Joint Variables (II)

Goal: Study expected values and variation when dealing with more than one r.v.

Suppose X, Y are two random variables

either both discrete or both continuous.

Let $p(x,y) \rightsquigarrow$ "joint pmf of X,Y" if both discrete

$f(x,y) \rightsquigarrow$ "joint pdf of X and Y" if both are continuous.

If $h(x,y)$ is a function of X and Y

can define the expected value of $h(x,y)$

$$\boxed{E(h(x,y)) = \begin{cases} \displaystyle\sum_{(x,y)\in \mathcal{X}\times\mathcal{Y}} h(x,y)\cdot p(x,y) & \text{if } X,Y \text{ are discrete} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\, f(x,y) & \text{if } X,Y \text{ are both continuous.} \end{cases}}$$

Covariance and Correlation

Want: A way to measure relationship betwn X and Y

Let

$\mu_X = E(X) \rightsquigarrow$ expected value of X

$\mu_Y = E(Y) \rightsquigarrow$ expected value of Y.

Let

$$\boxed{\begin{aligned} h(x,y) &= (X-\mu_X)(Y-\mu_Y) \\ &= \text{"product of the deviation of } X \text{ and } Y \text{ from their respective expected values"} \end{aligned}}$$

Note:

① If large values of X are associated with large values of Y $\rightsquigarrow$ ie $\boxed{(X-\mu_X)\geq 0 \Rightarrow (Y-\mu_Y)\geq 0}$

and

Small values of X are associated with small values of Y $\rightsquigarrow$ ie $\boxed{(X-\mu_X)\leq 0 \Rightarrow (Y-\mu_Y)\leq 0}$

We have

$$\boxed{h(x,y) = (X-\mu_X)(Y-\mu_Y) \geq 0}$$

as terms on right are both positive or both negative

② If large X values are associated with small Y values

ie $\boxed{(X-\mu_X)\geq 0 \Rightarrow (Y-\mu_Y)\leq 0}$

and

large Y values are associated with small X values

ie $\boxed{(Y-\mu_Y)\geq 0 \Rightarrow (X-\mu_X)\leq 0}$

We will have:

$$\boxed{h(x,y) = (X-\mu_X)\cdot(Y-\mu_Y) \leq 0}$$

since terms on the right are of opposite sign.

$\therefore$ $\boxed{\text{The expected value of } h(X,Y) = (X-\mu_X)(Y-\mu_Y)} \rightsquigarrow$ a good candidate to measure relationship betwn X and Y.

Define the Covariance of X and Y:

$$\boxed{\begin{aligned} Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E(Y). \\ &= \text{"the expected value of the product of deviation of } X \text{ from } \mu_X \text{ and } Y \text{ from } \mu_Y\text{"} \end{aligned}}$$

$$\boxed{Cov(X,Y) = \begin{cases} \displaystyle\sum_x\sum_y (x-\mu_X)(y-\mu_Y)\, p(x,y) & \text{if } X,Y \text{ are discrete} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)\, f(x,y)\,dx\,dy & \text{if } X,Y \text{ are continuous} \end{cases}}$$

Note:

① If $X=Y$, $\boxed{Cov(X,Y) = Cov(X,X) = V(X)}$

② If X and Y have "strong positive relationship" $\rightsquigarrow$ large values of X associated with large values of Y and small values of X are associated with small values of Y.

Cov(X,Y) will be positive.

② If X,Y have a "strong negative relationship" $\rightsquigarrow$ large values of X associated with small values of Y and large values of Y associated with small values of X

Cov(X,Y) will be negative.

③ If $Cov(X,Y)$ is positive $\rightsquigarrow$ can say X and Y are positively associated

$Cov(X,Y)$ is negative $\rightsquigarrow$ X and Y are negatively associated.

Not a good measure of association!

depends on the units in which X and Y are measured

can be made arbitrarily large/small by changinging units.

The Correlation Coefficient of X and Y

Let $\sigma_X = \sqrt{V(X)}$ and $\sigma_Y = \sqrt{V(Y)}$,

To make $Cov(X,Y)$ unitless $\rightsquigarrow$ divide by $\sigma_X \cdot \sigma_Y$

$\therefore$ The correlation of X and Y is

$$\boxed{\rho_{X,Y} = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y}}$$

Note:

① $\rho_{X,Y}$ does not depend on the units of X and Y.

② $\rho_{X,Y}$ satisfies $\boxed{-1 \leq \rho_{X,Y} \leq 1.}$

③ If a, c are both positive or both negative

$$\boxed{Corr(aX+b,\, cY+d) = Corr(X,Y)}$$

In particular, if $Y = aX+b$

$$\begin{aligned} Corr(X,Y) &= Corr(X, aX+b) = sign(a)\cdot Corr(X,X) \\ &= sign(a) = \pm 1. \end{aligned}$$

$\therefore$ $\boxed{Corr(X,Y) \rightsquigarrow \text{measures the degree of linear relationship betwn } X \text{ and } Y.}$

If $\rho_{X,Y}$ close to 1 $\rightsquigarrow$ strong positive linear association betwn X and Y.

close to -1 $\rightsquigarrow$ strong -ve linear association.

$\boxed{\rho_{X,Y}=0} \rightsquigarrow$ X,Y are uncorrelated $\rightsquigarrow$ no linear relationship.

does not mean that there is no relationship.

④ $\boxed{\text{If } X,Y \text{ are independent} \Rightarrow Cov(X,Y)=0 \Rightarrow Corr(X,Y)=0.}$

$Corr(X,Y)=0$ does not necessarily imply that X and Y are independent!

⑤ Good old adage:

$\boxed{\text{Correlation does not imply Causation!}}$

only implies association betwn variables

possible that there is an underlying variable that is "causing" the positive association.

⑥ Note:

$$\boxed{\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y} \Rightarrow Cov(X,Y) = \rho_{X,Y}\cdot\sigma_X\cdot\sigma_Y.}$$

Linear Combination of Random Variables

Suppose $X_1, X_2, X_3, \ldots, X_n$ are n-random variables such that:

$$\boxed{\begin{aligned} &(i)\ E(X_i) = \mu_i && i=1,2,3,\ldots,n \\ &(ii)\ V(X_i) = \sigma_i^2 && i=1,2,\ldots,n. \end{aligned}}$$

Let $a_1, a_2, \ldots, a_n \in \mathbb{R}$, can define the "linear combination" as

$$\boxed{Y = a_1X_1 + a_2X_2 + a_3X_3 + \cdots + a_nX_n}$$

What is $E(Y)$? $V(X)$?

Thm:

$$\boxed{\begin{aligned} (i)\ E(Y) &= E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) \\ &= a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n \end{aligned}}$$

$$\boxed{\begin{aligned} (ii)\ V(Y) &= V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j\, Cov(X_i,X_j) \end{aligned}}$$

In particular, if $X_1, X_2, X_3, \ldots, X_n$ are independent

$$\boxed{\begin{aligned} V(Y) &= a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n) \\ &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2 \end{aligned}}$$

For two random variables

$Z = aX+bY$, then

① $\boxed{E(Z) = a\mu_X + b\mu_Y}$

② $\boxed{V(Z) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,Cov(X,Y).}$

If $a=1$, $b=-1$; $Z = X-Y$

then

$$\boxed{\begin{aligned} E(X-Y) &= \mu_X - \mu_Y, \\ V(X-Y) &= \sigma_X^2 + \sigma_Y^2 + 2\rho_{X,Y}\cdot\sigma_X\cdot\sigma_Y. \end{aligned}}$$

Given $p(x,y) \longrightarrow$ joint pmf of X and Y.

$$h: \mathcal{X}\times\mathcal{Y} \longrightarrow \mathbb{R}$$
$$(x,y) \longmapsto h(x,y)$$

$$Z = h(X,Y)$$

$$E(Z) = ?$$

$$E(h(x,y)) = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} h(x,y)\cdot p(x,y)$$

$$= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} h(x,y)\cdot p(x,y)$$

If

① $h(X,Y) = X$
$\Downarrow$
$E(h(X,Y)) = E(X) = \mu_X$

② $\underline{h(X,Y) = Y} \longrightarrow E(h(X,Y)) = E(Y) = \mu_Y$

Let $\mu_X = E(X)$, $\mu_Y = E(Y)$

Define: $h(X,Y) = (X-\mu_X)(Y-\mu_Y)$

= product of the deviations of X and Y from their respective means.

large values of X are associated with large values of Y
$(X-\mu_X)\geq 0 \Rightarrow (Y-\mu_Y)\geq 0$
and
small values X are associated with small values of Y
$(X-\mu_X)\leq 0 \Rightarrow (Y-\mu_Y)\leq 0$

$\Downarrow$

$h(X,Y) = (X-\mu_X)(Y-\mu_Y) \geq 0$

$\Downarrow$

$\underline{E(h(X,Y)) \geq 0}$

$$\mathrm{Cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y))$$

= Expected value of the product of the deviations.

$\Downarrow$

If $\underline{\underline{\mathrm{Cov}(X,Y) \text{ is positive}}} \longrightarrow$ positive association between X and Y.

$\mathrm{Cov}(X,Y)$ is negative $\longrightarrow$ negative association between X and Y.

Note: $\mathrm{Cov}(X,Y)$ depends on the units in which X and Y are measured.

$\Downarrow$

units of $\mathrm{Cov}(X,Y)$ = product of the units of X and Y.

$$\boxed{\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}} = \rho_{X,Y}$$

$\Downarrow$

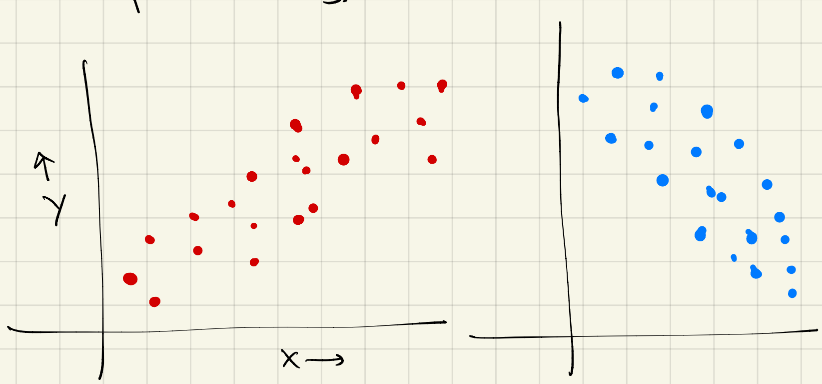
$\rho_{X,Y} \longrightarrow$ measures the extent of "linear relationship" between X and Y.

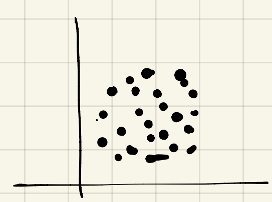
$-1 \leq \rho_{X,Y} \leq 1 \longrightarrow \rho_{X,Y} = \pm 1 \Longleftrightarrow Y = aX + b$

$\Downarrow$

perfect linear relationship between X and Y.

Scatter Plots $\rightsquigarrow (x,y)$ in $\mathbb{R}^2$





Date: April 9th:

Recall from earlier:

Joint dist $p(x,y)$

$x \backslash y$	1	2	3	4	5	6
0	$(1-p)\cdot(0.1)$	$(1-p)\,0.1$	$(1-p)\cdot 0.1$	$(1-p)\cdot 0.1$	$(1-p)\cdot 0.1$	$(1-p)(0.5)$
1	$p\cdot 1/6$	$p\cdot 1/6$	$p\cdot 1/6$	$p\cdot 1/6$	$p\cdot(1/6)$	$p\cdot(1/6)$

$\mathcal{Y} = \{1,2,3,4,5,6\}$

$p_Y(1) = p(0,1) + p(1,1) = (1-p)\cdot 0.1 + p\cdot\frac{1}{6}$

$p_Y(2) = p(0,2) + p(1,2) = (1-p)(0.1) + p\cdot\frac{1}{6}$

$p_Y(3) = p(0,3) + p(1,3) = (1-p)(0.1) + p\cdot\frac{1}{6}$

$p_Y(4) = \qquad =$

$p_Y(5) = \qquad =$

$p_Y(6) = p(0,6) + p(1,6) = (1-p)(0.5) + p\cdot\frac{1}{6}$.

x	0	1
$p_X(x)$	$1-p$	p

Y	1	2	. . . .	.	6
$p_Y(y)$	$(1-p)\,0.1 + p\cdot\frac{1}{6}$	$(1-p)\cdot 0.1 + p\cdot\frac{1}{6}$			$(1-p)(0.5) + p\cdot\frac{1}{6}$
if $p=\frac{1}{2}$	$\frac{1}{2}\cdot 0.1 + \frac{1}{2}\cdot\frac{1}{6}$	$\frac{1}{2}\cdot 0.1 + \frac{1}{12}$	$\frac{1}{2}(0.1) + \frac{1}{12}$	. .	$\frac{1}{2}\cdot 0.5 + \frac{1}{12}$

Want to calculate $Cov(X,Y)$

$$Cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

$$= E(XY) - E(X)\cdot E(Y)$$

$$E(X) = \sum_{x\in\mathcal{X}} x\cdot p_X(x) = 0\cdot p_X(0) + 1\cdot p_X(1)$$

$$= p = \mu_X$$

$$E(Y) = \sum_{y\in\mathcal{Y}} y\cdot p_Y(y)$$

$$= 1\cdot\left((1-p)\cdot 0.1 + p\tfrac{1}{6}\right) + 2\cdot\left((1-p)\cdot(0.1) + p\cdot\tfrac{1}{6}\right) + 3\cdot\left((1-p)(0.1) + p\tfrac{1}{6}\right)$$

$$+4\left((1-p)(0.1) + p\tfrac{1}{6}\right) + 5\left((1-p)\cdot(0.1) + p\cdot\tfrac{1}{6}\right) + 6\cdot\left((1-p)\cdot 0.5 + p\cdot\tfrac{1}{6}\right)$$

$$= (1-p)\cdot 0.1\,(1+2+3+4+5) + p\cdot\tfrac{1}{6}(1+2+3+4+5+6) + 6\cdot(1-p)(0.5)$$

$$= (1-p)\cdot(0.1)\left(\frac{5\times\cancel{6}^{\,3}}{\cancel{2}}\right) + p\cdot\tfrac{1}{6}\cdot\left(\frac{6\times 7}{2}\right) + 6\cdot(1-p)\cdot(0.5)$$

$$= (1-p)(0.1)(15) + p\cdot\tfrac{7}{2} + 6\cdot(1-p)\cdot 0.5$$

$$= (1-p)(0.1\times 15 + 6\times 0.5) + (3.5)\cdot p = (4.5)\cdot(1-p) + (3.5)p = 4.5 - (4.5p) + (3.5)p$$

$$= (1-p)\cdot\left(\tfrac{1}{10}\times 15 + \tfrac{30}{10}\right) + (3.5)p \qquad\qquad = 4.5 - p$$

Summarize

$E(X) = p, \quad E(Y) = 4.5 - p$

$$E(XY) = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} x\cdot y\cdot p(x,y)$$

$$= \sum_{y\in\mathcal{Y}} x\cdot y\cdot p(1,y)$$

$$= 1\cdot p\cdot\tfrac{1}{6} + 2\cdot p\cdot\tfrac{1}{6} + 3\cdot p\cdot\tfrac{1}{6} + \cdots + 6\cdot p\cdot\tfrac{1}{6}$$

$$= \left(p\cdot\tfrac{1}{6}\right)(1+2+3+\cdots+6) = p\cdot\frac{1}{\cancel{6}}\times\frac{\cancel{6}\times 7}{2}$$

$$= (3.5)\cdot p$$

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

$$= (3.5)\cdot p - \left(p\cdot(4.5-p)\right)$$

$$= 3.5p - 4.5p + p^2$$

$$= p^2 - p = p(p-1)$$

Since p satisfies $0<p<1 \Rightarrow (p-1)<0$

$\Rightarrow Cov(X,Y) = p(p-1) < 0$

$\Rightarrow$ X and Y are negatively associated.

Questions: (1) What if we rolled a fair die irrespective of the outcome of the coin toss?

(2) What if $p = \frac{1}{2}$.

Correllation Coefficient :

$$S_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

$\therefore$ We need to calculate σ_X and σ_X.

$$Var(X) = p \cdot (1-p) \qquad Var(Y) = E(Y^2) - (E(Y))^2$$

$$E(Y^2) = 1 \cdot 1 + 2^2 \cdot p_Y(2) + 3^2 \cdot p_Y(3) + 4^2 p_Y(4) + 5^2 p_Y(5) + 6^2 p_Y(6)$$

$$= 1^2 \cdot \left((1-p) \cdot 0.1 + p \cdot \tfrac{1}{6}\right) + 2^2 \cdot \left((1-p) \cdot (0.1) + p \cdot \tfrac{1}{6}\right) + 3^2 \left((1-p) \cdot (0.1) + p \cdot \tfrac{1}{6}\right)$$

$$+ 4^2 \left((1-p)(0.1) + p \cdot \tfrac{1}{6}\right) + 5^2 \left((1-p) \cdot (0.1) + p \cdot \tfrac{1}{6}\right) + 6^2 \left((1-p) \cdot 0.5 + p \cdot \tfrac{1}{6}\right)$$

$$= (1-p) \cdot 0.1 \left(1^2 + 2^2 + 3^2 + 4^2 + 5^2\right) + p \cdot \tfrac{1}{6}\left(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2\right) + 6^2 \cdot (1-p) \cdot (0.5)$$

$$1 + 2^2 + 3^2 + 4^2 + \cdots + n^2 \stackrel{?}{=} \frac{n(n+1)(2n+1)}{6}$$

$$= (1-p)(0.1) \cdot \left(\frac{5 \times \cancel{6} \times 11}{\cancel{6}}\right) + p \cdot \tfrac{1}{6}\left(\frac{\cancel{6} \times 7 \times 13}{\cancel{6}}\right) + 36(1-p)(0.5)$$

$$= (1-p)(0.1)(55) + p\left(\tfrac{91}{6}\right) + 18 \cdot (1-p)$$

$$= (1-p)(5.5 + 18) + p \cdot \left(\tfrac{91}{6}\right)$$

$$= (1-p)(23.5) + p\left(\tfrac{91}{6}\right)$$

$$= 23.5 - (8.333)\, p.$$

$$\therefore V(Y) = E(Y^2) - (E(Y))^2$$

$$= (23.5 - 8.33p) - (4.5 - p)^2$$

$$\therefore Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{p \cdot (1-p)} \cdot \sqrt{V(Y)}}$$