

Bivariate Expectation and Covariance

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

Then

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

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

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

$$\begin{aligned} \therefore E(g(X)) &= \sum_{i=1}^n g(x_i) p_{i.} \\ E(h(Y)) &= \sum_{j=1}^m h(y_j) p_{.j} \end{aligned} \quad \left| \quad \begin{aligned} E(g(X, Y)) \\ &= \sum_{j=1}^m \sum_{i=1}^n g(x_i, y_j) p_{ij} \end{aligned} \right.$$

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

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

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

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

Hence, $E(aX + bY) = aE(X) + bE(Y)$

Covariance of two random variable X and Y is given by

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \end{aligned}$$

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

If X and Y are two independent random variables
Then $E(XY) = E(X)E(Y)$
Hence $\text{Cov}(X, Y) = 0$.

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

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

$$= \sum_{i=1}^n \sum_{j=1}^m x_i y_j p_i \cdot p_j$$

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

$$= E(X)E(Y)$$

as X and Y
are two ind.
r.v.s. $p_{ij} = p_i \cdot p_j$

Proved

Correlation Coefficient

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

$$\rho_{xy} = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y}$$

where σ_x and σ_y are standard deviation of X and Y , respectively.

$$\text{i.e., } \sigma_x^2 = \text{Var}(X), \quad \sigma_y^2 = \text{Var}(Y).$$

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 $\rho_{uv} = \pm \rho_{xy}$.

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

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

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

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

$$\rho_{xy} = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y}$$

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

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

$$\text{Cov}(U, V) = E[(U - \bar{u})(V - \bar{v})]$$

$$= E[(aX + b - a\bar{x} - b)(cY + d - c\bar{y} - d)]$$

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

$$= ac \text{Cov}(X, Y).$$

$$\begin{aligned}\sigma_u^2 &= E(U - \bar{u})^2 = E(aX + b - a\bar{x} - b)^2 \\ &= a^2 E(X - \bar{x})^2 = a^2 \sigma_x^2.\end{aligned}$$

$$\Rightarrow \sigma_u = |a| \sigma_x$$

$$\text{Similarly, } \sigma_v = |c| \sigma_y$$

$$\therefore \rho_{uv} = \frac{\text{cov}(u, v)}{\sigma_u \sigma_v} = \frac{ca \text{cov}(x, y)}{|a| \sigma_x |c| \sigma_y}$$

$$\boxed{\rho_{uv} = \frac{ac}{|a||c|} \rho_{xy}}$$

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 $a < 0, c < 0$ 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., $-1 \leq \rho_{xy} \leq 1$.

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

$$\text{Var}(X) = \sigma_x^2, \text{Var}(Y) = \sigma_y^2$$

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

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

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

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

$$\Rightarrow \sigma_x^2 + k^2\sigma_y^2 + 2k\text{cov}(X, Y) \geq 0$$

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

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

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

$$\Rightarrow A + B \geq 0$$

~~A~~ is always non-neg term
then $\min A = 0$.

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

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

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

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

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

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

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

||
0

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

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

Assume
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 $1 - \rho_{xy}^2 \neq 0$

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

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

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

Let $Z = aX+bY$

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

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

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

$$- a^2 \bar{x}^2 - b^2 \bar{y}^2 - 2ab \bar{x}\bar{y}$$

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

$$= a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2ab \text{Cov}(X,Y)$$

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

$$\boxed{\text{Var}(aX+bY) = a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2ab \rho_{xy} \sigma_x \sigma_y}$$

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

$$\Rightarrow \text{Var}(aX+bY) = a^2 \sigma_x^2 + b^2 \sigma_y^2$$

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 $V = X+Y$.

$$\text{Ans. } \rho_{UV} = \frac{\sigma_x}{\sqrt{\sigma_x^2 + \sigma_y^2}}$$

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 $Y \sim P(n)$

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

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

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

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

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

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

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

$$\text{Now } P(X=x | Z=z) = \frac{P(X=x, Z=z)}{P(Z=z)}$$

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

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

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

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

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

Answer is

Lecture 11: P(7)

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

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

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

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

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

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

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

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

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

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

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

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

Abanajis

Lecture 11: P(8)