Bivariate Expectation and Covariance

Let (X,Y) be a 2D random variable with join pmf $P(X=i, Y=j) = p_{ij}$, $i = 1,2,\dots,n$
$j = 1,2,\dots,m$

Then

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X=i) = \sum_{i=1}^{n} x_i p_{i.}$$

$$E(Y) = \sum_{j=1}^{m} y_j P(Y=j) = \sum_{j=1}^{m} y_j p_{.j}$$

$$E(X^2) = \sum_{i=1}^{n} x_i^2 p_{i.} \; ; \; E(Y^2) = \sum_{j=1}^{m} y_j^2 p_{.j}$$

$$\therefore E(g(X)) = \sum_{i=1}^{n} g(x_i) p_{i.}$$

$$E(h(Y)) = \sum_{j=1}^{m} h(y_j) p_{.j}$$

$$E(g(X,Y)) = \sum_{j=1}^{m}\sum_{i=1}^{n} g(x_i, y_j) p_{ij}$$

$$Var(X) = E(X^2) - \{E(X)\}^2$$

$$Var(Y) = E(Y^2) - \{E(Y)\}^2$$

$$E(X+Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i + y_j) p_{ij}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i p_{ij} + \sum_{j=1}^{m}\sum_{i=1}^{n} y_j p_{ij}$$

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{m} p_{ij} + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} p_{ij}$$

$$= \sum_{i=1}^{n} x_i p_{i.} + \sum_{j=1}^{m} y_j p_{.j}$$

$$\boxed{E(X+Y) = E(X) + E(Y).}$$

Hence, $\boxed{E(aX+bY) = aE(X) + bE(Y)}$

Covariance of two random variable X and Y is given by

$$cov(X,Y) = E\left[(X-E(X))(Y-E(Y))\right]$$

$$= E\left[XY - XE(Y) - YE(X) + E(X)E(Y)\right]$$

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

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

If X and Y are two independent random variables
Then $E(XY) = E(X)E(Y)$

Hence $cov(X,Y) = 0$.

If X and Y are two independent random variables, then $E(XY) = E(X)E(Y)$.

Proof: $E(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_{ij}$

$= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_{i.} \times p_{.j}$ | as X and Y are two ind. r.v.s. $p_{ij} = p_{i.} \times p_{.j}$

$= \left(\sum_{i=1}^{n} x_i p_{i.}\right)\left(\sum_{j=1}^{m} y_j p_{.j}\right)$

$= E(X)E(Y)$

Proved

Correlation Coefficient

The correlation coefficient between X and Y is denoted by ρ_{XY} or r_{XY} and is defined by

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

where σ_X and σ_Y are standard deviation of X and Y, respectively.

i.e., $\sigma_X^2 = Var(X)$, $\sigma_Y^2 = Var(Y)$.

Result 1. The correlation coefficient is independent of origin and scale of the variables.

i.e if $U = aX + b$, $V = cY + d$. where a, b, c, d are constants, then ~~$\rho_{XY} = \pm \rho_{UV}$~~. $\rho_{UV} = \pm \rho_{XY}$.

Soln. $E(X) = \bar{x}$, $E(Y) = \bar{y}$

$$Var(X) = E(X-\bar{x})^2 = \sigma_X^2$$

$$Var(Y) = E(Y-\bar{y})^2 = \sigma_Y^2.$$

Let $Cov(X,Y) = E[(X-\bar{x})(Y-\bar{y})]$

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y}.$$

Now $U = aX+b \Rightarrow E(U) = \bar{u} = a\bar{x} + b$

$V = cY + d \Rightarrow E(V) = \bar{v} = c\bar{y} + d$

$$\begin{aligned} Cov(U,V) &= E[(U-\bar{u})(V-\bar{v})] \\ &= E[(aX+b-a\bar{x}-b)(cY+d-c\bar{y}-d)] \\ &= ac\,E[(X-\bar{x})(Y-\bar{y})] \\ &= ac\,Cov(X,Y). \end{aligned}$$

$$\sigma_U^2 = E(U-\bar{u})^2 = E(aX+b-a\bar{x}-b)^2$$
$$= a^2 E(X-\bar{x})^2 = a^2\sigma_X^2 .$$

$\Rightarrow \sigma_U = |a|\,\sigma_X$

Similarly, $\sigma_V = |c|\,\sigma_Y$

$$\therefore \rho_{UV} = \frac{cov(U,V)}{\sigma_U \sigma_V} = \frac{ca\,cov(X,Y)}{|a|\sigma_X\,|c|\sigma_Y}$$

$$\boxed{\rho_{UV} = \frac{ac}{|a||c|}\rho_{XY}}$$

Let $a>0, c>0$ then $\rho_{UV} = \rho_{XY}$

Let $a<0, c>0$ or $a>0, c<0$ then $\rho_{UV} = -\rho_{XY}$

Let $a<0, c<$ then $\rho_{UV} = \rho_{XY}$

Hence $\rho_{UV} = \pm\rho_{XY}$.

Q. Prove that the correlation coefficient ρ_{XY} between two jointly distributed random variables X and Y lies between -1 and $+1$, i.e., $-1 \le \rho_{XY} \le 1$.

Soln. Let $E(X) = \bar{x}$, $E(Y) = \bar{y}$

$Var(X) = \sigma_X^2$, $Var(Y) = \sigma_Y^2$

Then $\{(X-\bar{x}) + k(Y-\bar{y})\}^2 \geq 0$ for any real k.

$\Rightarrow E\{(X-\bar{x}) + k(Y-\bar{y})\}^2 \geq 0.$

$\Rightarrow E\left[(X-\bar{x})^2 + 2k(X-\bar{x})(Y-\bar{y}) + k^2(Y-\bar{y})^2\right] \geq 0$

$\Rightarrow E(X-\bar{x})^2 + 2k\,E\left[(X-\bar{x})(Y-\bar{y})\right] + k^2 E(Y-\bar{y})^2 \geq 0$

$\Rightarrow \sigma_X^2 + k^2\sigma_Y^2 + 2k\,cov(X,Y) \geq 0$

$\Rightarrow \sigma_x^2 + k^2\sigma_y^2 + 2k\rho_{xy}\sigma_x\sigma_y \geq 0$.

$\Rightarrow \sigma_x^2 + 2k\rho_{xy}\sigma_x\sigma_y + k^2\rho_{xy}^2\sigma_y^2 + k^2\sigma_y^2 - k^2\rho_{xy}^2\sigma_y^2 \geq 0$

$\Rightarrow (\sigma_x + k\rho_{xy}\sigma_y)^2 + k^2\sigma_y^2(1-\rho_{xy}^2) \geq 0$.

$\Rightarrow \quad A + B \geq 0$ — A is always non-neg term then min $A = 0$.

Now $A = 0 \Rightarrow k = -\dfrac{\sigma_x}{\rho_{xy}\sigma_y}$

At, $k = -\dfrac{\sigma_x}{\sigma_y\rho_{xy}}$, $k^2\sigma_y^2(1-\rho_{xy}^2) \geq 0$

$\Rightarrow 1 - \rho_{xy}^2 \geq 0$

$\Rightarrow 1 \geq \rho_{xy}^2$

$\Rightarrow -1 \leq \rho_{xy} \leq 1$.

$$\frac{(\sigma_x + k\rho_{xy}\sigma_y)^2}{k^2\sigma_y^2} + (1-\rho_{xy}^2) \geq 0$$

$$\Rightarrow \underbrace{\left(\frac{\sigma_x}{k\sigma_y} + \rho_{xy}\right)^2}_{=0} + (1-\rho_{xy}^2) \geq 0.$$

$1 - \rho_{xy}^2 \geq 0$

$\Rightarrow -1 \leq \rho_{xy} \leq 1$.

ABanerjee.
Lecture 11 : P(5)

$A + B \geq 0$

$A \geq 0$ (always true)

$\hookrightarrow 0 \leq A < |1-\rho_{xy}^2| \cup |1-\rho_{xy}^2| \leq A$

In this range $A+B$ may be <0 if $1-\rho_{xy}^2 \not< 0$ $\therefore 1-\rho_{xy}^2 \geq 0$

In this range $A+B$ always ≥ 0. whether $1-\rho_{xy}^2 < 0$ or $1-\rho_{xy}^2 \geq 0$

Theorem: $Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\, Cov(X,Y)$.

Let $Z = aX + bY$

$E(Z) = a E(X) + b E(Y) = a\bar{x} + b\bar{y}$

$E(Z^2) = E(a^2X^2 + 2abXY + b^2Y^2)$
$= a^2 E(X^2) + 2ab E(XY) + b^2 E(Y^2)$

$Var(Z) = E(Z^2) - \{E(Z)\}^2$

$= a^2 E(X^2) + 2ab E(XY) + b^2 E(Y^2) - a^2\bar{x}^2 - b^2\bar{y}^2 - 2ab\,\bar{x}\bar{y}$

$= a^2 [E(X^2) - \bar{x}^2] + b^2 [E(Y^2) - \bar{y}^2] + 2ab[E(XY) - \bar{x}\bar{y}]$

$= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\, Cov(X,Y)$.

$$\boxed{Var(aX+bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\rho_{XY}\,\sigma_X\sigma_Y}$$

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

$\Rightarrow Var(aX+bY) = a^2\sigma_X^2 + b^2\sigma_Y^2$.

Q. If $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ and $U = X+Y$, $V = X-Y$, then find ρ_{UV}. Ans. $\rho_{UV} = 0$

Q. Let X and Y are two random variables with s.d. σ_X and σ_Y, then find ρ_{UV}, where $U = X$ and $V = X+Y$.

Ans. $\rho_{UV} = \dfrac{\sigma_X}{\sqrt{\sigma_X^2 + \sigma_Y^2}}$

Q. Let $X \sim B(m,p)$ and $Y \sim B(n,p)$. then find the distribution for $X+Y$. (X and Y are two independent random variables).

Ans. $X+Y \sim B(m+n, p)$.

Q. Let X and Y are independent Poisson variates, then find the conditional probability distribution of X given X+Y. What are the mean and variance of this conditional distribution.?

Soln Let $X \sim P(m)$ and $Y \sim P(n)$

Then $P(X=x) = \dfrac{e^{-m} m^x}{x!}, \quad x = 0,1,2,\dots$

$P(Y=y) = \dfrac{e^{-n} n^y}{y!}, \quad y = 0,1,2,\dots$

$M_X(t) = e^{m(e^t-1)}, \quad M_Y(t) = e^{n(e^t-1)}$

$\Rightarrow M_{X+Y}(t) = e^{(m+n)(e^t-1)}$

$\Rightarrow X+Y \sim P(m+n)$

$\therefore$ The r.v $Z = X+Y \sim P(m+n)$

Hence $P(Z=z) = \dfrac{e^{-(m+n)}(m+n)^z}{z!}, \quad z = 0,1,2,\dots$

Now $P(X=x \mid Z=z) = \dfrac{P(X=x, Z=z)}{P(Z=z)}$

$= \dfrac{P(X=x, Y=z-x)}{P(Z=z)}, \quad x = 0,1,2,\dots,z$

$= \dfrac{P(X=x)\,P(Y=z-x)}{P(Z=z)}$

$= \dfrac{\dfrac{e^{-m} m^x}{x!} \cdot \dfrac{e^{-n} n^{z-x}}{(z-x)!}}{\dfrac{e^{-(m+n)}(m+n)^z}{z!}}$

$= \dfrac{m^x n^{z-x}}{x!(z-x)!} \cdot \dfrac{z!}{(m+n)^z}$

$= {}^zC_x \left(\dfrac{m}{m+n}\right)^x \left(\dfrac{n}{m+n}\right)^{z-x}, \quad x = 0,1,\dots,z$

ABanerjee

$\therefore \; P(X=x \mid Z=3) = {}^3C_x \, p^x q^{3-x} \; ; \; x = 0, 1, \ldots, 3$

where ~~p~~ $p = \frac{m}{m+n}$ & $q = 1 - \frac{m}{m+n} = \frac{n}{m+n}$

Hence $P(X=x \mid Z=3) \sim B\left(3, \frac{m}{m+n}\right)$

Mean $= 3 \frac{m}{m+n}$

Variance $= 3 \cdot \frac{m}{m+n} \cdot \frac{n}{m+n}$

$= \frac{3mn}{(m+n)^2}$.

Q. Show that the correlation coefficient of X and Y is zero if X and Y are two independent random variables. Is the converse is also true. Justify your answer.

Soln. $Cov(X,Y) = E[(X-\bar{x})(Y-\bar{y})] = E[XY - X\bar{x} - Y\bar{y} + \bar{x}\bar{y}]$

$= E(XY) - \bar{x}E(X) - \bar{y}E(Y) + \bar{x}\bar{y}$

$= E(XY) - \bar{x}\bar{y} = E(X)E(Y) - \bar{x}\bar{y}$

$= \cancel{\bar{x}\bar{y}} - \cancel{\bar{x}\bar{y}} = 0.$

Hence $\rho_{xy} = \frac{cov(X,Y)}{\sigma_x \sigma_y} = \frac{0}{\sigma_x \sigma_y} = 0.$

The converse is not necessarily always true. For justification find an example.

ABanerjee