Lecture - 6 (1)

Expected Values :- Let X and Y be jointly r.v with pmf $p_{X,Y}(x,y)$ or pdf $f_{X,Y}(x,y)$ then the expected values of a function $h(X,Y)$ is denoted as $E(h(X,Y))$ and defined as

$$E(h(X,Y)) = \begin{cases} \sum\limits_{x \in R_X} \sum\limits_{y \in R_Y} h(x,y) \cdot p_{X,Y}(x,y); & \text{discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) \cdot f_{X,Y}(x,y)\, dx\, dy; & \text{Continuous} \end{cases}$$

Covariance : In covariance, we are discussing that, how two r.v X and Y related.

(*) If covariance between two r.v is positive, it means they are positively related. That is by increasing one variable, other is also increased.

e.g population and pollution are positively related.

(**) If covariance between two r.v is negative, it means they are negatively related i.e by increasing one variable other is decreases.

e.g plantation and pollution are negatively related.

Def^n

(2)

Let X and Y be the r.v then covariance between X and Y is denoted as $Cov(X,Y)$ and defined as

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$$

$$= E[XY - X\underbrace{E(Y)}_{constant} - Y\underbrace{E(X)}_{constant} + \underbrace{E(X).E(Y)}_{constant}]$$

$\because$ $E(X) = \begin{cases} \sum x \cdot p_X(x) ; & \text{discrete} \\ \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx ; & \text{continuous} \end{cases}$

and $E(Y)$ are constant values and $E(aX+b) = aE(X)+b$ then

$$Cov(X,Y) = E(XY) - E(X).E(Y) - \cancel{E(Y).E(X)} + \cancel{E(X).E(Y)}$$

$$\boxed{Cov(X,Y) = E(XY) - E(X).E(Y)}$$

See Example 5.16 ; (in book)

Note! If $cov(X,Y) > 0 \rightarrow$ positively related

$< 0 \rightarrow$ Negatively related

Note: If X and Y are independent then

$$E(XY) = E(X).E(Y).$$

In this case $cov(X,Y) = 0$. But converse of this statement need not be true. i.e if $cov(X,Y) = 0$ then X and Y may be dependent.

Ex Let $X \sim U[-1, 1]$ (Uniform distn)
and $Y = X^2$ then we need to show that
$cov(X,Y) = 0$ but X and Y are not independent

Soln $\because X \sim U[-1,1]$

$f_X(x) = \frac{1}{1-(-1)} = \frac{1}{2}$; $x \in [-1,1]$

$X \sim U[a,b]$
$f_X(x) = \frac{1}{b-a}$; $a<x<b$
$E(X) = \frac{a+b}{2}$
$V(X) = \frac{(b-a)^2}{12}$

$E(X) = \frac{1+(-1)}{2} = 0$

$cov(X,Y) = E XY - E(X).E(Y)$

($\because Y = X^2$)

Now; $E XY = E(X \cdot X^2)$
$= E X^3 = \int_{-1}^{1} x^3 \cdot \frac{1}{2} dx = 0$
↳ odd integral

Thus $Cov(X,Y) = 0 - 0 . EY$
$= 0$

but X and Y are not independent as they are written in terms of each other. i.e $Y = X^2$.

Correlation: The correlation or correlation coefficient between two r.v X and Y is defined as

$$Corr(X,Y) \text{ or } \rho_{X,Y} = \frac{cov(X,Y)}{\sigma_x . \sigma_y}$$

σ_x → standard deviation of X
σ_y → standard deviation of Y

$Corr(X,Y) = 1$ → perfectly linear correlated in +ve direction
$= -1$ → perfectly linear correlation in -ve direction

Scanned by CamScanner

Example 5.15 The joint and marginal pmf's for X = automobile policy deductible amount and Y = homeowner policy deductible amount in Example 5.1 were

$p(x, y)$		y = 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

x	100	250
$p_X(x)$	.5	.5

y	0	100	200
$p_Y(y)$	.25	.25	.5

from which $\mu_X = \Sigma x p_X(x) = 175$ and $\mu_Y = 125$. Therefore,

$$\begin{aligned}
\text{Cov}(X, Y) &= \sum_{(x,y)}\sum (x - 175)(y - 125)p(x, y) \\
&= (100 - 175)(0 - 125)(.20) + \cdots \\
&\quad + (250 - 175)(200 - 125)(.30) \\
&= 1875
\end{aligned}$$

Example 5.16 *(Example 5.5 continued)* The joint and marginal pdf's of X = amount of almonds and Y = amount of cashews were

$$f(x, y) = \begin{cases} 24xy & 0 \le x \le 1, 0 \le y \le 1, x + y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = \begin{cases} 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

with $f_Y(y)$ obtained by replacing x by y in $f_X(x)$. It is easily verified that $\mu_X = \mu_Y = \frac{2}{5}$, and

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x, y)\, dx\, dy = \int_0^1 \int_0^{1-x} xy \cdot 24xy\, dy\, dx$$

$$= 8\int_0^1 x^2(1-x)^3\, dx = \frac{2}{15}$$

Thus $\text{Cov}(X, Y) = \frac{2}{15} - \left(\frac{2}{5}\right)\left(\frac{2}{5}\right) = \frac{2}{15} - \frac{4}{25} = -\frac{2}{75}$. A negative covariance is reasonable here because more almonds in the can implies fewer cashews. ■

It might appear that the relationship in the insurance example is quite strong since $\text{Cov}(X, Y) = 1875$, whereas $\text{Cov}(X, Y) = -\frac{2}{75}$ in the nut example would seem to imply quite a weak relationship. Unfortunately, the covariance has a serious defect that makes it impossible to interpret a computed value. In the insurance example, suppose we had expressed the deductible amount in cents rather than in dollars. Then $100X$ would replace X, $100Y$ would replace Y, and the resulting covariance would be $\text{Cov}(100X, 100Y) = (100)(100)\text{Cov}(X, Y) = 18{,}750{,}000$. If, on the other hand, the deductible amount had been expressed in hundreds of dollars, the computed covariance would have been $(.01)(.01)(1875) = .1875$. *The defect of covariance is that its computed value depends critically on the units of measurement.* Ideally, the choice of units should have no effect on a measure of strength of relationship. This is achieved by scaling the covariance.