§2　統計量と標本分布

2.0　多次元確率変数

2 つの確率変数の組 (X, Y) を 2 次元確率変数という。

以下、しばらく離散型の場合を考える。X のとり得る値を $x_1, x_2, \ldots, x_m$, Y のとり得る値を $y_1, y_2, \ldots, y_n$ とする。

確率変数の組 (X, Y) に対し

$$P(X = x_i, Y = y_j) = p_{ij} \quad (1 \leq i \leq m, 1 \leq j \leq n)$$

とする。これを (X, Y) の同時確率分布といい、それを表にしたものを同時確率分布表という。ここで、

$$p_{ij} \geq 0, \quad \sum_{j=1}^{n} \sum_{i=1}^{m} p_{ij} = 1$$

$X \backslash Y$	y_1	y_2	$\cdots$	y_n	計
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	p_1
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	p_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_m	p_{n1}	p_{n2}	$\cdots$	p_{mn}	p_m
計	q_1	q_2	$\cdots$	q_n	1

同時確率分布表

である。このとき、

$$P(X = x_i) = \sum_{j=1}^{n} p_{ij} = p_i \quad (1 \leq i \leq m) \qquad \text{を } X \text{ の周辺分布,}$$

$$P(Y = y_j) = \sum_{i=1}^{m} p_{ij} = q_j \quad (1 \leq j \leq n) \qquad \text{を } Y \text{ の周辺分布,}$$

という。同時確率分布表では右の列が X の周辺分布、下の行が Y の周辺分布となっている。

例 2.1 袋の中に 1, 2, 3 の数字の書かれた球がそれぞれ 5 個, 3 個, 2 個入っている。この袋から 1 個ずつ球を取り出すとき、1 個め, 2 個めに出た球に書かれていた数字をそれぞれ

(1) 非復元抽出のとき X_1, Y_1 とし、

(2) 復元抽出のとき X_2, Y_2 とする。

このとき、(X_1, Y_1) と (X_2, Y_2) の同時確率分布を調べ、同時確率分布表をつくれ。

解: (1) $P(X_1 = 1, Y_1 = 1) = \dfrac{5}{10}\dfrac{4}{9} = \dfrac{2}{9}$,

$P(X_1 = 1, Y_1 = 2) = \dfrac{5}{10}\dfrac{3}{9} = \dfrac{1}{6}$,

$P(X_1 = 1, Y_1 = 3) = \dfrac{5}{10}\dfrac{2}{9} = \dfrac{1}{9}$,

$\vdots$

と求めると、(X_1, Y_1) について右の同時分布表を得る。

(2) $P(X_2 = 1, Y_2 = 1) = \dfrac{5}{10}\dfrac{5}{10} = \dfrac{1}{4}$,

$P(X_2 = 1, Y_2 = 2) = \dfrac{5}{10}\dfrac{3}{10} = \dfrac{3}{20}$,

$P(X_2 = 1, Y_2 = 3) = \dfrac{5}{10}\dfrac{2}{10} = \dfrac{1}{10}$,

$\vdots$

と求めると、(X_2, Y_2) について右の同時分布表を得る。　□

$X_1 \backslash Y_1$	1	2	3	計
1	$\frac{2}{9}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{2}$
2	$\frac{1}{6}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{3}{10}$
3	$\frac{1}{9}$	$\frac{1}{15}$	$\frac{1}{45}$	$\frac{1}{5}$
計	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{5}$	1

(1) 非復元抽出

$X_2 \backslash Y_2$	1	2	3	計
1	$\frac{1}{4}$	$\frac{3}{20}$	$\frac{1}{10}$	$\frac{1}{2}$
2	$\frac{3}{20}$	$\frac{9}{100}$	$\frac{3}{50}$	$\frac{3}{10}$
3	$\frac{1}{10}$	$\frac{3}{50}$	$\frac{1}{25}$	$\frac{1}{5}$
計	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{5}$	1

(2) 復元抽出

注意 これより、X_1 と X_2, Y_1 と Y_2 の周辺分布はともに等しいが、(X_1, Y_1) と (X_2, Y_2) の同時確率分布は異なることがわかる。このように、同時確率分布を考察することは確率分布を理解するうえで重要である。

(**確率変数の独立性**) 確率変数 X, Y が独立であるとは

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad (1 \le i \le m, 1 \le j \le n)$$

となるときにいう。例 2.1 では (2) の X_2, Y_2 は独立であるが、(1) の X_1, Y_1 は独立ではない。

2.1 確率変数の関数

(X, Y) の同時確率分布が前ページの同時確率分布表で与えられるとき、関数 $\phi(x, y)$ に対して

$$E[\phi(X, Y)] = \sum_{i=1}^{m}\sum_{j=1}^{n} \phi(x_i, y_j)P(X = x_i, Y = y_j)$$

と定める。また、$\phi(X, Y)$ の分散を

$$V[\phi(X, Y)] = E[\Big(\phi(X, Y) - E[\phi(X, Y)]\Big)^2] \tag{2.1}$$

で定める。このとき次が成立する。(証明は省略する。)

定理 2.1 (平均と分散の性質) (1) a, b, c が定数のとき $E[aX + bY + c] = aE[X] + bE[Y] + c$ となる。
さらに、X, Y が独立であれば、関数 $\varphi(x), \psi(y)$ に対して $E[\varphi(X)\psi(Y)] = E[\varphi(X)]E[\psi(Y)]$.
(2) a, b, c が定数のとき、X, Y が独立であれば、$V[aX + bY + c] = a^2V[X] + b^2V[Y]$ となる。

連続型の場合: 2 次元確率変数 (X, Y) が連続型であるとは、ある関数 $f(x, y)$ があって、

$$P(a \le X \le b, c \le Y \le d) = \int_a^b \Big(\int_c^d f(x, y)\,dy\Big)\,dx$$

と表されるときをいい、この $f(x, y)$ を (X, Y) の同時密度関数という。このとき、

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy \quad \text{を } X \text{ の周辺密度関数,} \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx \quad \text{を } Y \text{ の周辺密度関数}$$

という。また、X, Y が独立であるとは、$f(x, y) = f_X(x)f_Y(y)$ が成立するときにいう。関数 $\phi(x, y)$ に対して

$$E[\phi(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(x, y)f(x, y)\,dxdy$$

と定め、分散 $V[\phi(X, Y)]$ を (2.1) で定める。このとき、離散型のときと同様に**定理 2.1** が成立する。

3 つ以上の確率変数 $X_1, X_2, \ldots, X_n$ を考える。$X_1, X_2, \ldots, X_n$ が独立であるとは (離散型、連続型も含んだ形で)、任意の実数 $a_1, b_1, \ldots, a_n, b_n$ $(a_i \le b_i,\ 1 \le i \le n)$ に対して

$$P(a_1 \le X_1 \le b_1, \cdots, a_n \le X_n \le b_n) = P(a_1 \le X_1 \le b_1) \cdots P(a_n \le X_n \le b_n)$$

が成り立つときと定義する。また、2 次元確率変数の場合と同様に、同時確率分布や同時密度関数を用いて、$X_1, X_2, \ldots, X_n$ の関数の平均や分散が定義でき、さらに、定理 2.1 と同様に次の定理が成立する。

定理 2.2 (平均と分散の性質) (1) $a_1, a_2, \ldots, a_n$ が定数のとき

$$E[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1E[X_1] + a_2E[X_2] + \cdots + a_nE[X_n].$$

(2) $a_1, a_2, \ldots, a_n$ が定数で $X_1, X_2, \ldots, X_n$ が独立のとき

$$V[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = {a_1}^2V[X_1] + {a_2}^2V[X_2] + \cdots + {a_n}^2V[X_n].$$