§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、

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

である。このとき、

$$P(X = x_i) = \sum_{j=1}^{n} p_{ij} = p_i \quad (1 \leq i \leq m) \qquad \text{を } X \text{ の周辺分布}, \tag{2.1}$$

$$P(Y = y_j) = \sum_{i=1}^{m} p_{ij} = q_j \quad (1 \leq j \leq n) \qquad \text{を } Y \text{ の周辺分布}, \tag{2.2}$$

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。

$X \backslash Y$	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

表 2.1 同時確率分布表

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)

$$P(X_1 = 1, Y_1 = 1) = \frac{5}{10}\frac{4}{9} = \frac{2}{9},$$
$$P(X_1 = 1, Y_1 = 2) = \frac{5}{10}\frac{3}{9} = \frac{1}{6},$$
$$P(X_1 = 1, Y_1 = 3) = \frac{5}{10}\frac{2}{9} = \frac{1}{9},$$
$$\vdots$$

と求めると、(X_1, Y_1) について右の同時分布表を得る。

(2)

$$P(X_2 = 1, Y_2 = 1) = \frac{5}{10}\frac{5}{10} = \frac{1}{4},$$
$$P(X_2 = 1, Y_2 = 2) = \frac{5}{10}\frac{3}{10} = \frac{3}{20},$$
$$P(X_2 = 1, Y_2 = 3) = \frac{5}{10}\frac{2}{10} = \frac{1}{10},$$
$$\vdots$$

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表 2.1 同時確率分布表で与えられるとき、関数 $\phi(x, y)$ に対して

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

と定める。また、$\phi(X, Y)$ の分散を

$$V[\phi(X,Y)] = E[\Big(\phi(X,Y) - E[\phi(X,Y)]\Big)^2] \tag{2.3}$$

で定める。このとき次が成立する。

定理 2.1 X, Y が独立であれば、$E[XY] = E[X]E[Y]$.

証明: (X, Y) の同時分布が表 2.1 で与えられているとすると、$p_{ij} = p_i q_j$ なので、

$$\begin{aligned} E[XY] &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_i q_j \\ &= \sum_{i=1}^{m} x_i p_i \sum_{j=1}^{n} y_j q_j = E[X]E[Y]. \quad \square \end{aligned}$$

定理 2.2 (平均と分散の性質) 定数 a, b, c に対して、$E[aX + bY + c] = aE[X] + bE[Y] + c$ となる。さらに、X, Y が独立であれば、$V[aX + bY + c] = a^2V[X] + b^2V[Y]$ となる。

証明: (X, Y) の同時分布が表 2.1 で与えられているとすると、

$$\begin{aligned} E[aX + bY + c] &= \sum_{i=1}^{m}\sum_{j=1}^{n}(ax_i + by_j + c)p_{ij} = a\sum_{i=1}^{m} x_i \sum_{j=1}^{n} p_{ij} + b\sum_{j=1}^{n} y_j \sum_{i=1}^{m} p_{ij} + c\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \\ &= a\sum_{i=1}^{m} x_i p_i + b\sum_{j=1}^{n} y_j q_j + c \cdot 1 = aE[X] + bE[Y] + c. \end{aligned}$$

$\mu_1 = E[X]$, $\mu_2 = E[Y]$ とおく。X, Y が独立であれば、

$$E[(X - \mu_1)(Y - \mu_2)] = E[XY] - \mu_2 E[X] - \mu_1 E[Y] + \mu_1\mu_2 = E[XY] - \mu_1\mu_2 = 0$$

となることに注意する。したがって、

$$\begin{aligned} &V[aX + bY + c] = E[\{aX + bY + c - (a\mu_1 + b\mu_2 + c)\}^2] = E[\{a(X - \mu_1) + b(Y - \mu_2)\}^2] \\ &= a^2E[(X - \mu_1)^2] + 2abE[(X - \mu_1)(Y - \mu_2)] + b^2E[(Y - \mu_2)^2] = a^2V[X] + b^2V[Y]. \quad \square \end{aligned}$$

注意 $\mathrm{Cov}[X.Y] = E[(X - \mu_1)(Y - \mu_2)] = E[XY] - \mu_1\mu_2$ と定め、これを X, Y の**共分散**という。

例題 2.2 (X, Y) の同時確率分布表が右で与えられるとする。
(1) X の周辺分布, Y の周辺分布、$E[X], E[Y]$ を求めよ。
(2) $E[XY], \mathrm{Cov}(X, Y)$ を求めよ。

X \ Y	1	2	3
1	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{3}$

解: 授業で説明する。　$\square$