Department of Mathematics
College of William & Mary

MATH 451/551

Chapter 6. Joint Distribution

6.3 Covariance & Correlation

GuanNan Wang
gwang01@wm.edu

1

Theorem 6.5

If X and Y are random variables with finite population variances and covariance, then

$$V(X+Y) = V(X) + V(Y) + 2Cov(X,Y).$$

Proof: $V(X+Y) = E\left[\{(X+Y) - E(X+Y)\}^2\right]$

$$= E\left[\{(X+Y) - (\mu_X+\mu_Y)\}^2\right]$$

$$= E\left[\{(X-\mu_X) + (Y-\mu_Y)\}^2\right]$$

$$= E\left\{(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)\right\}$$

$$= E\{(X-\mu_X)^2\} + E\{(Y-\mu_Y)^2\} + 2E\{(X-\mu_X)(Y-\mu_Y)\}$$

$$= V(X) + V(Y) + 2Cov(X,Y)$$

The population variance of X+Y must be 0 because the sum of # head and # tails is always 2.

Example 1

A fair coin is tossed twice. Let X be the number of heads that appear and Y be the number of tails that appear. Find the population variance of $X + Y$.

$$V(X+Y) = V(X) + V(Y) + 2Cov(X,Y) = \frac{1}{2} + \frac{1}{2} + 2\cdot(-\frac{1}{2}) = 0$$

① $\mathcal{A} = \{(x,y) \mid (0,2), (1,1), (2,0)\}$

$$f(x,y) = \begin{cases} \frac{1}{4} & x=0, y=2 \\ \frac{1}{2} & x=1, y=1 \\ \frac{1}{4} & x=2, y=0 \end{cases} \Rightarrow f_X(x) = \begin{cases} \frac{1}{4} & x=0 \\ \frac{1}{2} & x=1 \\ \frac{1}{4} & x=2 \end{cases}, \quad f_Y(y) = \begin{cases} \frac{1}{4} & y=0 \\ \frac{1}{2} & y=1 \\ \frac{1}{4} & y=2 \end{cases}$$

② $V(X) = E(X^2) - \{E(X)\}^2 = 0^2 \times \frac{1}{4} + 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} - (0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4})^2$

$= \frac{3}{2} - 1^2 = \frac{1}{2}$

$V(Y) = \frac{1}{2}$

$Cov(X,Y) = E(XY) - \mu_X \mu_Y = 0 \cdot 2 \cdot \frac{1}{4} + 1 \cdot 1 \cdot \frac{1}{2} + 2 \cdot 0 \cdot \frac{1}{4} - 1^2 = \frac{1}{2} - 1 = -\frac{1}{2}$

Theorem

Theorem 6.6

If X and Y are independent random variables, then $Cov(X, Y) = 0$.

$Cov(X, Y) = E(XY) - \mu_X \mu_Y$.

X & Y are independent $\Rightarrow E(XY) = E(X)E(Y) = \mu_X \mu_Y$

$\Rightarrow Cov(X, Y) = 0$.

the converse of THM 6.6 is NOT true

Example 2

X & Y are indep $\Rightarrow$ Cov(X, Y) = 0

$\not\Leftarrow$

Show that the random variables X and Y that are uniformly distributed over the support $\mathcal{A} = \{(x, y) | 1 < x^2 + y^2 < 4\}$ have a population covariance 0 and are dependent random variables.

$\pi 2^2 - \pi 1^2 = 3\pi$

$f(x,y) = \frac{1}{3\pi}$, $\boxed{1 < x^2+y^2 < 4}$

$Cov(X,Y) = E(XY) - \mu_X \mu_Y$

$E(XY) = \int\int_{\mathcal{A}} xy f(x,y)\, dy\, dx$

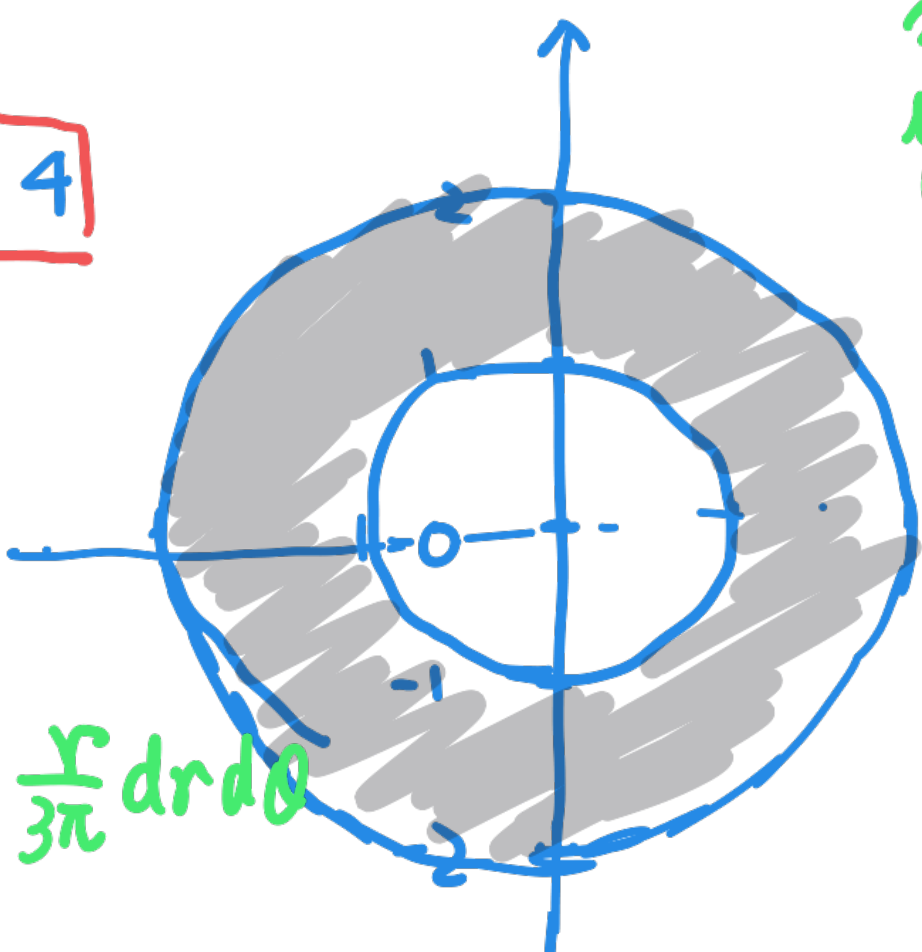
$= \int_0^{2\pi} \int_1^2 r\cos\theta \, r\sin\theta \, \frac{r}{3\pi} \, dr\, d\theta$

$= 0$

$x = r \cdot \cos\theta$

$y = r \cdot \sin\theta$

$dy\, dx = r\, dr\, d\theta$



$\mu_X = E(X) = 0$

$\mu_Y = E(Y) = 0$

$\Rightarrow Cov(X,Y) = 0$

Theorem

Theorem 6.7

If X and Y are independent random variables, then

$$V(X+Y) = V(X) + V(Y).$$

In general $V(X+Y) = V(X) + V(Y) + 2\,Cov(X,Y)$

X & Y are indep $\Rightarrow Cov(X,Y) = 0$

$\Rightarrow V(X+Y) = V(X) + V(Y)$

Correlation

$Z = \frac{X - \mu_X}{\delta_X}$

Correlation

Let X and Y be random variables with finite population means μ_X and μ_Y, and finite population variances $\sigma_X^2 > 0$ and $\sigma_Y^2 > 0$, respectively. The **population correlation** between X and Y is

$$\rho = \frac{E\{(X - \mu_X)(Y - \mu_Y)\}}{\sigma_X \sigma_Y}. \quad \frac{Cov(X, Y)}{\delta_X \delta_Y}$$

ρ_{XY}, $Corr(X, Y)$

Theorem 6.8

If X and Y are independent random variables, then $\rho = 0$.

X & Y are indep $\Rightarrow Cov(X,Y) = 0$.

$$\rho = \frac{Cov(X,Y)}{\underbrace{\delta_X \delta_Y}} = 0.$$

$$Cov\left(\frac{X}{\delta_X}, \frac{Y}{\delta_Y}\right) = E\left(\frac{XY}{\delta_X \delta_Y}\right) - \frac{\mu_X \mu_Y}{\delta_X \delta_Y} = \frac{E(XY) - \mu_X\mu_Y}{\delta_X \delta_Y} = \frac{Cov(X,Y)}{\delta_X \delta_Y}$$

Theorem 6.9

If X and Y are random variables with population correlation ρ, then $-1 \le \rho \le 1$.

$\frac{X}{\delta_X} + \frac{Y}{\delta_Y}$:

$$V\left(\frac{X}{\delta_X} + \frac{Y}{\delta_Y}\right) = V\left(\frac{X}{\delta_X}\right) + V\left(\frac{Y}{\delta_Y}\right) + 2Cov\left(\frac{X}{\delta_X}, \frac{Y}{\delta_Y}\right)$$

$$= \frac{V(X)}{\delta_X^2} + \frac{V(Y)}{\delta_Y^2} + \frac{2Cov(X,Y)}{\delta_X \delta_Y}$$

$$= 1 + 1 + 2\rho = 2(1+\rho) \ge 0 \Rightarrow \rho \ge -1$$

$\frac{X}{\delta_X} - \frac{Y}{\delta_Y}$:

$$V\left(\frac{X}{\delta_X} - \frac{Y}{\delta_Y}\right) = V\left(\frac{X}{\delta_X}\right) + V\left(\frac{Y}{\delta_Y}\right) - 2Cov\left(\frac{X}{\delta_X}, \frac{Y}{\delta_Y}\right)$$

$$= \frac{V(X)}{\delta_X^2} + \frac{V(Y)}{\delta_Y^2} - \frac{2Cov(X,Y)}{\delta_X \delta_Y}$$

$$= 1 + 1 - 2\rho \ge 0 \Rightarrow \rho \le 1$$

Example 3

Let the discrete random variables X and Y have joint probability mass function $f(x, y)$ given by the entries in the table. Find the population correlation between X and Y.

	1	2	3	$f_X(x)$
1	0.2	0.1	0.3	0.6
2	0.1	0.1	0.2	0.4
$f_Y(y)$	0.3	0.2	0.5	1

$$\rho = \frac{Cov(X,Y)}{\delta_X \delta_Y} = \frac{0.02}{\sqrt{0.24}\sqrt{0.76}} = 0.0468$$

① $\mu_X = E(X) = 1\times0.6 + 2\times0.4 = 1.4$

$\delta_X^2 = E(X^2) - \{E(X)\}^2 = 1^2\times0.6 + 2^2\times0.4 - 1.4^2 = 0.24$

② $\mu_Y = E(Y) = 1\times0.3 + 2\times0.2 + 3\times0.5 = 2.2$

$\delta_Y^2 = E(Y^2) - \{E(Y)\}^2 = 1^2\times0.3 + 2^2\times0.2 + 3^2\times0.5 - 2.2^2 = 0.76$

③ $Cov(X,Y) = E(XY) - \mu_X\mu_Y = 1\cdot1\cdot0.2 + 1\cdot2\cdot0.1 + 1\cdot3\cdot0.3 + 2\cdot1\cdot0.1 + 2\cdot2\cdot0.1 + 2\cdot3\cdot0.2$
$- 1.4\times2.2 = 0.02$

Theorem

Theorem 6.10

Let X and Y be random variables with population correlation ρ. The population coorrelation ρ equals -1 iff the support of X and Y lies on a line with negative slope. The population correlation ρ equals 1 iff the support of X and Y lies on a line with positive slope.

- A population correlatioon of -1 between the random variables X and Y is often known as a **perfect negative correlation**.
- A population correlation oof 1 between the random variable X and Y is often known as a **perfect positive correlation**.

Example 4

A fair coin is tossed twice. Let X be the number of heads that appear and Y be the number of tails that appear. Find the population correlation between X and Y.

$\mathcal{A} = \{(x,y) \mid (0,2),(1,1),(2,0)\}$

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{4} & x=0 \quad y=2 \\ \frac{1}{2} & x=1 \quad y=1 \\ \frac{1}{4} & x=2 \quad y=0 \end{cases}$$

① $\mu_X = E(X) = 1$

$\delta_X^2 = E(X^2) - \{E(X)\}^2 = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} - 1$

$= \frac{1}{2}$

② $\mu_Y = E(Y) = 1$

$\delta_Y^2 = \frac{1}{2}$

③ $Cov(X,Y) = E(XY) - \mu_X\mu_Y = -\frac{1}{2}$

$$\rho = \frac{Cov(X,Y)}{\delta_X \delta_Y} = \frac{-\frac{1}{2}}{\sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}}} = -1$$

$Y = 2 - X$

Thank You

THANK YOU!