Unit-2

Covariance

Covariance:

Definition:

$\boxed{\text{Covariance}}$

For two jointly distributed real-valued random variables X and Y with finite second moments i.e., $E(X^2) < \infty$, the covariance is defined as

$$Cov(X,Y) = E[(X - \overline{X})(Y - \overline{Y})] \quad \text{—} (1)$$

The covariance is also sometimes denoted by σ_{XY} or $\sigma(X,Y)$.

2nd form:

$$\begin{aligned}
Cov(X,Y) &= E[(X - \bar{X})(Y - \bar{Y})] \\
&= E[XY - X\bar{Y} - \bar{X}Y + \bar{X}\bar{Y}] \\
&= E(XY) - E(X\bar{Y}) - E(\bar{X}Y) + E(\bar{X}\bar{Y}) \\
&= E(XY) - \bar{Y}E(X) - \bar{X}E(Y) + \bar{X}\bar{Y}E(1) \\
&= E(XY) - \bar{Y}\bar{X} - \cancel{\bar{X}\bar{Y}} + \cancel{\bar{X}\bar{Y}} \\
&\boxed{= E(XY) - E(X)E(Y)} \quad \text{—} (2)
\end{aligned}$$

Properties:

Property 1: Covariance with itself.

For a random variable X;

$$Cov(X,X) = Var(X)$$

$Y = X \longrightarrow$ Var

$\neq X \longrightarrow$ Cova

Proof: By definition,

$$Cov(X,X) = E(XX) - E(X)E(X)$$

$$= E(X^2) - [E(X)]^2$$

$$= Var(X)$$

i.e., the variance is a special case of the covariance in which the two variables are identical

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

Covariance Properties:

Property 2:

For real-valued random variable X and a is real-valued constant then

$$Cov(X, a) = 0$$

Proof: By definition,

$$\begin{aligned} Cov(X, a) &= E(Xa) - E(X)E(a) \\ &= aE(X) - aE(X) \\ &= 0 \end{aligned}$$

Covariance Properties:

Property 3: Covariance is symmetric

If X, Y are real-valued random variables then

$$Cov(X,Y) = Cov(Y,X)$$

Proof: By definition,

$$\begin{aligned} &Cov(X,Y) \\ &= E(XY) - E(X)E(Y) \\ &= E(YX) - E(Y)EX(X) \\ &= Cov(Y,X) \end{aligned}$$

Property 4: If X, Y, Z are real-valued random variables then

$$Cov(X+Y,Z) = Cov(X,Z) + Cov(Y,Z)$$

Proof: By definition

$$\begin{aligned} &Cov(X+Y,Z) \\ &= E[(X+Y)Z] - E(X+Y)E(Z) \\ &= E(XZ+YZ) - \\ &\qquad [E(X)+E(Y)]E(Z) \\ &= E(XZ) - E(X)E(Z) \\ &\qquad + E(YZ) - E(Y)E(Z) \\ &= Cov(X,Z) + Cov(Y,Z) \end{aligned}$$

Covariance Properties:

Property 5: If X, Y are real-valued random variables and a, b are real-valued constants then

$$Cov(aX, bY) = abCov(X, Y) \quad : \quad Cov(X + a, Y + b) = Cov(X, Y)$$

Proof: By definition,

$$\begin{aligned} &Cov(aX, bY) \\ &= E[(aX)(bY)] - E(aX)E(bY) \\ &= E(abXY) - aE(X)bE(Y) \\ &= ab[E(XY) - E(X)E(Y)] \\ &= abCov(X, Y) \end{aligned}$$

$$\begin{aligned} &E[(X+a)(Y+b)] - E(X+a)E(Y+b) \\ &= E[XY + bX + aY + ab] - [E(X)+a][E(Y)+b] \\ &= E(XY) \cancel{+bE(X)} \cancel{+aE(Y)} \cancel{+ab} \\ &\quad - E(X)E(Y) \cancel{-bE(X)} \cancel{-aE(Y)} \cancel{-ab} \\ &= Cov(X, Y) \end{aligned}$$

Covariance Properties:

Property 6: For real-valued random variables X, Y and real-valued constants a, b, c, d we have

$$Cov(aX + bY, cX + dY) = acVar(X) + bdVar(Y) + (ad + bc)Cov(X, Y)$$

Proof:

$$\begin{aligned}
&Cov(aX + bY, cX + dY) \\
&\quad = E[(aX + bY)(cX + dY)] - E(aX + bY)E(cX + dY) \\
&\quad = E[acX^2 + bdY^2 + (ad + bc)XY] - [aE(X) + bE(Y)][cE(X) + dE(Y)] \\
&\quad = ac\left[E(X^2) - (E(X))^2\right] + bd\left[E(Y^2) - (E(Y))^2\right] \\
&\qquad\qquad + (ad + bc)[E(XY) - E(X)E(Y)] \\
&\quad = acVar(X) + bdVar(Y) + (ad + bc)Cov(X, Y)
\end{aligned}$$

Covariance Properties:

Property 8: Covariance Relationship to inner products

Many properties of the covariance are similar to the inner product

1) Symmetric:

$$Cov(X,Y) = Cov(Y,X)$$

2) Positive semi-definite:

$$Cov(X,X) = Var(X) \geq 0 \text{ and}$$

$$Cov(X,X) = 0 \text{ implies that } X \text{ is constant almost surely.}$$

3) Bilinear:

For constants a, b and random variables X, Y, Z.

$$Cov(aX + bY, Z) = aCov(X,Z) + bCov(Y,Z)$$

Covariance Questions:

Discrete variables

Let X and Y be two random variables

X \ Y	-1	0	1
-1	0	0.1	0.1
0	0.2	0.2	0.2
1	0	0.1	0.1

Prove that they are uncorrelated.

Continuous variables

$$f(x,y) = \begin{cases} 8xy & ; 0 \le y \le x \le 1 \\ 0 & ; \ elsewhere \end{cases}$$

Find the covariance of X and Y.

Covariance Questions:

Two Steps Approach:

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

Step 1:

Find Marginal density of X and Y

For discrete variables	For continuous variables
$p(x) = \sum_y p(x,y)$	$f(x) = \int f(x,y)dy$
$p(y) = \sum_x p(x,y)$	$f(y) = \int f(x,y)dx$

Step 2:

Find expectation of X, Y and XY

For discrete variables	For continuous variables
$E(X) = \sum_y x\, p(x,y)$	$E(X) = \int x\, f(x)dx$
$E(Y) = \sum_x y\, p(x,y)$	$E(Y) = \int y f(y)dy$
$E(XY) = \sum\sum xy\, p(x,y)$	$E(XY) = \int\int xyf(x,y)dydx$

Example: Let X and Y be two random variables each taking three values $-1, 0, 1$ and having the joint probability distribution is

$X \backslash Y$	-1	0	1
-1	0	0.1	0.1
0	0.2	0.2	0.2
1	0	0.1	0.1

Prove that X and Y have different expectations and they are uncorrelated.

$Cov(X,Y) = 0$

Solution:

(i) $E(X) \neq E(Y)$

(ii) $Cov(X,Y) = 0$

Solution:

Marginal of X is

x	-1	0	1
$p(x)$			

$p(x) = \sum_y p(x,y)$

Marginal of Y is

y	-1	0	1
$p(y)$			

X \ Y	-1	0	1
-1	0	0.1	0.1
0	0.2	0.2	0.2
1	0	0.1	0.1

Marginal of X is

x	-1	0	1
$p(x)$	0.2	0.6	0.2

Marginal of Y is

y	-1	0	1
$p(y)$	0.2	0.4	0.4

$$E(X) = \sum x p(x)$$
$$= (-1)(0.2) + (0)(0.6) + (1)(0.2)$$
$$= 0$$

$$E(Y) = \sum y p(y)$$
$$= (-1)(0.2) + (0)(0.4) + (1)(0.4)$$
$$= 0.2$$

Target is to prove $Cov(X,Y) = 0$

i.e., $E(XY) - E(X)E(Y) = 0$

$$E(XY) = \sum\sum xyp(x,y)$$

$$= (-1)(-1)(0)$$

X \ Y	-1	0	1
-1	0	0.1	0.1
0	0.2	0.2	0.2
1	0	0.1	0.1

$$\begin{aligned} E(XY) &= \sum\sum xyp(x,y) \\ &= (-1)(-1)(0) + (-1)(0)(0.1) + (-1)(1)(0.1) \\ &\qquad + \cdots \ldots + (1)(0)(0.1) + (1)(1)(0.1) \\ &= 0 \end{aligned}$$

Thus,

$$\begin{aligned} Cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= 0 - 0(0.2) \\ &= 0 \end{aligned}$$

Hence, X and Y are uncorrelated.

Example...

Example: The joint probability mass function of X and Y is given below:

Find the covariance of (X, Y).

X \ Y	−1	1
0	1/8	3/8
1	2/8	2/8

Solution:

The marginal of X is

x	0	1
$p(x)$	½	½

$$E(X) = \sum x p(x)$$
$$= (0)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{2}\right)$$
$$= 1/2$$

The marginal of Y is

y	-1	1
$p(y)$	3/8	5/8

$$E(Y) = \sum y p(y)$$
$$= (-1)\left(\frac{3}{8}\right) + (1)\left(\frac{5}{8}\right)$$
$$= 1/4$$

Solution....

$$E(XY) = \sum\sum xyp(x,y)$$
$$= (0)(-1)\left(\frac{1}{8}\right) + (0)(1)\left(\frac{3}{8}\right)$$
$$+(1)(-1)\left(\frac{2}{8}\right) + (1)(1)\left(\frac{2}{8}\right)$$
$$= 0$$

Hence,

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$
$$= 0 - \left(\frac{1}{2}\right)\left(\frac{1}{4}\right)$$
$$= -\frac{1}{8}$$

Example: Let X and Y be two random variables each taking three values $-1, 0, 1$ and having the joint probability distribution

Prove that X and Y are uncorrelated.

X \ Y	-1	0	1
-1	2/16	1/16	2/16
0	2/16	2/16	2/16
1	2/16	1/16	2/16

Solution: The marginal of X is

x	-1	0	1
$p(x)$	5/16	6/16	5/16

$$E(X) = \sum xp(x) = -\frac{5}{16} + 0 + \frac{5}{16} = 0$$

The marginal of Y is

y	-1	0	1
$p(y)$	6/16	4/16	6/16

$$E(Y) = \sum yp(y) = -\frac{6}{16} + 0 + \frac{6}{16} = 0$$

Solutions

$$
\begin{aligned}
E(XY) &= \Sigma\Sigma xyp(x,y) \\
&= (-1)(1)\left(\frac{2}{16}\right) + (-1)(0)\left(\frac{1}{16}\right) \\
&\quad + \cdots \ldots \ldots + (1)(1)\left(\frac{2}{16}\right) \\
&= \frac{2}{16} - \frac{2}{16} - \frac{2}{16} + \frac{2}{16} \\
&= 0
\end{aligned}
$$

$$\therefore \; Cov(X,Y) = E(XY) - E(X)E(Y) = \underline{\underline{0}}$$

Thus, X and Y are uncorrelated.

Example: If X and Y are two independent random variables with means 5 and 10 and standard deviations 2 and 3 respectively. Find the covariance between $3X+4Y$ and $3X-Y$.

Target

$$Cov(U,V) = E(UV) - E(U)E(V)$$

$$E(XY) = E(X)E(Y) = 5\times 10 = 50$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$\Rightarrow E(X^2) = 4 + 25 = 29$$

Solution: Given that $E(X) = 5$; $E(Y) = 10$; $Var(X) = 4$; $Var(Y) = 9$

Let $U = 3X + 4Y$ and $V = 3X - Y$

$$E(U) = 3E(X) + 4E(Y) = 15 + 40 = 55$$

$$E(V) = 3E(X) - E(Y) = 15 - 10 = 5$$

$$E(UV) = E(9X^2 + 9XY - 4Y^2)$$

$$= 9E(X^2) + 9E(XY) - 4E(Y^2)$$

$$= 9[4+25] + 9(5)(10) - 4[1\cancel{6}\,9 + 100]$$

$$= 247$$

$\therefore Cov(U,V)$

$$= E(UV) - E(U)E(V)$$

$$= 247 - (55)(5)$$

$$= -28$$

Example: If X, Y and Z are uncorrelated random variables with zero means and Standard deviations 5, 12 and 9 respectively; $U = X + Y$ and $V = Y + Z$. Find the covariance between U and V.

Solution: Given that X, Y and Z are uncorrelated random variables

$$Cov(X,Y) = 0;\ Cov(X,Z) = 0;\ Cov(Y,Z) = 0$$
$$E(X) = E(Y) = E(Z) = 0;$$
$$Var(X) = 25;\ Var(Y) = 144;\ Var(Z) = 81$$

Target

$$Cov(U,V) = E(UV) - E(U)E(V)$$

$$Cov(X,Y) = 0$$
$$\Rightarrow E(XY) - E(X)E(Y) = 0$$
$$\Rightarrow E(XY) - 0 \times 0 = 0$$

$$E(U) = E(X+Y) = E(X) + E(Y) = 0$$

$$E(V) = E(Y+Z) = E(Y) + E(Z) = 0$$

$$E(UV) = E[XY + XZ + Y^2 + YZ] = E(XY) + E(XZ) + E(Y^2) + E(YZ) = 0 + 0 + E(Y^2) + 0$$

because Cov(X,Y) = 0 implies E(XY) = E(X)E(Y) $= 0 \times 0 = 0$

$$Var(Y) = E(Y^2) - [E(Y)]^2$$
$$\Rightarrow 144 = E(Y^2) - 0$$

$$Cov(U,V) = E(UV) - E(U)E(V) = 144 - 0 = 144$$

Example: Two independent random variables X and Y have p.d.f's defined by

$$f(x) = \begin{cases} 4ax & ; \ 0 \le x \le 1 \\ 0 & ; elsewhere \end{cases} \quad \text{and} \quad f(y) = \begin{cases} 4by & ; \ 0 \le y \le 1 \\ 0 & ; elsewhere \end{cases}.$$

Show that $X + Y$ and $X - Y$ are uncorrelated.

Solution: Take $U = X + Y \ ; \ V = X - Y$

Target

To show U and V are uncorrelated,

it means $Cov(U,V) = 0$

For p.d.f. X, we have

$$\int_0^1 f(x)dx = 1$$

$$\Rightarrow 4a\int_0^1 xdx = 1$$

$$\Rightarrow 2a = 1$$

$$\Rightarrow a = \frac{1}{2}$$

For p.d.f. Y, we have

$$\int_0^1 f(y)dy = 1$$

$$\Rightarrow 4b\int_0^1 ydy = 1$$

$$\Rightarrow 2b = 1$$

$$\Rightarrow b = \frac{1}{2}$$

$$Cov(U,V) = \underline{\underline{E(UV)}} - E(U)E(V)$$

$$f(x) = 2x \ ; \ 0 \le x \le 1$$

$$f(y) = 2y \ ; \ 0 \le y \le 1$$

$E(U)$

$= E(X+Y)$

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

$= \int_0^1 xf(x)dx + \int_0^1 yf(y)dy$

$= 2\int_0^1 x^2dx + 2\int_0^1 y^2dy$

$= \frac{2}{3} + \frac{2}{3}$

$= \frac{4}{3}$

$E(V)$

$= E(X-Y)$

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

$= \int_0^1 xf(x)dx - \int_0^1 yf(y)dy$

$= 2\int_0^1 x^2dx - 2\int_0^1 y^2dy$

$= \frac{2}{3} - \frac{2}{3}$

$= 0$

$E(UV)$

$= E(X^2 - Y^2)$

$= E(X^2) - E(Y^2)$

$= \int_0^1 x^2f(x)dx - \int_0^1 y^2f(y)dy$

$= 2\int_0^1 x^3dx - 2\int_0^1 y^3dy$

$= \frac{2}{4} - \frac{2}{4}$

$= 0$

$Cov(U,V) = E(UV) - E(U)E(V)$

$= 0 - \frac{4}{3}(0) = 0$

Hence, U and V are uncorrelated.