相関係数

前回までの補足

復習

$$V[aX+bY] = a^2V[X] + 2ab\,Cov(X,Y] + b^2V[Y]$$

Special Case

- $Y=1$ (constant)
  $V[aX+b] = a^2V(X]$

(∵)
$V[aX+bY]$
$= a^2V[X] + 2ab\,Cov[X,1] + b^2V[1]$
$= a^2V(X]$. //

Special Case

- $b=0$
  $V[aX] = a^2V[X]$

Th

X : r.v. with $E(X)=\mu$, $V[X]=\sigma^2>0$

$$Z_X \equiv \frac{X-\mu_X}{\sigma_X} \quad (\text{標準化 r.v.})$$

$\Longrightarrow$ ① $E[Z_X]=0$

② $V[Z_X]=1$

③ $E[Z_X^2]=1$

Proof

$$① \ E[Z_X]=E\left(\frac{X-\mu_X}{\sigma_X}\right]$$

$$=\frac{1}{\sigma_X}\left(E[X]-\mu_X\right)=0$$

$$② \ V[Z_X]=V\left[\frac{X-\mu_X}{\sigma_X}\right]$$

$$=V\left[\frac{1}{\sigma_X}X-\frac{\mu_X}{\sigma_X}\right]$$

$$=\frac{1}{\sigma_X^2}V[X]=\frac{1}{\sigma_X^2}\sigma_X^2=1.$$

③ $E[Z_X^2]$

$$=E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^2\right]$$

$$=\frac{1}{\sigma_X^2}E\left[(X-\mu_X)^2\right]$$

$$=\frac{1}{\sigma_X^2}V[X]=1.$$

//

相関係数

X, Y : r.v.

$V[X], V[Y] > 0$ ← 前提

Def.

$$\boxed{\rho[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sqrt{V[X]}\sqrt{V[Y]}}}$$

X と Y の **相関係数**

correlation coefficient

$\rho[X,Y] > 0$ 　 X と Y は 正の相関関係がある.

< 0 　 〃 　 負の相関関係がある.

$= 0$ 　 〃 　 無相関である.

定義からただちにわかること

① $\rho[X,X] = 1$

② $\rho[X,-X] = -1$

③ $\rho[X,Y] = \rho[Y,X]$

確率変数 X,Y
$V[X],\ V[Y]>0$とする。

このとき、相関係数$\rho[X,Y]$の計算手順
(まずは分散と共分散を求める。)

Step 1. 周辺確率(密度)関数を求める。

Step 2. 期待値$E[X],\ E[Y],\ E\left[X^2\right],\ E\left[Y^2\right],\ E[XY]$を計算する。

$$E[X]=\begin{cases}\displaystyle\sum_i x_i P(X=x_i) & X\text{が離散確率変数のとき}\\ \displaystyle\int_{-\infty}^{\infty} x\ f_X(x)\,dx & X\text{が連続確率変数のとき}\end{cases}$$

Step 3. 分散、共分散を計算する。

$$V[X]=E\left[X^2\right]-\left(E[X]\right)^2$$
$$V[Y]=E\left[Y^2\right]-\left(E[Y]\right)^2$$
$$Cov[X,Y]=E[XY]-E[X]E[Y]$$

Step 4. 標準偏差と共分散を用いて相関係数を計算する。

$$\rho[X,Y]=\frac{Cov[X,Y]}{\sqrt{V[X]}\sqrt{V[Y]}}$$

Th.

X, Y : r.v.

$V[X], V[Y] > 0$

$a, b, c, d \in \mathbb{R}$

$a, c \neq 0$

$$\Longrightarrow \rho[aX+b, cY+d] = \frac{ac}{|ac|} \rho[X, Y]$$

Proof

It follows that

$$\begin{aligned}
&\rho[aX+b, cY+d] \\
&= \frac{Cov[aX+b, cY+d]}{\sqrt{V[aX+b]} \cdot \sqrt{V[cY+d]}} \\
&= \frac{ac\, Cov(X, Y)}{\sqrt{a^2 V(X)} \sqrt{c^2 V[Y]}} \\
&= \frac{ac}{|ac|} \frac{Cov[X, Y]}{\sqrt{V(X)} \sqrt{V(Y)}} \\
&= \frac{ac}{|ac|} \rho[X, Y].
\end{aligned}$$

//

※ $Cov[aX+b, cY+d] = ac\, Cov(X, Y]$

$V[aX+b] = a^2 V(X]$

← 復習

考察

- $ac>0$ (aとcが同符号)のとき

$$\rho[aX+b, cY+d] = \rho[X, Y]$$

- $ac<0$ (aとcが異符号)のとき

$$\rho[aX+b, cY+d] = -\rho[X, Y]$$

例 $\rho[2X-5, -3Y+1] = -\rho[X, Y]$

- $X=Y$とすると

$$\rho[aX+b, cX+d] = \frac{ac}{|ac|} = \begin{cases} 1 & \text{if } ac>0 \\ -1 & \text{if } ac<0 \end{cases}$$

- $X=Y$, $c=1$, $d=0$とすると

$$\rho[X, aX+b] = \begin{cases} 1 & \text{if } a>0 \\ -1 & \text{if } a<0 \end{cases}$$

※ r.v. XとYに1次の関係がある場合、
それらの相関係数$\rho[X,Y]$は 1 or −1.
正の関係の場合は1, 負の関係の場合は−1になる。

例 $\rho[X, 2X-7] = 1$

$\rho[X, -3X+10] = -1$

Th

X, Y : r.v.

$V[X] = \sigma_X^2 > 0$

$V[Y] = \sigma_Y^2 > 0$

$$Z_X \equiv \frac{X - \mu_X}{\sigma_X}$$

$$Z_Y \equiv \frac{Y - \mu_Y}{\sigma_Y}$$

$$\Rightarrow \rho[Z_X, Z_Y] = \rho[X, Y]$$

Proof

$$\rho[Z_X, Z_Y]$$

$$= \rho\left[\frac{1}{\sigma_X}X - \frac{\mu_X}{\sigma_X}, \frac{1}{\sigma_Y}Y - \frac{\mu_Y}{\sigma_Y}\right]$$

$$= \frac{\frac{1}{\sigma_X}\cdot\frac{1}{\sigma_Y}}{\left|\frac{1}{\sigma_X}\frac{1}{\sigma_Y}\right|}\rho[X, Y]$$

$\sigma_X, \sigma_Y > 0$

$$= \rho[X, Y].$$

//

cf.

X, Y : r.v.

$V[X], V[Y] > 0$

$$Z_X = \frac{X - \mu_X}{\sigma_X}$$

$$Z_Y = \frac{Y - \mu_Y}{\sigma_Y}$$

$$\Longrightarrow Cov[Z_X, Z_Y] = \frac{1}{\sigma_X \sigma_Y} Cov[X, Y]$$

Proof

$$Cov[Z_X, Z_Y]$$

$$= Cov\left[\frac{1}{\sigma_X} X - \frac{\mu_X}{\sigma_X}, \frac{1}{\sigma_Y} Y - \frac{\mu_Y}{\sigma_Y}\right]$$

$$= \frac{1}{\sigma_X \sigma_Y} Cov[X, Y]$$

//

復習

$$Cov[aX+b, cY+d] = ac\, Cov[X, Y]$$

次の目標：

Th

X, Y : r.v. with $V[X], V[Y] > 0$

$\Longrightarrow \rho[X, Y] \in [-1, 1]$

i.e. $|\rho[X, Y]| \leqq 1$

Th

X, Y : r.v. with $V[X], V[Y] > 0$

$\Longrightarrow |\mathrm{Cov}[X, Y]| \leqq \sqrt{V[X]}\sqrt{V[Y]}$

Lemma

X, Y : r.v.

$\Rightarrow |E[XY]| \leq \sqrt{E[X^2]}\sqrt{E[Y^2]}$

Proof

Let $t \in \mathbb{R}$.

It holds that

$$0 \leq E[(tX+Y)^2]$$
$$= E[t^2X^2 + 2tXY + Y^2]$$
$$= t^2E[X^2] + 2tE[XY] + E[Y^2] \quad \forall t \in \mathbb{R}.$$

Therefore,

$$D/4 = (E[XY])^2 - E[X^2]E[Y^2] \leq 0.$$

$$\therefore |E[XY]| \leq \sqrt{E[X^2]}\sqrt{E[Y^2]}.$$

//

Lemma

X, Y : r.v. with $V[X], V[Y] > 0$

$$Z_X \equiv \frac{X - \mu_X}{\sigma_X}$$

$$Z_Y \equiv \frac{Y - \mu_Y}{\sigma_Y}$$

$$\Longrightarrow \rho[Z_X, Z_Y] = E(Z_X Z_Y)$$

Proof

$$\rho(Z_X, Z_Y)$$

$$= \frac{Cov[Z_X, Z_Y]}{\sqrt{V[Z_X]}\sqrt{V[Z_Y]}}$$

$$= Cov[Z_X, Z_Y]$$

$$= E(Z_X Z_Y] - E[Z_X]E(Z_Y]$$

$$= E[Z_X Z_Y].$$

//

※ $\rho[Z_X, Z_Y] = E(Z_X Z_Y]$

$\quad \parallel$

$\rho[X, Y]$ ← 前の定理

Th

X, Y : r.v. with $V[X], V[Y] > 0$

$\Rightarrow \rho[X, Y] \in [-1, 1]$

Proof

We show that $|\rho[X, Y]| \leq 1$.

Define
$$\begin{cases} Z_X \equiv \dfrac{X - \mu_X}{\sigma_X}, \\ Z_Y \equiv \dfrac{Y - \mu_Y}{\sigma_Y}. \end{cases}$$

We know that

① $E[Z_X^2] = E[Z_Y^2] = 1$

② $\rho[X, Y] = \rho[Z_X, Z_Y] = E[Z_X Z_Y]$

③ $|E[AB]| \leq \sqrt{E[A^2]}\sqrt{E[B^2]}$ $\quad \forall A, B$: r.v.

Substituting $\begin{cases} A = Z_X \\ B = Z_Y \end{cases}$ in ③, we obtain

$$\underset{\substack{\overset{②}{\|} \\ |\rho[Z_X, Z_Y]| \\ \overset{②}{\|} \\ |\rho[X, Y]|}}{|E[Z_X Z_Y]|} \overset{}{\underset{③}{\leq}} \sqrt{E[Z_X^2]}\sqrt{E[Z_Y^2]} \overset{①}{=} 1.$$

$\therefore |\rho[X, Y]| \leq 1.$

//

Th

X, Y : r.v. with $V(X]>0$, $V(Y]>0$

$$\Longrightarrow \left|Cov[X,Y]\right| \leqq \sqrt{V[X]}\sqrt{V[Y]}$$

Proof.

We know that

$$|\rho(X,Y]| \leqq 1$$

i.e. $$\left|\frac{Cov[X,Y]}{\sqrt{V[X]}\sqrt{V[Y]}}\right| \leqq 1.$$

Thus,

$$\left|Cov[X,Y]\right| \leqq \sqrt{V[X]}\sqrt{V[Y]}.$$

//

Remark

$\rho[X, Y] \in [-1, 1]$ だが、
X と Y に 1次の関係がある場合
$\rho[X, Y] = 1$ or -1 になる。

相関係数は 2つの r.v. の 1次結合の強さを表す尺度とも言える。

2つの r.v. の間に 2次の関係があっても
相関係数はゼロ(つまり無相関)に
なることがある。

($\rho[X, Y] = 0$ のとき、X と Y は「無相関」
というが、必ずしも関連がないことを
意味するわけではない。)

例 (2つのr.v.の間に2次の関係があるにもかかわらず無相関になる例)

$X \sim U[-1,1]$ とする。

$\therefore$ p.d.f. for X

$$f(x) = \begin{cases} 1/2 & \text{if } x \in [-1,1] \\ 0 & \text{if } x \notin [-1,1] \end{cases}$$

$Y \equiv X^2$ とする。(2次の関係)

Then, $E[X] = \int_{-1}^{1} x \cdot \frac{1}{2} dx = \left[\frac{x^2}{4}\right]_{-1}^{1} = 0$

$E[Y] = E[X^2]$

$= \int_{-1}^{1} x^2 \cdot \frac{1}{2} dx = \frac{1}{6}[x^3]_{-1}^{1} = \frac{1}{3}$

$E[XY] = E[X^3]$

$= \int_{-1}^{1} x^3 \cdot \frac{1}{2} dx = \frac{1}{8}[x^4]_{-1}^{1} = 0.$

$\therefore Cov[X,Y] = E[XY] - E[X]E[Y]$

$= 0.$

$\therefore \rho[X,Y] = 0.$

$\therefore$ X と Y は2次の関係 $(Y = X^2)$ にあるが、

無相関である。

//

相関係数

問題1. Xを$\sigma_X > 0$である確率変数とする。また、それ標準化した確率変数をZ_Xとおく。すなわち、

$$Z_X = \frac{X - \mu_X}{\sigma_X}$$

である。このとき、(1) $E[Z_X] = 0$, (2) $V[Z_X] = 1$, (3) $E[Z_X^2] = 1$となることを示しなさい。

問題2. a, b, c, d $(a, c \neq 0)$を定数とするとき、

$$\rho[aX + b,\ cY + d] = \frac{ac}{|ac|}\rho[X, Y]$$

を示しなさい。この結果により、確率変数Xに対して$Y = aX + b$（ただし、$a \neq 0$）とするとき、

$$\rho[X, Y] = \begin{cases} 1 & \text{if } a > 0 \\ -1 & \text{if } a < 0 \end{cases}$$

となることを確認しなさい。

問題3. X, Yを標準化した確率変数をZ_X, Z_Yとおく。すなわち、

$$Z_X = \frac{X - \mu_X}{\sigma_X}, \qquad Z_Y = \frac{Y - \mu_Y}{\sigma_Y}$$

である。もちろん、$\sigma_X, \sigma_Y > 0$であることは前提である。このとき、$\rho[Z_X, Z_Y] = \rho[X, Y]$を示しなさい。

問題4. A, Bを確率変数とするとき、$|E[AB]| \leq \sqrt{E[A^2]}\sqrt{E[B^2]}$を示しなさい。

問題5. $\rho[Z_X, Z_Y] = E[Z_X Z_Y]$を示しなさい。ここで、記号は問題3と同じである。

問題6. X, Yを、それぞれ、平均μ_X, μ_Y、分散$\sigma_X^2, \sigma_Y^2 (> 0)$の確率変数とする。以下を示しなさい。

(1) $\rho[X, Y] \in [-1, 1]$
(2) $|Cov[X, Y]| \leq \sqrt{V[X]}\sqrt{V[Y]}$

問題7. 2つの確率変数X, Yの分布が下表のように与えられているとする。

$X \backslash Y$	1	2	3	$P(X = x)$
1	0	1/10	2/10	
2	1/10	2/10	4/10	
$P(Y = y)$				

このとき、XとYの相関係数$\rho[X, Y]$を求めよ。

解答
問題7.
Step1. まず、XとYの周辺確率関数を求める。

$X\backslash Y$	1	2	3	$P(X=x)$
1	0	1/10	2/10	**3/10**
2	1/10	2/10	4/10	**7/10**
$P(Y=y)$	**1/10**	**3/10**	**6/10**	

Step2. $E[X], E[Y], E[X^2], E[Y^2], E[XY]$を求める。

$$E[X] = \frac{3}{10}\cdot 1 + \frac{7}{10}\cdot 2 = \frac{17}{10}$$

$$E[Y] = \frac{1}{10}\cdot 1 + \frac{3}{10}\cdot 2 + \frac{6}{10}\cdot 3 = \frac{25}{10}$$

$$E[X^2] = \frac{3}{10}\cdot 1^2 + \frac{7}{10}\cdot 2^2 = \frac{31}{10}$$

$$E[Y^2] = \frac{1}{10}\cdot 1^2 + \frac{3}{10}\cdot 2^2 + \frac{6}{10}\cdot 3^2 = \frac{67}{10}$$

$$\begin{aligned} E[XY] &= 0\cdot 1\cdot 1 + \frac{1}{10}\cdot 1\cdot 2 + \frac{2}{10}\cdot 1\cdot 3 \\ &\quad + \frac{1}{10}\cdot 2\cdot 1 + \frac{2}{10}\cdot 2\cdot 2 + \frac{4}{10}\cdot 2\cdot 3 \\ &= \frac{42}{10} \end{aligned}$$

Step3. $V[X], V[Y], Cov[X,Y]$を求める。

$$\begin{aligned} V[X] &= E[X^2] - (E[X])^2 \\ &= \frac{31}{10} - \left(\frac{17}{10}\right)^2 = \frac{21}{100} \\ V[Y] &= E[Y^2] - (E[Y])^2 \\ &= \frac{67}{10} - \left(\frac{25}{10}\right)^2 = \frac{45}{100} \\ Cov[X,Y] &= E[XY] - E[X]E[Y] \\ &= \frac{42}{10} - \frac{17}{10}\frac{25}{10} = -\frac{5}{100} \end{aligned}$$

Step4. $\rho[X,Y]$を求める。

$$\begin{aligned} \rho[X,Y] &= \frac{Cov[X,Y]}{\sqrt{V[X]}\sqrt{V[Y]}} \\ &= \frac{-\frac{5}{100}}{\sqrt{\frac{21}{100}}\sqrt{\frac{45}{100}}} = \frac{-5}{\sqrt{21}\sqrt{45}} = \underline{-0.16265} \end{aligned}$$