3.4 共分散と相関係数

X と Y の 2 次までの積率としてそれぞれの平均と分散：

$$\mathbb{E}[X], \qquad \mathbb{E}[Y], \qquad \mathbb{VAR}[X], \qquad \mathbb{VAR}[Y]$$

に加えて，平均まわりの相互積率を X と Y の**共分散**といい，

$$\mathbb{COV}[X, Y] = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

と記すこと[6]にする．

共分散の公式

$$\mathbb{COV}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

証明

$$\begin{aligned}\mathbb{COV}[X, Y] = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] &= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

がわかる． □

定理 3.4 X と Y が独立ならば，

$$\mathbb{COV}[X, Y] = 0$$

である．

証明

$$\begin{aligned}\mathbb{COV}[X, Y] = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

がわかる． □

注意 3.6 定理 3.4 の逆は必ずしも成立しないことに注意せよ．

定理 3.5 a, b を定数として，$aX + bY$ の期待値と分散はつぎのようになる：

$$\begin{aligned}\mathbb{E}[aX + bY] &= a\mathbb{E}[X] + b\mathbb{E}[Y], \\ \mathbb{VAR}[aX + bY] &= a^2\mathbb{VAR}[X] + b^2\mathbb{VAR}[Y] + 2ab\mathbb{COV}[X, Y]\end{aligned}$$

となる．特に，X と Y が独立ならば，

$$\mathbb{VAR}[aX + bY] = a^2\mathbb{VAR}[X] + b^2\mathbb{VAR}[Y]$$

である．

[6] $\mathbb{E}[X^2]$ と $\mathbb{E}[Y^2]$ の期待値の存在を仮定すれば，ここで考えている期待値はすべて存在する．

例 3.5 (X, Y) は同時確率密度関数

$$f_{X,Y} = \begin{cases} 1, & 0 < x < 1,\, x < y < x+1, \\ 0, & (\text{その他}), \end{cases}$$

を持つとする．このときの X と Y の共分散を求めよう．まず，X と Y の周辺確率密度関数を求めよう．$0 < x < 1$ のとき，$f_{X,Y}(x, y) > 0$ であることに注意すれば，$0 < x < 1$ のとき，

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_x^{x+1} dy = 1$$

となる．それ以外では $f_{X,Y}(x, y) = 0$ なので，$f_X(x) = 0$ となる．したがって

$$f_X(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & (\text{その他}), \end{cases}$$

となる．また，$0 < y < 2$ のとき，$f_{X,Y}(x, y) > 0$ となることに注意すれば，

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx \begin{cases} \int_0^y 1\, dx, & 0 < y < 1, \\ \int_{y-1}^1 1\, dx, & 1 \le y < 2, \end{cases} = \begin{cases} y, & 0 < y < 1, \\ 2-y & 1 \le y < 2, \end{cases}$$

となる．したがって

$$f_Y(y) = \begin{cases} y, & 0 < y < 1, \\ 2-y & 1 \le y < 2, \\ 0, & (\text{その他}) \end{cases}$$

となる．これらに注意すれば，

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)\, dx \int_0^1 x\, dx = \frac{1}{2}, \\ \mathbb{E}[Y] &= \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_0^1 y^2\, dy + \int_1^2 y(2-y)\, dy = \frac{1}{3} + \frac{2}{3} = 1, \\ \mathbb{E}[XY] &= = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x, y)\, dx\, dy = \int_0^1 \int_x^{x+1} xy\, dy\, dx = \int_0^1 x \left[\frac{y^2}{2}\right]_x^{x+1} dx \\ &= \int_0^1 \frac{x(2x+1)}{2}\, dx = \left[\frac{x^3}{3} - \frac{x^2}{4}\right]_0^1 = \frac{1}{12} \end{aligned}$$

となる．したがって，

$$\mathbb{COV}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{1}{12} - \frac{1}{2} = -\frac{5}{12}$$

となる．

相関係数

X と Y の標準化

$$\frac{X-\mathbb{E}[X]}{\sqrt{\mathbb{VAR}[X]}}, \qquad \frac{Y-\mathbb{E}[Y]}{\sqrt{\mathbb{VAR}[Y]}}$$

の積の平均を X と Y の**相関係数**といい,

$$\rho[X, Y] = \mathbb{E}\left[\frac{X-\mathbb{E}[X]}{\sqrt{\mathbb{VAR}[X]}} \times \frac{Y-\mathbb{E}[Y]}{\sqrt{\mathbb{VAR}[Y]}}\right] = \frac{\mathbb{COV}[X, Y]}{\sqrt{\mathbb{VAR}[X]}\sqrt{\mathbb{VAR}[Y]}}$$

で記すことにする.

$\rho[X, Y] > 0$ のときには X と Y に**正の相関**, $\rho[X, Y] < 0$ のときには X と Y に**負の相関**があるといい, $\rho[X, Y] = 0$ のときには X と Y は**無相関**であるという. 定義と定理 3.4 から, X と Y が独立ならば, X と Y は無相関である. また, 注意 3.6 から X と Y は無相関であっても X と Y が必ずしも独立ではない.

定理 3.6 X と Y は 2 次の積率をもつ任意の確率変数とする. このとき,

(i) $-1 \leq \rho[X, Y] \leq 1$.

(ii) $|\rho[X, Y]| = 1$ となるための必要十分条件は定数 $a, b\,(a \neq 0)$ が存在して,

$$P(Y = aX + b) = 1$$

が成立することである. 特に, $\rho[X, Y] = 1$ ならば, $a > 0$ となり, $\rho[X, Y] = -1$ ならば, $a < 0$ となる.

証明 (i) を示すために,

$$h(t) = \mathbb{E}[\{(X-\mu_X)t + (Y-\mu_Y)\}^2]$$

とする. ただし, $\mu_X = \mathbb{E}[X], \mu_Y = \mathbb{E}[Y]$ である. 期待値の中を展開すれば,

$$\begin{aligned} h(t) &= t^2\mathbb{E}[(X-\mu_X)^2] + 2t\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] + \mathbb{E}[(Y-\mu_Y)^2] \\ &= t^2\mathbb{VAR}[X] + 2\mathbb{COV}[X, Y] + \mathbb{VAR}[Y] \\ &= \mathbb{VAR}[X]\left(t + \frac{\mathbb{COV}[X, Y]}{\mathbb{VAR}[X]}\right)^2 + \frac{\mathbb{VAR}[X]\mathbb{VAR}[Y] - \{\mathbb{COV}[X, Y]\}^2}{\mathbb{VAR}[X]} \end{aligned}$$

となる. しかし, $h(t) \geq$ から

$$\mathbb{VAR}[X]\mathbb{VAR}[Y] - \{\mathbb{COV}[X, Y]\}^2 \geq 0$$

を得る. したがって,

$$\frac{\{\mathbb{COV}[X, Y]\}^2}{\mathbb{VAR}[X]\mathbb{VAR}[Y]} \leq 1 \Longleftrightarrow \left|\frac{\mathbb{COV}[X, Y]}{\sqrt{\mathbb{VAR}[X]}\sqrt{\mathbb{VAR}[Y]}}\right| \leq 1$$

となり，(i) は示せた．
(ii) 上の議論から $|\rho[X, Y]| = 1$ となるための必要十分条件は

$$\mathbb{VAR}[X]\mathbb{VAR}[Y] - \{\mathbb{COV}[X, Y]\}^2 = 0 \Longleftrightarrow \ t \text{ の二次方程式 } h(t) = \text{ は単根である．}$$

である．この解を t_0 とする．しかし，$\mathbb{E}[(X - \mu_X)t_0 + (Y - \mu_Y)] = 0$ に注意すれば，

$$0 = h(t_0) = \mathbb{E}[\{(X - \mu_X)t_0 + (Y - \mu_Y)\}^2] = \mathbb{VAR}[(X - \mu_X)t_0 + (Y - \mu_Y)]$$

となり，分散の性質 (iii) から

$$P((X - \mu_X)t_0 + (Y - \mu_Y) = 0) = 1$$

となる．したがって，$a = -t_0, b = \mu_X t_0 + \mu_Y$ とおけば，$P(aX + b = Y) = 1$ が成立することがわかる．また，

$$h(t) = \mathbb{VAR}[X]\left(t - \frac{\mathbb{COV}[X, Y]}{\mathbb{VAR}[X]}\right)^2 = 0$$

に注意すれば，方程式

$$h(t) = \mathbb{VAR}[X]\left(t + \frac{\mathbb{COV}[X, Y]}{\mathbb{VAR}[X]}\right)^2 = 0$$

の解は

$$t = t_0 = -\frac{\mathbb{COV}[X, Y]}{\mathbb{VAR}[X]}$$

なので，

$$a = \frac{\mathbb{COV}[X, Y]}{\mathbb{VAR}[X]}$$

となる．したがって，a の符号と $\rho[X, Y]$ の符号は同じになる．よって，(ii) は示された． □