，離散的確率変数 X が x 以下となる確率

$$P(X\leq x)=\sum_{y\leq x}P(X=y) \tag{1}$$

を x の関数とみなして**分布関数**と呼ぶ．分布関数は，離散的データの場合の累積相対度数折れ線と同様に，0 から 1 へ階段状に増加する関数となる．ただし，ジャンプする点においてはジャンプ後の値をとる．次の図は，ゆがみのないサイコロの目の確率分布の分布関数である．



分布関数による確率計算

確率分布は，X のとりうる値を $x_1<x_2<\cdots<x_{k-1}<x_k<\cdots$ とするとき，

$$P(X=x_k)=P(X\leq x_k)-P(X\leq x_{k-1}) \tag{2}$$

により求めることができる．

次の例題では，暗黙のうちに「独立性」の概念を使っているが，これについては第 5 節で説明する．

〔**例題 5.1**〕 3 個のサイコロを投げて出る目の最大値と最小値の確率分布を求めよ．

（解）3 個のサイコロの出る目を，それぞれ X，Y，Z，目の最大値を表す確率変数を $X_{\max}$ とする．このとき，次のことが成り立つ．

$$X_{\max}\leq k \quad\Leftrightarrow\quad X\leq k,\quad Y\leq k,\quad Z\leq k$$

このことから，$k=1,2,\cdots,6$ に対して，

$$P(X_{\max}\leq k)=P(X\leq k,Y\leq k,Z\leq k)=P(X\leq k)P(Y\leq k)P(Z\leq k)$$
$$=(k/6)^3$$

となる．したがって，(2) を使って，

$$P(X_{\max}=k)=P(X_{\max}\leq k)-P(X_{\max}\leq k-1)$$
$$=(k/6)^3-((k-1)/6)^3=(3k^2-3k+1)/216$$

を得る．他方，目の最小値を表す確率変数を $X_{\min}$ とすると，

$$X_{\min} > k \quad \Leftrightarrow \quad X > k,\quad Y > k,\quad Z > k$$

が成り立つから，

$$P(X_{\min} \le k) = 1 - P(X_{\min} > k) = 1 - P(X > k)P(Y > k)P(Z > k)$$

$$= \left(1 - \frac{k}{6}\right)^3$$

となる．したがって，

$$P(X_{\min} = k) = P(X_{\min} \le k) - P(X_{\min} \le k-1)$$

$$= (1-(k-1)/6)^3 - (1-k/6)^3 = (3k^2 - 39k + 127)/216$$

を得る（**図 5-4** 95 ページ参照）．具体的には，次の確率分布をもつ．

k	1	2	3	4	5	6
$P(X_{\max} = k)$	0.0046	0.0324	0.0880	0.1713	0.2824	0.4213
$P(X_{\min} = k)$	0.4213	0.2824	0.1713	0.0880	0.0324	0.0046

明らかに，$P(X_{\min} = k) = P(X_{\max} = 7 - k)$ が成り立つ．すなわち，$X_{\min}$ の確率分布は，$7 - X_{\max}$ の確率分布に等しい．このとき，$X_{\min}$ の分布は $X_{\max}$ の分布の**鏡像**（**ミラー・イメージ**）であるという．

〔**例題** 5.2〕 1 の目が出るまでサイコロを投げ続けるとき，投げる回数の確率分布を求めよ．

（解）投げる回数を X とするとき，$X = k$ となる確率は，最初の $k-1$ 回は 1 の目以外が出て，最後に 1 の目が出る確率である．したがって，X の確率分布は，

$$P(X = k) = \left(\frac{5}{6}\right)^{k-1}\frac{1}{6} \qquad (k = 1,\ 2,\ \cdots)$$

となる．これが確率分布となっていることは，等比数列の和の公式

$$\sum_{k=1}^{\infty} x^k = \frac{x}{1-x} \qquad |x| < 1$$

から明らかである．

連続的確率変数の場合　連続的確率変数の確率分布は**密度関数**と呼ばれ，離散的な場合と異なり連続的な関数となる．連続的確率変数は特定の値をとる確率が 0 であるから，とりうる値の各々に対して確率を考えることは無意味であり，密度関数そのものも特定の値をとる確率を与えるわけではない．

密度関数の定義　$f(x) = \lim_{h \to 0} \dfrac{P(x < X < x+h)}{h}$

意味合い　h が微小のとき，$P(x < X < x+h) \cong h \times f(x)$

分布関数の計算　$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt \qquad F'(x) = f(x)$

確率の計算　$P(a < X < b) = \int_a^b f(t)dt$

密度関数の性質

(a) すべての x に対して $f(x) \geq 0$

(b) $\int_{-\infty}^{\infty} f(x)dx = 1$

〔**例題 5.3**〕 0.5 以上 6.5 以下の区間で一様に実現する確率変数を 3 個取り出したとき，最大値と最小値の密度関数を求めよ.

（解）例題 5.1 と同様にして，3 個の確率変数を X，Y，Z，最大値を表す確率変数を $Y_{\max}$ とする．このとき，次のことが成り立つ.

$$Y_{\max} \leq x \quad \Leftrightarrow \quad X \leq x, \quad Y \leq x, \quad Z \leq x$$

さらに，X，Y，Z は，すべて同一の分布関数

$$F(x) = P(X \leq x) = \frac{x - 0.5}{6} \qquad (0.5 \leq x \leq 6.5)$$

をもつ．これらのことから，$0.5 \leq x \leq 6.5$ に対して，

$$P(Y_{\max} \leq x) = P(X \leq x, Y \leq x, Z \leq x) = \{F(x)\}^3$$
$$= \frac{(x-0.5)^3}{216}$$

となる．したがって，$Y_{\max}$ の密度関数 $f_{\max}(x)$ は，

$$f_{\max}(x) = \frac{d}{dx}\{F(x)\}^3 = \frac{(x-0.5)^2}{72} \qquad (0.5 \leq x \leq 6.5)$$

となる．他方，最小値 $Y_{\min}$ の密度関数 $f_{\min}(x)$ も，例題 5.1 と同様に考えれば，

$$f_{\min}(x) = \frac{(x-6.5)^2}{72} \qquad (0.5 \leq x \leq 6.5)$$

を得る．この場合も，$Y_{\min}$ の分布は $7 - Y_{\max}$ の分布と同一であり，$Y_{\max}$ の鏡像であることがわかる（$Y_{\max}$ の密度関数は**図 5-7** を参照）．

4　確率分布の特性値

$\boxed{\text{期待値}}$　X を離散的確率変数とするとき，

$$E(X) = \sum_x xP(X = x)$$

を X の**期待値**，あるいは X の**確率分布の平均**という．期待値は確率分布の重心である．期待値の記号としては μ（ミュー）を使うことが多い.

$\boxed{\text{分散}}$

離散的確率変数 X の**分散**を $V(X)$ で表し，

$$V(X) = \sum_x (x - E(X))^2 P(X = x)$$
$$= \sum_x x^2 P(X = x) - \{E(X)\}^2$$

で定義する.

$\boxed{\text{標準偏差}}$

$$SD(X)=\sqrt{V(X)}$$

で定義される．分散は σ^2（シグマ 2 乗），標準偏差は σ を使うことが多い．

データの平均や分散は，データ数が有限であるから常に存在するが，母集団の場合は，なかみが無限にあるならば存在しないこともある．平均が存在しない場合には，自動的に分散も存在しない．

〔**例題** 5.4〕 次の確率分布をもつ確率変数 X の期待値は存在しないことを示せ．

$$P(X=n)=\frac{6}{n^2\pi^2} \qquad (n=1,2,\cdots) \tag{8}$$

（解）まず，

$$\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\pi^2}{6}$$

が成り立つことが知られているので，(8) は確かに確率分布であることに注意する．このとき，

$$E(X)=\sum_{n=1}^{\infty}nP(X=n)=\frac{6}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n}$$

となり，最後の和は無限大に発散する．したがって，期待値は存在しない．

データの平均と標準偏差の間に成立したチェビシェフの不等式は，確率変数 X の場合にも，期待値 μ と分散 σ^2 が存在するならば次の形で成立する．

$$P\left(|X-\mu|\geq k\sigma\right)\leq\frac{1}{k^2} \tag{9}$$

〔**例題** 5.5〕 離散的確率変数に対して，チェビシェフの不等式 (9) を証明せよ．

（解）X の分散の定義式を次のように 2 つの和に分ける．

$$\sigma^2=\left\{\sum_{A(x)}+\sum_{B(x)}\right\}(x-\mu)^2P(X=x)$$

ここで，$A(x)$ は不等式 $|x-\mu|\geq k\sigma$ をみたす x の集まりである．他方，$B(x)$ は，この不等式をみたさない x の集まりである．したがって，

$$\begin{aligned}\sigma^2 &\geq \sum_{A(x)}(x-\mu)^2P(X=x)\geq k^2\sigma^2\sum_{A(x)}P(X=x)\\ &=k^2\sigma^2P(A(X))=k^2\sigma^2P\left(|X-\mu|\geq k\sigma\right)\end{aligned}$$

を得る．この不等式の最左辺と最右辺から，(9) の関係を導くことができる．

確率変数に関するチェビシェフの不等式は，母集団におけるすべての値の中で，平均から標準偏差の $\pm k$ 倍以上離れる割合は高々 $1/k^2$ である，ということをいっている．特に，母集団分布が単峰で対称と考えられるならば，平均から標準偏差の ±3 倍以上離れる確率は 0 に非常に近いのが普通である．

〔**例題 5.6**〕　ゆがみのないサイコロを振るとき，出る目の期待値および標準偏差を求めよ．また，期待値から標準偏差の ±2 倍以上離れる確率を求めよ．

（解）出る目を確率変数 X で表すと，期待値 μ は確率分布の重心であることから，即座に 3.5 を得る．分散は，(6) の最後の表現から

$$\sigma^2 = \sum_{x=1}^{6} x^2 P(X = x) - \mu^2$$
$$= \frac{1}{6}\sum_{x=1}^{6} x^2 - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

となる．したがって，$\sigma = \sqrt{35/12} = 1.71$ となる．このとき

$$P(|X-\mu| \geq 2\sigma) = P(|X-3.5| \geq 3.42) = 0$$

を得る．(9) のチェビシェフの不等式から得られる確率の上限は 1/4 であり，今の場合にはあまりにも保守的な値であることがわかる．

一般の確率変数の期待値　X が確率変数ならば，関数 $g(X)$ も確率変数であり，その期待値は

$$E(g(X)) = \sum_x g(x) P(X = x) \qquad (10)$$

により定義される．特に，$g(x) = a$（定数）ならば，

$$E(a) = \sum_x a \times P(X = x) = a \sum_x P(X = x) = a$$

となる．また，$g(x) = (x - E(X))^2$ とおくことにより，

$$E\left[(X - E(X))^2\right] = \sum_x (x - E(X))^2 P(X = x) = V(X)$$

が成立する．

〔**例題 5.7**〕　下の確率分布をもつ確率変数 X に対して，$E(X^2)$ の値を，(10) を使う方法と，$Y = X^2$ の確率分布に基づく方法の 2 通りで求めよ．

x	-2	-1	1	2	3
$P(X = x)$	1/5	1/5	1/5	1/5	1/5

（解）(10) を使う方法では，$g(x) = x^2$，$P(X = x) = 1/5$ であるから，

$$E(X^2) = \sum_x x^2 P(X = x) = \frac{1}{5}\left((-2)^2 + (-1)^2 + 1^2 + 2^2 + 3^2\right)$$
$$= \frac{19}{5}$$

を得る．他方，$Y = X^2$ のとりうる値は 1，4，9 であり，これらの値をとる確率を考えることにより，Y の確率分布は次のようになる．

y	1	4	9
$P(Y = y)$	2/5	2/5	1/5

したがって，

$$E(Y) = 1 \times \frac{2}{5} + 4 \times \frac{2}{5} + 9 \times \frac{1}{5} = \frac{19}{5}$$

となり，最初の方法と同一の解を得る．

期待値と分散の性質

(a) $E(aX+b)=aE(X)+b$

(b) $V(aX+b)=a^2V(X)$

(c) $V(X)=E\left[\left(X-E(X)\right)^2\right]=E\left(X^2\right)-\left(E(X)\right)^2$

５　多次元の確率変数と特性値

離散的な場合： n **次元確率分布** $P(X_1=x_1,\cdots,X_n=x_n)$

連続的な場合： n **次元密度関数** $f(x_1,\cdots,x_n)$

$$P(x_1<X_1<x_1+h_1,\cdots,x_n<X_n<x_n+h_n)\cong h_1\times\cdots\times h_n\times f(x_1,\cdots,x_n)$$

表 5-1　2 次元確率分布の例

		Y			
		１	２	３	
X	１	0	0.1	0.2	0.3
	２	0.1	0.2	0.4	0.7
		0.1	0.3	0.6	1.0

この例では，確率変数 X がとりうる値は 1，2 の 2 通り，Y がとりうる値は 1，2，3 の 3 通りである．そして，同時確率は，例えば P(X=1，Y=2)=0.1 のように与えられている．

周辺確率分布

$$P(X=x)=\sum_y P(X=x,Y=y) \quad ；表の最後の列$$

$$P(Y=y)=\sum_x P(X=x,Y=y) \quad ；表の最後の行$$

$E(X)=1.7$,　　　　$V(X)=0.21$

$E(Y)=2.5$,　　　　$V(Y)=0.45$

多次元確率変数の期待値

X と Y が確率変数であれば，実数値関数 $g(X,Y)$ は 1 次元の確率変数となり，その期待値は，

$$E\left(g(X,Y)\right)=\sum_x\sum_y g(x,y)P(X=x,Y=y)$$

で定義される．

〔**例題** 5.8〕 確率変数 X と Y の 2 次元確率分布が**表** 5-1 で与えられるとき，$E(XY)$ の値を求めよ．

（解）

$$E(XY)=\sum_{x=1}^{2}\sum_{y=1}^{3}xyP(X=x,Y=y)$$
$$=1\times(1\times 0+2\times 0.1+3\times 0.2)+2\times(1\times 0.1+2\times 0.2+3\times 0.4)$$
$$=4.2$$

を得る.

〔**例題** 5.9〕 a と b を定数とするとき,

$$E(aX+bY)=aE(X)+bE(Y)$$

となることを示せ.

（解）次の計算から明らかである.

$$E(aX+bY)=\sum_{x}\sum_{y}(ax+by)P(X=x,Y=y)$$
$$=a\sum_{x}x\sum_{y}P(X=x,Y=y)+b\sum_{y}y\sum_{x}P(X=x,Y=y)$$
$$=a\sum_{x}xP(X=x)+b\sum_{y}yP(Y=y)$$
$$=aE(X)+bE(Y)$$

〔**例題** 5.10〕 ゆがみのない 2 個のサイコロを同時に投げるとする. このとき, 出る目の和の期待値を求めよ.

（解）2 個のサイコロの出る目を, それぞれ X と Y とするとき, $E(X+Y)$ を求めればよい. このとき, 例題 5.6 と 5.9 より, 即座に

$$E(X+Y)=E(X)+E(Y)=3.5+3.5=7$$

を得る.

共分散

確率変数 X と Y の**共分散**は,

$$Cov(X,Y)=E\left[\left(X-E(X)\right)\left(Y-E(Y)\right)\right]$$
$$=E(XY)-E(X)E(Y)$$

〔**例題** 5.11〕 a と b を定数とするとき,

$$V(aX+bY)=a^2V(X)+b^2V(Y)+2abCov(X,Y)$$

となることを示せ.

（解） $Z=aX+bY$ は 1 次元の確率変数であるから, その分散を考えると次のようになる.

$$V(aX+bY)=E\left[\left\{aX+bY-E\left(aX+bY\right)\right\}^2\right]$$
$$=E\left[\left\{a\left(X-E(X)\right)+b\left(Y-E(Y)\right)\right\}^2\right]$$

あとは, 大きなカッコの中の 2 乗を展開して, 項ごとに期待値をとればよい.

相関係数

確率変数 X と Y の**相関係数**は,

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{Cov(X,Y)}{SD(X)SD(Y)}$$

により定義される．その値は，データの場合の相関係数と同様に，絶対値が 1 以下となる1．

〔**例題 5.12**〕 **表 5-1** の X と Y の 2 次元確率分布から，共分散と相関係数を求めよ．

（解）X と Y の期待値と分散はすでに求めた結果を利用する．共分散は,

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$
$$= 4.2 - 1.7 \times 2.5 = -0.05$$

となる．また，相関係数は

$$\rho_{XY} = \frac{-0.05}{\sqrt{0.21 \times 0.45}} = -0.16$$

となる．X と Y の間には弱い負の相関関係があることになる．

〔**例題 5.13**〕 2 個のサイコロを投げたときに出る目のうち，小さい方を X，大きい方を Y とする．ただし，同じ目が出たときは $X=Y$ とする．このとき，X と Y の 2 次元確率分布，および相関係数を求めよ

（解）2 次元確率分布は $P(X=x, Y=y)$ の形で表される．ただし，(x,y) のとりうる値は，1 から 6 までの整数値で，$x \le y$ となるものである．ここで，x と y に大小関係を考えないと，(x,y) の組み合わせは全部で 36 通りとなるが，このうち，$x > y$ となるペアは x と y を入れ替える．このように考えて，求める 2 次元確率分布と周辺分布は**表 5-2**（112 ページ）のようになる．

この表から,

$$E(X) = 91/36, \qquad E(Y) = 161/36$$

$$V(X) = V(Y) = \frac{2555}{1296}, \qquad E(XY) = \frac{441}{36}$$

を得る．したがって，相関係数は

$$\rho_{XY} = \frac{\frac{441}{36} - \frac{91 \times 161}{36 \times 36}}{\frac{2555}{1296}} = \frac{441 \times 36 - 91 \times 161}{2555} = \frac{1225}{2555}$$

$$=0.479$$

となる．

6　確率変数の独立性と同一分布性

$\boxed{\text{条件付き確率分布}}$ 2 次元の離散的確率変数 X と Y に対して，$P(X=x)$ が正なら

ば $X=x$ が与えられたときに $Y=y$ となる確率を

$$P(Y=y|X=x)=\frac{P(X=x,Y=y)}{P(X=x)} \tag{16}$$

で定義する．そして，この確率を y のすべてについて考えるとき，$X=x$ が与えられたときの Y の**条件付き確率分布**という．条件付き確率分布は，同時分布と周辺分布の比となっている．

〔**例題 5.14**〕 **表 5-1** の X と Y の 2 次元確率分布から，$X=1$ が与えられたときの Y の条件付き確率分布を求めよ．

（解）$X=1$ が与えられたとき，Y がとりうる値は 2 か 3 である．したがって，条件付き確率分布は，(16) を使って，

$$P(Y=2|X=1)=\frac{P(X=1,Y=2)}{P(X=1)}=\frac{0.1}{0.3}=\frac{1}{3}$$

$$P(Y=3|X=1)=\frac{P(X=1,Y=3)}{P(X=1)}=\frac{0.2}{0.3}=\frac{2}{3}$$

独立性　X のとりうるどのような値を与えても Y の条件付き確率分布が変わらないならば，すなわち，

$$P(Y=y|X=x)=P(Y=y)$$

が，X と Y のとりうる値 x，y のすべてに対して成り立つならば，X と Y は**互いに独立**であるという．このことは，

$$P(X=x,Y=y)=P(X=x)\times P(Y=y)$$

が，X と Y のとりうる値 x，y のすべてに対して成り立つことと同値である．

X と Y が互いに独立ならば，X の関数 $g(X)$ と Y の関数 $h(Y)$ も互いに独立である．そして，さらに次のことが成立する．

〔**例題 5.15**〕 X と Y が互いに独立ならば，X の関数 $g(X)$ と Y の関数 $h(Y)$ に対して，

$$E[g(X)h(Y)]=E[g(X)]E[h(Y)]$$

となることを示せ．

（解）独立性の仮定を使って，

$$\begin{aligned}E[g(X)h(Y)]&=\sum_x\sum_y g(x)h(y)P(X=x)P(Y=y)\\&=\sum_x g(x)P(X=x)\sum_y h(y)P(Y=y)\\&=E[g(X)]E[h(Y)]\end{aligned}$$

となる．

このことから，

X と Y が互いに独立 $\Rightarrow$ 共分散は 0

なぜなら，

$$Cov(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

独立確率変数の性質
$X_1,\cdots,X_n$ が互いに独立 $\Rightarrow$ $V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V\left(X_i\right)$

ところで，サイコロを n 回投げるときに出る目 $X_1,\cdots,X_n$ は，互いに独立であるだけでなく，周辺確率分布も同一と考えられる．このとき，$X_1,\cdots,X_n$ は**独立，同一分布**に従うという．そして，共通の期待値を μ，分散を σ^2 とするとき，このことを，以下では

$$X_1,\cdots,X_n \quad \sim \quad i.i.d.(\mu,\sigma^2)$$

と表す．ここで，$i.i.d.$ は independent and identically distributed の略である．このとき，1 次結合の期待値と分散は，次のようにさらに簡単になる．

独立，同一分布に従う確率変数の性質
$X_1,\cdots,X_n \quad \sim \quad i.i.d.(\mu,\sigma^2) \quad \Rightarrow \quad E(\sum_{i=1}^{n} a_i X_i) = \mu\sum_{i=1}^{n} a_i$ (17)
$V(\sum_{i=1}^{n} a_i X_i) = \sigma^2 \sum_{i=1}^{n} a_i^2$ (18)

上の関係式の 1 次結合において，特に $a_i = 1/n$ $(i = 1,\cdots,n)$ とすれば，確率変数

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

を考えることができる．これは n 個の確率変数の平均であり，特定の n 個のデータに基づく平均 $\bar{x}$ とは明確に区別されるべきである．$\overline{X}$ の期待値と分散は (17)，(18) から次のようになる．

確率変数の平均 $\overline{X}$ の性質（独立，同一分布の場合）
$X_1,\cdots,X_n \quad \sim \quad i.i.d.(\mu,\sigma^2) \quad \Rightarrow \quad E(\overline{X}) = \mu$ (19)
$V(\overline{X}) = \dfrac{\sigma^2}{n}$ (20)

関係式 (19)，(20) は重要なメッセージを含んでいる．それは，$\overline{X}$ の期待値は個々の確率変数の期待値に等しいが，分散はその $1/n$ になるということである．このことは，あとの章で議論する統計的推測の有効性を支持する根拠ともなる重要な性質である．