Joint Variables (II)

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

Suppose X, Y are two random variables
$\Updownarrow$
either both discrete or both continuous.

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

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

If $h(x,y)$ is a function of X and Y
$\Updownarrow$
can define the expected value of $h(x,y)$

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

Covariance and Correlation

Want: A way to measure relationship betwn X and Y.

Let

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

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

and

large Y values are associated with small X values
$\Updownarrow$
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:

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

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

Note:

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

② If X and Y have "strong positive relationship" $\longrightarrow$ large values of X associated with large values of Y and small values of X are associated with small values of Y.
$\Updownarrow$
$\mathrm{Cov}(X,Y)$ will be positive.

③ If X,Y have a "strong negative relationship" $\longrightarrow$ large values of X associated with small values of Y and large values of Y associated with small values of X
$\Updownarrow$
$\mathrm{Cov}(X,Y)$ will be negative.

③ If $\mathrm{Cov}(X,Y)$ is positive $\longrightarrow$ can say X and Y are positively associated.

$\mathrm{Cov}(X,Y)$ is negative $\longrightarrow$ X and Y are negatively associated.

$\Updownarrow$

Not a good measure of association!
depends on the units in which X and Y are measured
$\Updownarrow$
Can be made arbitrarily large/small by changing units.

The Correlation Coefficient of X and Y

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

To make $\mathrm{Cov}(X,Y)$ unitless $\longrightarrow$ divide by $\sigma_X\cdot\sigma_Y$

$\therefore$ The correlation of X and Y is

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

Note:

① ρ_{XY} does not depend on the units of X and Y.

② ρ_{XY} satisfies $\boxed{-1\leq \rho_{XY}\leq 1.}$

③ If a, c are both positive or both negative

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

In particular, if $Y = aX+b$

$\mathrm{Corr}(X,Y) = \mathrm{Corr}(X, aX+b) = \mathrm{sign}(a)\cdot\mathrm{Corr}(X,X)$

$= \mathrm{sign}(a) = \pm 1$

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

If ρ_{XY} close to $1 \longrightarrow$ strong positive linear association betwn X and Y.

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

$\boxed{\rho_{XY}=0} \longrightarrow X,Y$ are uncorrelated $\longrightarrow$ no linear relationship.
$\Updownarrow$
does not mean that there is no relationship.

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

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

⑤ Good old adage:

$\boxed{\text{Correlation does not imply Causation!}}$
$\Updownarrow$
only implies association betwn variables
$\Updownarrow$
possible that there is an underlying variable that is "causing" the positive association.

⑥ Note:

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

Linear Combination of Random Variables

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

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

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

$$\boxed{Y = a_1X_1 + a_2X_2 + a_3X_3 + \cdots + a_nX_n}$$
$\Updownarrow$
What is $E(Y)$? $V(Y)$?

Thm:

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

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

$$\boxed{\begin{aligned} &\text{In particular, if } X_1, X_2, X_3, \ldots, X_n \text{ are independent} \\ &V(Y) = a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n) \\ &\quad\;\; = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2 \end{aligned}}$$

For two random variables

$Z = aX + bY$, then

$\boxed{\text{① } E(Z) = a\mu_X + b\mu_Y}$

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

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

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