Sri SAI RAM ENGINEERING COLLEGE
INSTITUTE OF TECHNOLOGY
West Tambaram, Chennai - 44

Sairam INSTITUTIONS

SAIRAM DIGITAL RESOURCES

YEAR II SEM IV

MA8391

PROBABILITY AND STATISTICS (IT)

UNIT II

TWO – DIMENSIONAL RANDOM VARIABLES

2.3 COVARIANCE, CORRELATION

SCIENCE & HUMANITIES

COVARIANCE

A common measure of the relationship between two random variables is the covariance.

If X and Y are random variables, then co-variance between them is defined as,

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

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

Note : If X and Y are independent, then $Cov\,(X,Y) = 0$

If X and Y are independent, then $E[XY] = E(X).E(Y) \Rightarrow Cov\,(X,Y) = 0$

Properties of Covariance

(i) $Cov\,(aX,bY) = ab\,Cov(X,Y)$

(ii) $Cov\,(X+a,Y+b) = Cov(X,Y)$

$(iii)\ \ Cov\,(aX+b, cY+d)\ = ac\ Cov(X,Y)$

$(iv)\ \ V(X_1+X_2)\ = V(X_1)+V(X_2)+2Cov\,(X_1,X_2)$

$(v)\ \ V(X_1-X_2)\ = V(X_1)+V(X_2)-2Cov\,(X_1,X_2)$

1. Two random variables X and Y have the following joint probability density function

$$\boldsymbol{f(x,y) = \begin{cases} 2-x-y, 0\le x\le 1, 0\le y\le 1 \\ 0\ , otherwise \end{cases}}$$ **Show that $Cov\,(X,Y)\ = -\frac{1}{144}$.**

Solution :

Given the joint probability density function $f(x,y) = \begin{cases} 2-x-y, 0\le x\le 1, 0\le y\le 1 \\ 0\ , otherwise \end{cases}$

Marginal density function of X is $f_X(x) = \int_{\infty}^{\infty} f\,(x,y)dy$

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

$$= \left[2y - \text{xy} - \frac{y^2}{2}\right]_0^1$$

$$= 2 - x - \frac{1}{2}$$

$$f_X(x) = \begin{cases} \frac{3}{2} - x, 0 < x \le 1 \\ 0\ otherwise \end{cases}$$

Marginal density function of Y is $f_Y(y) = \int_0^1 (2 - x - y)\, dx$

$$= \left[2x - \frac{x^2}{2} - xy\right]_0^1$$

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

$$f_Y(y) = \begin{cases} \frac{3}{2} - y, & 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$

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

$$E(X) = \int_0^1 x\, f_X(x) dx = \int_0^1 x\left(\frac{3}{2} - x\right) dx = \left[\frac{3}{2}\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \left[\frac{3}{4} - \frac{1}{3}\right] = \frac{5}{12}$$

$$E(Y) = \int_0^1 y\, f_Y(y) dy = \int_0^1 y\left(\frac{3}{2} - y\right) dy = \left[\frac{3}{2}\frac{y^2}{2} - \frac{y^3}{3}\right]_0^1 = \left[\frac{3}{4} - \frac{1}{3}\right] = \frac{5}{12}$$

$$E(XY) = \int_0^1 \int_0^1 xy\, f(x,y) dxdy$$

$$= \int_0^1 \int_0^1 x\, y(2 - x - y) dxdy$$

$$= \int_0^1 \int_0^1 (2xy - x^2y - xy^2)\, dxdy$$

$$= \int_0^1 \left[\frac{2x^2y}{2} - \frac{x^3}{3}y - \frac{x^2}{2}y^2\right]_0^1 dy$$

$$= \int_0^1 \left(y - \frac{1}{3} - \frac{y^2}{2}\right) dy$$

$$= \left[\frac{y^2}{2} - \frac{y}{3} - \frac{y^3}{6}\right]_0^1 = \frac{1}{6}$$

$$Cov(X,Y) = \left(\frac{1}{6}\right) - \left(\frac{5}{12}\right)\left(\frac{5}{12}\right) = \left(\frac{1}{6}\right) - \left(\frac{25}{144}\right) = -\frac{1}{144} \qquad \therefore\ Cov(X,Y) = -\frac{1}{144}$$

CORRELATION

Let X and Y be two random variables. The coefficient of correlation between X and Y is denoted by ρ_{XY} *or* r_{XY} and is defined by $\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY)-E(X)E(Y)}{\sigma_X \sigma_Y}$, if $\sigma_X \neq 0, \sigma_Y \neq 0$.

Note :

(i) This is called Karl Pearson's Correlation Coefficient.

(ii) If X and Y are independent, then $Cov\ (X,Y) = 0 \therefore\ \rho_{XY} = 0$.

(iii) If $\rho_{XY} = 0$, we say X and Y are uncorrelated.

(iv) Correlation coefficient always lies between -1 and 1. $i.e. -1 \leq \rho_{XY} \leq 1$.

The Karl Pearson's coefficient of correlation

Let X and Y be given random variables. The Karl Pearson's coefficient of correlation is denoted by r_{XY} or $r(X,Y)$ and defined as $r(X,Y) = r_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$

where $\quad cov(X,Y) = \frac{1}{n}\sum XY - \bar{X}\bar{Y}$

$$\sigma_X = \sqrt{\frac{1}{n}\sum X^2 - \bar{X}^2}, \quad \sigma_Y = \sqrt{\frac{1}{n}\sum Y^2 - \bar{Y}^2} \quad \text{and} \quad \bar{X} = \frac{\sum X}{n}, \ \bar{Y} = \frac{\sum Y}{n}$$

1. **Calculate the correlation coefficient for the following heights (in inches) of fathers *X* and their Sons Y.**

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

Solution:

X	Y	XY	X^2	Y^2
65	67	4355	4225	4489
66	68	4488	4356	4624
67	65	4355	4489	4225
67	68	4556	4489	4624
68	72	4896	4624	5184
69	72	4968	4761	5184
70	69	4830	4900	4761
72	71	5112	5184	5041
∑ X=544	∑ Y=552	∑ XY=37560	∑ X^2 = 37028	∑ Y^2 = 38132

$$\overline{X} = \frac{\Sigma X}{n} = \frac{544}{8} = 68\,; \quad \overline{Y} = \frac{\Sigma Y}{n} = \frac{552}{8} = 69$$

$$\overline{X}\,\overline{Y} = 68 \times 69 = 4692$$

$$\sigma_X = \sqrt{\frac{1}{n}\Sigma x^2 - X^2 -} = \sqrt{\frac{1}{8}(37028) - 68^2} = \sqrt{46285 - 4624} = 2.121$$

$$\sigma_{\mathrm{y}} = \sqrt{\frac{1}{n}\Sigma y^2 - y^2} = \sqrt{\frac{1}{8}(38132) - 69^2} = \sqrt{47665 - 4761} = 2.345$$

$$Cov(X,Y) = \frac{1}{n}\Sigma XY - \overline{X}\,\overline{Y} = \frac{1}{8}(37650) - 68 \times 69 = 4695 - 4692 = 3$$

The correlation coefficient of X and Y is given by

$$r(X,Y) = \frac{Cov(X,\mathrm{Y})}{\sigma_X \sigma_{\mathrm{y}}} = \frac{3}{(2.121)\,(2.345)} = \frac{3}{4.973} = 0.6032.$$

2. The joint pdf of X and y is

Y \ X	-1	1
0	1/8	3/8
1	2/8	2/8

Find the correlation coefficient of (X, Y).

Solution:

Marginal probability mass function of X is

When $X = 0,\ P(X) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$ $\quad X = 1,\ P(X) = \frac{2}{8} + \frac{2}{8} = \frac{4}{8}$

Marginal probability mass function of Y is

When $Y = -1,\ P(Y) = \frac{1}{8} + \frac{2}{8} = \frac{3}{8}$ $\quad Y = 1,\ P(Y) = \frac{3}{8} + \frac{2}{8} = \frac{5}{8}$

Y \ X	-1	1	P(x)
0	1/8	3/8	4/8
1	2/8	2/8	4/8
P(y)	3/8	5/8	1

$$E(X) = \sum x_i p(x_i) = (-1)\left(\frac{3}{8}\right) + (1)\left(\frac{5}{8}\right) = \frac{2}{8} = \frac{1}{4}$$

$$E(X^2) = \sum {x_i}^2 p(x_i) = (1)\left(\frac{3}{8}\right) + (1)\left(\frac{5}{8}\right) = 1$$

$$E(Y) = \sum y_j p(y_j) = (0)\left(\frac{4}{8}\right) + (1)\left(\frac{4}{8}\right) = \frac{4}{8} = \frac{1}{2}$$

$$E(Y^2) = \sum {y_j}^2 p(y_j) = (0)\left(\frac{4}{8}\right) + (1)\left(\frac{4}{8}\right) = \frac{4}{8} = \frac{1}{2}$$

$$Var(X) = E(X^2) - \left(E(X)\right)^2 = 1 - \left(\frac{1}{4}\right)^2 = 1 - \frac{1}{16} = \frac{15}{16}$$

$$Var(Y) = E(Y^2) - \left(E(Y)\right)^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

$$E(XY) = \sum_i \sum_j x_i y_j \, p(x_i, y_j)$$

$$= (0)(-1)\left(\frac{1}{8}\right) + (0)(1)\left(\frac{3}{8}\right) + (1)(-1)\left(\frac{2}{8}\right) + (1)(1)\left(\frac{2}{8}\right)$$

$$= 0 + 0 - \left(\frac{2}{8}\right) + \left(\frac{2}{8}\right) = 0$$

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

$$r = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{-\left(\frac{1}{8}\right)}{\sqrt{\frac{1}{4}}\sqrt{\frac{15}{16}}} = \frac{-\left(\frac{1}{8}\right)}{\frac{\sqrt{15}}{8}} - 0.258.$$

3. Two random variables X and Y have the following joint probability density function

$$f(x,y)=\begin{cases}\frac{6-x-y}{8}, & 0\le x\le 2, 2\le y\le 4\\ 0, & otherwise\end{cases}$$

Find the correlation coefficient.

Solution :

Correlation coefficient $\rho_{XY}=\frac{E(XY)-E(X)E(y)}{\sigma_X\sigma_y}$

Marginal density function of X is

$$f_X(x)=\int_{-\infty}^{\infty} f\,(x,y)dy=\int_2^4\left(\frac{6-x-y}{8}\right)dy=\frac{6-2x}{8}$$

Marginal density function of Y is

$$f_Y(y)=\int_{-\infty}^{\infty} f\,(x,y)dx=\int_0^2\left(\frac{6-x-y}{8}\right)dx=\frac{10-2y}{8}$$

Then $E(X) = \int_0^2 x\, f_X(x)dx = \int_0^2 x\left(\frac{6-2x}{8}\right)dx = \frac{1}{8}\left[\frac{6x^2}{2} - \frac{2x^3}{3}\right]_0^2$

$$= \frac{1}{8}\left[3(2^2 - 0^2) - \frac{2}{3}(2^3 - 0^3)\right] = \frac{1}{8}\left[12 - \frac{16}{3}\right] = \frac{1}{8}\left(\frac{20}{3}\right) = \frac{5}{6}$$

$$E(Y) = \int_2^4 y\left(\frac{10-2y}{8}\right)dy = \frac{1}{8}\left[\frac{10y^2}{2} - \frac{2y^3}{3}\right]_2^4 = \frac{1}{8}\left[5(4^2 - 2^2) - \frac{2}{3}(4^3 - 2^3)\right]$$

$$= \frac{1}{8}\left[5(12) - \frac{2}{3}(56)\right] = \frac{1}{8}\left[\frac{68}{3}\right] = \frac{17}{6}$$

$$E(X^2) = \int_0^2 x^2\, f_X(x)dx = \int_0^2 x^2\left(\frac{6-2x}{8}\right)dx = \frac{1}{8}\left[\frac{6x^3}{3} - \frac{2x^4}{4}\right]_0^2$$

$$= \frac{1}{8}\left[2(2^3 - 0^3) - \frac{2}{4}(2^4 - 0^4)\right] = \frac{1}{8}\,[16 - 8] = 1$$

$$E(Y^2) = \int_2^4 y^2\left(\frac{10-2y}{8}\right)dy = \frac{1}{8}\left[\frac{10y^3}{3} - \frac{2y^4}{4}\right]_2^4 = \frac{1}{8}\left[\frac{10}{3}(4^3 - 2^3) - \frac{1}{2}(4^4 - 2^4)\right]$$

$$= \frac{1}{8}\left[\frac{10}{3}(64 - 8) - \frac{1}{2}(256 - 16)\right] = \frac{1}{8}\left[\frac{560}{3} - 120\right] = \frac{25}{3}$$

$$Var(X) = \sigma_X^2 = E(X^2) - [E(X)]^2 = 1 - \left(\frac{5}{6}\right)^2 = 1 - \frac{25}{36} = \frac{11}{36}$$

$$Var(Y) = \sigma_Y^2 = E(Y^2) - [E(Y)]^2 = \frac{25}{3} - \left(\frac{17}{6}\right)^2 = \frac{25}{3} - \frac{289}{36} = \frac{11}{36}$$

$$E(XY) = \int_2^4 \int_0^2 xy\left(\frac{6-x-y}{8}\right) dxdy$$

$$= \int_2^4 \int_0^2 \left(\frac{6xy}{8} - \frac{x^2y}{8} - \frac{xy^2}{8}\right) dxdy$$

$$= \frac{1}{8}\int_2^4 \left[6y\left(\frac{x^2}{2}\right) - y\left(\frac{x^3}{3}\right) - y^2\left(\frac{x^2}{2}\right)\right]_0^2 dy$$

$$= \frac{1}{8}\int_2^4 \left[6y\left(\frac{2^2}{2} - 0\right) - y\left(\frac{2^3}{3} - 0\right) - y^2\left(\frac{2^2}{2} - 0\right)\right] dy$$

$$= \frac{1}{8}\int_2^4 (12y - \frac{8}{3}y - 2y^2)\, dy = \frac{1}{8}\left[12\left(\frac{y^2}{2}\right) - \frac{8}{3}\left(\frac{y^2}{2}\right) - 2\left(\frac{y^3}{3}\right)\right]_2^4$$

$$= \frac{1}{8}\left[6(4^2 - 2^2) - \frac{4}{3}(4^2 - 2^2) - \frac{2}{3}(4^3 - 2^3)\right]$$

$$= \frac{1}{8}\left[6(12) - \frac{4}{3}(12) - \frac{2}{3}(56)\right] = \frac{1}{8}\left[56 - \frac{112}{3}\right] = \frac{7}{3} \qquad \therefore E(XY) = \frac{7}{3}$$

$$\rho_{XY} = \frac{E(XY)-E(X)E(Y)}{\sigma_X\sigma_y} = \frac{\frac{7}{3}-\left(\frac{5}{6}\right)\left(\frac{17}{6}\right)}{\left(\frac{\sqrt{11}}{6}\right)\left(\frac{\sqrt{11}}{6}\right)} = -\frac{1}{11}$$

4. Let X_1 and X_2 be two independent random variables with mean 5 and 10 and standard deviations 2 and 3 respectively. Obtain the correlation coefficient of UV where $U = 3X_1 + 4X_2$ and $V = 3X_1 - X_2$.

Solution:

Given $E(X_1) = 5, E(X_2) = 10$ $\quad \sigma_{X_1} = 2,\ \sigma_{X_2} = 3$ $\quad \therefore\ V(X_1) = 4, V(X_2) = 9$

Since X and Y are independent $E(XY) = E(X)E(Y)$

Correlation coefficient $= \frac{E(UV)-E(U)E(V)}{\sqrt{Var(U)}\sqrt{Var(V)}}$

$E(U) = E(3X_1 + 4X_2) = 3E(X_1) + 4E(X_2) = (3)(5) + (4)(10) = 15 + 40 = 55.$

$E(V) = E(3X_1 - X_2) = 3E(X_1) - E(X_2) = (3)(5) - 10 = 15 - 10 = 5$

$$E(UV) = E[(3X_1 + 4X_2)(3X_1 - X_2)]$$
$$= E[9X_1^2 - 3X_1X_2 + 12X_1X_2 - 4X_2^2]$$
$$= 9E(X_1^2) - 3E(X_1X_2) + 12E(X_1X_2) - 4E(X_2^2)$$
$$= 9E(X_1^2) + 9E(X_1X_2) - 4E(X_2^2)$$
$$= 9E(X_1^2) + 9E(X_1)E(X_2) - 4E(X_2^2)$$
$$= 9E(X_1^2) + 450 - 4E(X_2^2) \qquad \because E(X_1)E(X_2) = (5)(10) = 50$$

$$V(X_1) = E(X_1^2) - [E(X_1)]^2$$

$$E(X_1^2) = V(X_1) + [E(X_1)]^2 = 4 + 25 = 29$$

$$E(X_2^2) = V(X_2) + [E(X_2)]^2 = 9 + 100 = 109$$

$$\therefore E(UV) = (9 \times 29) + 450 - (4 \times 109) = 261 + 450 - 436 = 275$$

$$Cov(U,V) = E(UV) - E(U)E(V) = 275 - (5 \times 55) = 0$$

Since $Cov(U,V) = 0$, Correlation coefficient $= 0$.

5. Two random variables X and Y are defined by,

$$f(x) = \begin{cases} 4ax\,, & 0 \le x \le 1 \\ 0\,, & otherwise \end{cases}\,, \qquad f(y) = \begin{cases} 4by\,, & 0 \le y \le 1 \\ 0\,, & otherwise \end{cases}$$

Show that $U = X + Y$ and $V = X - Y$ are uncorrelated.

Solution:

$$f(x) = \begin{cases} 4ax\,, & 0 \le x \le 1 \\ 0\,, & otherwise \end{cases}$$

$f(x)$ is the density function of X

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

$$\int_0^1 4ax\,dx = 1$$

$$4a\left[\frac{x^2}{2}\right]_0^1 = 1$$

$$2a = 1\,; \quad a = \frac{1}{2}$$

$$f(y) = \begin{cases} 4by\,, & 0 \le y \le 1 \\ 0\,, & otherwise \end{cases}$$

$f(y)$ is the density function of Y

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

$$\int_0^1 4by\,dy = 1$$

$$4b\left[\frac{y^2}{2}\right]_0^1 = 1$$

$$2b = 1\,; \quad b = \frac{1}{2}$$

To prove $U = X + Y$ and $V = X - Y$ are uncorrelated

$i.e.,$ to prove $Cov(U,V) = 0$

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

$$E(U) = E(X+Y) = E(X) + E(Y) \qquad E(V) = E(X-Y) = E(X) - E(Y)$$

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

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^1 x\,(4ax)\, dx = 4a \int_0^1 x^2 dx = 4a\left[\frac{x^3}{3}\right]_0^1 = \frac{4a}{3} = \frac{2}{3}$$

$$E(Y) = \int_{-\infty}^{\infty} y\, f(y)\, dy = \int_0^1 y\,(4by)\, dy = 4b \int_0^1 y^2 dy = 4b\left[\frac{y^3}{3}\right]_0^1 = \frac{4b}{3} = \frac{2}{3}$$

$$E(XY) = E(X)E(Y) = \left(\frac{2}{3}\right).\left(\frac{2}{3}\right) = \frac{4}{9}$$

$$E(U) = E(X+Y) = E(X) + E(Y) = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$$

$$E(V) = E(X - Y) = E(X) + E(Y) = \frac{2}{3} - \frac{2}{3} = 0$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \, f(x)\, dx = \int_0^1 x^2\,(2x)\,dx = 2\int_0^1 x^3 dx = 2\left[\frac{x^4}{4}\right]_0^1 = \frac{2}{4}[1-0] = \frac{1}{2}$$

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 \, f(y)\, dy = \int_0^1 y^2\,(2y)\,dy = 2\int_0^1 y^3 dy = 2\left[\frac{y^4}{4}\right]_0^1 = \frac{2}{4}[1-0] = \frac{1}{2}$$

$$E(UV) = E(X^2 - Y^2) = E(X^2) - E(Y^2) = \frac{1}{2} - \frac{1}{2} = 0$$

$$Cov(U,V) = E(UV) - E(U)E(V) = 0 - \left(\frac{4}{3}\right)(0) = 0$$

$\therefore$ U and V are uncorrelated.