SMTA1402 - Probability and Statistics

Unit-4 Correlation and Regression

21).a). Two random variables X and Y have the following joint probability density function $f(x,y)=\begin{cases}2-x-y, & 0\le x\le 1,\ 0\le y\le 1\\ 0 & , otherwise\end{cases}$. Find the marginal probability density function of X and Y. Also find the covariance between X and Y.

b). If $f(x,y)=\dfrac{6-x-y}{8}, 0\le x\le 2, 2\le y\le 4$ for a bivariate X,Y, find the correlation coefficient

Solution:

a) 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 $f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$Cov(X,Y) = \frac{1}{6} - \frac{5}{12} \times \frac{5}{12}$$

$$= \frac{1}{6} - \frac{25}{144} = -\frac{1}{144}.$$

b). Correlation coefficient $\rho_{XY} = \dfrac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}$

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 $E(X) = \int_0^2 x f_X(x)\,dx = \int_0^2 x\left(\frac{6-2x}{8}\right) dx$

$$= \frac{1}{8}\left[\frac{6x^2}{2} - \frac{2x^3}{3}\right]_0^2$$

$$= \frac{1}{8}\left[12 - \frac{16}{13}\right] = \frac{1}{8} \times \frac{20}{3} = \frac{5}{6}$$

$$E(Y) = \int_2^4 y\left(\frac{10-2y}{8}\right) dy = \frac{1}{8}\left[\frac{10y^2}{2} - \frac{2y^3}{3}\right]_2^4 \quad = \frac{17}{6}$$

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

$$E(Y^2) = \int_2^4 y^2\left(\frac{10-2y}{8}\right) dy = \frac{1}{8}\left[\frac{10y^3}{3} - \frac{2y^4}{4}\right]_2^4 = \frac{25}{3}$$

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

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

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

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

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

$$= \frac{1}{8}\left[96 - \frac{64}{3} - \frac{128}{3} - 24 + \frac{16}{3} + \frac{16}{3}\right] = \frac{1}{8}\left[\frac{56}{3}\right]$$

$$E(XY) = \frac{7}{3}$$

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

$$\rho_{XY} = -\frac{1}{11}.$$

22.a). Let the random variables X and Y have pdf $f(x, y) = \frac{1}{3}$, $(x, y) = (0,0),(1,1),(2,0)$. Compute the correlation coefficient.

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

Solution:

a). The probability distribution is

Y \ X	0	1	2	$P(Y)$
0	$\frac{1}{3}$	0	0	$\frac{1}{3}$
1	0	$\frac{1}{3}$	0	$\frac{1}{3}$
0	0	0	$\frac{1}{3}$	$\frac{1}{3}$
$P(X)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

$$E(X) = \sum_i x_i p_i(x_i) = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{3}\right) + \left(2 \times \frac{1}{3}\right) = 1$$

$$E(Y) = \sum_j y_i p_j(y_j) = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{3}\right) + \left(0 \times \frac{1}{3}\right) = \frac{1}{3}$$

$$E(X^2) = \sum_i x_i^2 p(x_i) = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{3}\right) + \left(4 \times \frac{1}{3}\right) = \frac{5}{3}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{5}{3} - 1 = \frac{2}{3}$$

$$E(Y^2) = \sum_j y_j^2 p(y_j) = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{3}\right) + \left(0 \times \frac{1}{3}\right) = \frac{1}{3}$$

$$\therefore V(Y) = E(Y^2) - [E(Y)]^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}$$

Correlation coefficient $\rho_{XY} = \dfrac{E(XY)-E(X)E(Y)}{\sqrt{V(X)V(Y)}}$

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

$$=0.0.\frac{1}{3}+0.1.0+1.0.0+1.1.\frac{1}{3}+1.2.0+0.0.0+0.1.0+0.2.\frac{1}{3}=\frac{1}{3}$$

$$\rho_{XY}=\frac{\frac{1}{3}-(1)\left(\frac{1}{3}\right)}{\sqrt{\frac{2}{3}\times\frac{2}{9}}}=0$$

Correlation coefficient $=0$.

b). Given $E(X_1)=5,\quad E(X_2)=10$

$V(X_1)=4,\quad V(X_2)=9$

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

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

$$E(U)=E(3X_1+4X_2)=3E(X_1)+4E(X_2)$$

$$=(3\times5)+(4\times10)=15+40=55.$$

$$E(V)=E(3X_1-X_2)=3E(X_1)-E(X_2)$$

$$=(3\times5)-10=15-10=5$$

$$E(UV)=E\left[(3X_1+4X_2)(3X_1-X_2)\right]$$

$$=E\left[9X_1^2-3X_1X_2+12X_1X_2-4X_2^2\right]$$

$$=9E(X_1^2)-3E(X_1X_2)+12E(X_1X_2)-4E(X_2^2)$$

$$=9E(X_1^2)+9E(X_1X_2)-4E(X_2^2)$$

$$=9E(X_1^2)+9E(X_1)E(X_2)-4E(X_2^2)$$

$$=9E(X_1^2)+450-4E(X_2^2)$$

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

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

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

$$\therefore E(UV)=(9\times29)+450-(4\times109)$$

$$=261+450-436=275$$

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

$$=275-(5\times 55)=0$$

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

23.a). Let the random variable X has the marginal density function $f(x)=1,\ -\frac{1}{2}<x<\frac{1}{2}$ and let the conditional density of Y be $f\left(\frac{y}{x}\right)=\begin{cases}1,\ x<y<x+1,\ -\frac{1}{2}<x<0\\ 1,\ -x<y<1-x,\ 0<x<\frac{1}{2}\end{cases}$. Prove that the variables X and Y are uncorrelated.

b). Given $f(x,y)=xe^{-x(y+1)},\ x\geq 0,\ y\geq 0$. Find the regression curve of Y on X.

Solution:

a). We have $E(X)=\int_{-\frac{1}{2}}^{\frac{1}{2}} xf(x)dx=\int_{-\frac{1}{2}}^{\frac{1}{2}} xdx=\left[\frac{x^2}{2}\right]_{-\frac{1}{2}}^{\frac{1}{2}}=0$

$$E(XY)=\int_{-\frac{1}{2}}^{0}\int_{x}^{x+1} xydxdy+\int_{0}^{\frac{1}{2}}\int_{-x}^{1-x} xydxdy$$

$$=\int_{-\frac{1}{2}}^{0} x\left[\int_{x}^{x+1} ydy\right]dx+\int_{0}^{\frac{1}{2}} x\left[\int_{-x}^{1-x} ydy\right]dx$$

$$=\frac{1}{2}\int_{-\frac{1}{2}}^{0} x(2x+1)dx+\frac{1}{2}\int_{0}^{\frac{1}{2}} x(1-2x)dx$$

$$=\frac{1}{2}\left[\frac{2x^3}{3}+\frac{x^2}{2}\right]_{-\frac{1}{2}}^{0}+\frac{1}{2}\left[\frac{x^2}{2}-\frac{2x^3}{3}\right]_{0}^{\frac{1}{2}}=0$$

Since $Cov(X,Y)=E(XY)-E(X)E(Y)=0$, the variables X and Y are uncorrelated.

b). Regression curve of Y on X is $E\left(\frac{y}{x}\right)$

$$E\left(\frac{y}{x}\right)=\int_{-\infty}^{\infty} yf\left(\frac{y}{x}\right)dy$$

$$f(y/x)=\frac{f(x,y)}{f_X(X)}$$

Marginal density function $f_X(x)=\int_{0}^{\infty} f(x,y)dy$

$$= x\int_{0}^{\infty} e^{-x(y+1)}dy$$

$$= x\left[\frac{e^{-x(y+1)}}{-x}\right]_0^{\infty} = e^{-x},\ \ x \ge 0$$

Conditional pdf of Y on X is $f\left(y/x\right) = \frac{f(x,y)}{f_X(x)} = \frac{xe^{-xy-x}}{e^{-x}} = xe^{-xy}$

The regression curve of Y on X is given by

$$E\left(y/x\right) = \int_{0}^{\infty} yxe^{-xy}dy$$

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

$E\left(y/x\right) = \frac{1}{x} \Rightarrow y = \frac{1}{x}$ and hence $xy = 1$.

24.a). Given $f(x,y) = \begin{cases} \frac{x+y}{3}, & 0 < x < 1,\ 0 < y < 2 \\ 0 & , \textit{otherwise} \end{cases}$, obtain the regression of Y on X and X on Y.

b). Distinguish between correlation and regression Analysis

Solution:

a). Regression of Y on X is $E\left(Y/X\right)$

$$E\left(Y/X\right) = \int_{-\alpha}^{\alpha} yf\left(y/x\right)dy$$

$$f\left(Y/X\right) = \frac{f(x,y)}{f_X(x)}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

$$= \int_{0}^{2}\left(\frac{x+y}{3}\right)dy = \frac{1}{3}\left[xy + \frac{y^2}{2}\right]_0^2$$

$$= \frac{2(x+1)}{3}$$

$$f\left(Y/X\right) = \frac{f(x,y)}{f_X(x)} = \frac{x+y}{2(x+1)}$$

Regression of Y on X = $E\left(Y/X\right) = \int_{0}^{2} \frac{y(x+y)}{2(x+1)}dy$

$$= \frac{1}{2(x+1)}\left[\frac{xy^2}{2}+\frac{y^3}{3}\right]_0^2$$

$$= \frac{1}{2(x+1)}\left[2x+\frac{8}{3}\right] = \frac{3x+4}{3(x+1)}$$

$$E\left(X/Y\right) = \int_{-\infty}^{\infty} xf\left(x/y\right)dx$$

$$f\left(x/y\right) = \frac{f(x,y)}{f_Y(y)}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

$$= \int_0^1 \left(\frac{x+y}{3}\right)dx = \frac{1}{3}\left[\frac{x^2}{2}+xy\right]_0^1$$

$$= \frac{1}{3}\left[\frac{1}{2}+y\right]$$

$$f\left(x/y\right) = \frac{2(x+y)}{2y+1}$$

Regression of X on $Y = E\left(X/Y\right) = \int_0^1 \frac{x+y}{2y+1}dx$

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

$$= \frac{\frac{1}{2}+y}{2y+1} = \frac{1}{2}.$$

b).

1. Correlation means relationship between two variables and Regression is a Mathematical Measure of expressing the average relationship between the two variables.
2. Correlation need not imply cause and effect relationship between the variables. Regression analysis clearly indicates the cause and effect relationship between Variables.
3. Correlation coefficient is symmetric i.e. $r_{xy} = r_{yx}$ where regression coefficient is not symmetric
4. Correlation coefficient is the measure of the direction and degree of linear relationship between two variables. In regression using the relationship between two variables we can predict the dependent variable value for any given independent variable value.

25.a). X any Y are two random variables with variances σ_x^2 and σ_y^2 respectively and r is the coefficient of correlation between them. If $U = X + KY$ and $V = X + \frac{y\sigma_x}{\sigma_y}$, find the value of k so that U and V are uncorrelated.

b). Find the regression lines:

X	6	8	10	18	20	23
Y	40	36	20	14	10	2

Solution:

Given $U = X + KY$

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

$$V = X + \frac{\sigma_X}{\sigma_Y} Y$$

$$E(V) = E(X) + \frac{\sigma_X}{\sigma_Y} E(Y)$$

If U and V are uncorrelated, $Cov(U,V) = 0$

$$E\left[\left(U - E(U)\right)\left(V - E(V)\right)\right] = 0$$

$$\Rightarrow E\left[\left(X + KY - E(X) - KE(Y)\right) \times \left(X + \frac{\sigma_X}{\sigma_Y} Y - E(X) - \frac{\sigma_X}{\sigma_Y} E(Y)\right)\right] = 0$$

$$\Rightarrow E\left\{\left[\left(X - E(X)\right) + K\left(Y - E(Y)\right)\right] \times \left[\left(X - E(X)\right) + \frac{\sigma_X}{\sigma_Y}\left(Y - E(Y)\right)\right]\right\} = 0$$

$$\Rightarrow E\left\{\left(X - E(X)\right)^2 + \frac{\sigma_X}{\sigma_Y}\left(X - E(X)\right)\left(Y - E(Y)\right) + K\left(Y - E(Y)\right)\left(X - E(X)\right) + K\frac{\sigma_X}{\sigma_Y}\left(Y - E(Y)\right)^2\right\} = 0$$

$$\Rightarrow V(X) + \frac{\sigma_X}{\sigma_Y} Cov(X,Y) + KCov(X,Y) + K\frac{\sigma_X}{\sigma_Y} V(Y) = 0$$

$$K\left[Cov(X,Y) + \frac{\sigma_X}{\sigma_Y} V(Y)\right] = -V(X) - \frac{\sigma_X}{\sigma_Y} Cov(x,y)$$

$$K = \frac{-V(X) - \frac{\sigma_X}{\sigma_Y} r\sigma_X\sigma_Y}{r\sigma_X\sigma_Y + \frac{\sigma_X}{\sigma_Y} V(Y)} = \frac{-\sigma_X^2 - r\sigma_X^2}{r\sigma_X\sigma_Y + \sigma_X\sigma_Y}$$

$$= \frac{-\sigma_X^2(1+r)}{\sigma_X\sigma_Y(1+r)} = -\frac{\sigma_X}{\sigma_Y}.$$

b).

X	Y	X^2	Y^2	XY
6	40	36	1600	240

8	36	64	1296	288
10	20	100	400	200
18	14	324	196	252
20	10	400	100	200
23	2	529	4	46
$\sum X = 85$	$\sum Y = 122$	$\sum X^2 = 1453$	$\sum Y^2 = 3596$	$\sum XY = 1226$

$\overline{X} = \frac{\sum x}{n} = \frac{85}{6} = 14.17$, $\overline{Y} = \frac{\sum y}{n} = \frac{122}{6} = 20.33$

$$\sigma_x = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{1453}{6} - \left(\frac{85}{6}\right)^2} = 6.44$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2} = \sqrt{\frac{3596}{6} - \left(\frac{122}{6}\right)^2} = 13.63$$

$$r = \frac{\frac{\sum xy}{n} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y} = \frac{\frac{1226}{6} - (14.17)(20.33)}{(6.44)(13.63)} = -0.95$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = -0.95 \times \frac{6.44}{13.63} = -0.45$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = -0.95 \times \frac{13.63}{6.44} = -2.01$$

The regression line X on Y is

$x - \overline{x} = b_{xy}\left(y - \overline{y}\right) \Rightarrow x - 14.17 = -0.45\left(y - \overline{y}\right)$

$\Rightarrow x = -0.45y + 23.32$

The regression line Y on X is

$y - \overline{y} = b_{yx}\left(x - \overline{x}\right) \Rightarrow y - 20.33 = -2.01\left(x - 14.17\right)$

$\Rightarrow y = -2.01x + 48.81$

26. a) Using the given information given below compute $\overline{x}, \overline{y}$ and r. Also compute σ_y when $\sigma_x = 2,\ 2x + 3y = 8$ and $4x + y = 10$.

b) The joint pdf of X and Y is

Y \ X	-1	1

0	$\frac{1}{8}$	$\frac{3}{8}$
1	$\frac{2}{8}$	$\frac{2}{8}$

Find the correlation coefficient of X and Y.

Solution:

a). When the regression equation are Known the arithmetic means are computed by solving the equation.

$2x+3y=8$ ------------ (1)

$4x+y=10$ ------------ (2)

$(1)\times 2 \Rightarrow 4x+6y=16$ ------- (3)

$(2)-(3) \Rightarrow -5y=-6$

$\Rightarrow y=\frac{6}{5}$

Equation $(1) \Rightarrow 2x+3\left(\frac{6}{5}\right)=8$

$\Rightarrow 2x=8-\frac{18}{5}$

$\Rightarrow x=\frac{11}{5}$

i.e. $\bar{x}=\frac{11}{5}$ & $\bar{y}=\frac{6}{5}$

To find r, Let $2x+3y=8$ be the regression equation of X on Y.

$2x=8-3y \Rightarrow x=4-\frac{3}{2}y$

$\Rightarrow b_{xy}=$Coefficient of Y in the equation of X on $Y=-\frac{3}{2}$

Let $4x+y=10$ be the regression equation of Y on X

$\Rightarrow y=10-4x$

$\Rightarrow b_{yx}=$coefficient of X in the equation of Y on $X=-4$.

$$r=\pm\sqrt{b_{xy}b_{yx}}$$

$$=-\sqrt{\left(-\frac{3}{2}\right)(-4)} \qquad (\because b_{xy} \text{ \& } b_{yx} \text{ are negative })$$

$$=-2.45$$

Since r is not in the range of $(-1\le r\le 1)$ the assumption is wrong.

Now let equation (1) be the equation of Y on X

$\Rightarrow y=\frac{8}{3}-\frac{2x}{3}$

$\Rightarrow b_{yx}=$Coefficient of X in the equation of Y on X

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

from equation (2) be the equation of X on Y

$$b_{xy} = -\frac{1}{4}$$

$$r = \pm\sqrt{b_{xy}b_{yx}} \qquad = \sqrt{-\frac{2}{3}\times -\frac{1}{4}} = 0.4081$$

To compute σ_y from equation (4) $b_{yx} = -\frac{2}{3}$

But we know that $b_{yx} = r\frac{\sigma_y}{\sigma_x}$

$$\Rightarrow -\frac{2}{3} = 0.4081\times\frac{\sigma_y}{2}$$

$$\Rightarrow \sigma_y = -3.26$$

b). Marginal probability mass function of X is

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

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

Marginal probability mass function of Y is

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

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

$$E(X) = \sum_x x\,p(x) = 0\times\frac{4}{8}+1\times\frac{4}{8} = \frac{4}{8}$$

$$E(Y) = \sum_y y\,p(y) = -1\times\frac{3}{8}+1\times\frac{5}{8} = -\frac{3}{8}+\frac{5}{8} = \frac{2}{8}$$

$$E(X^2) = \sum_x x^2\,p(x) = 0^2\times\frac{4}{8}+1^2\times\frac{4}{8} = \frac{4}{8}$$

$$E(Y^2) = \sum_y y^2\,p(y) = (-1)^2\times\frac{3}{8}+1^2\times\frac{5}{8} = \frac{3}{8}+\frac{5}{8} = 1$$

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

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

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

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

$$E(XY)=\sum_x\sum_y xy\,p(x,y)$$

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

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

$$r=\frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}=\frac{-\frac{1}{8}}{\sqrt{\frac{1}{4}}\sqrt{\frac{15}{16}}}=-0.26\,.$$

27. a) 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

b) If X and Y are independent exponential variates with parameters 1, find the pdf of $U=X-Y$.

Solution:

X	Y	XY	X^2	Y^2
65	67	4355	4225	4489
66	68	4488	4359	4624
67	65	4355	4489	4285
68	72	4896	4624	5184
69	72	4968	4761	5184
70	69	4830	4900	4761
72	71	5112	5184	5041
$\sum X=544$	$\sum Y=552$	$\sum XY=37560$	$\sum X^2=37028$	$\sum Y^2=38132$

$$\overline{X}=\frac{\sum x}{n}=\frac{544}{8}=68$$

$$\overline{Y}=\frac{\sum y}{n}=\frac{552}{8}=69$$

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

$$\sigma_X=\sqrt{\frac{1}{n}\sum x^2-\overline{X}^2}=\sqrt{\frac{1}{8}(37028)-68^2}=\sqrt{4628.5-4624}=2.121$$

$$\sigma_Y=\sqrt{\frac{1}{n}\sum y^2-y^2}=\sqrt{\frac{1}{8}(38132)-69^2}=\sqrt{4766.5-4761}=2.345$$

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

The correlation coefficient of X and Y is given by

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

b). Given that X and Y are exponential variates with parameters 1

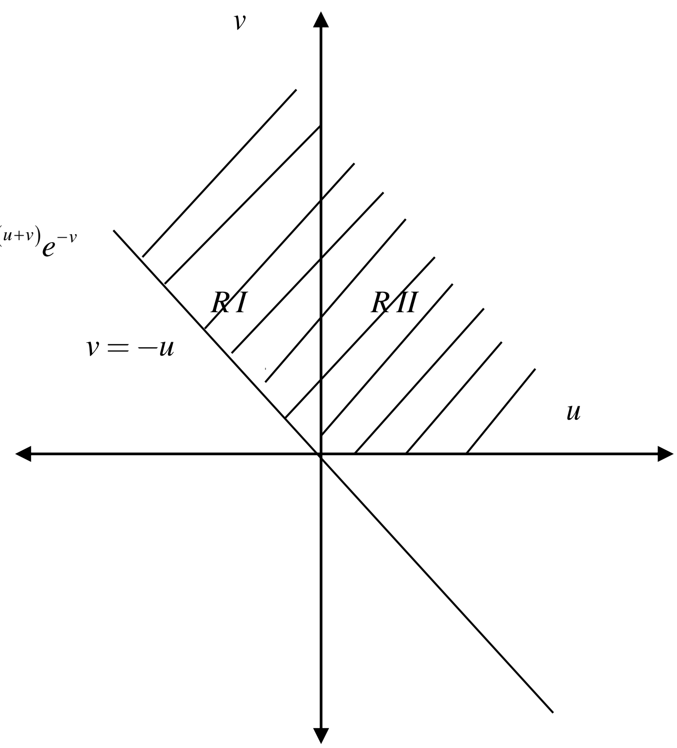
$$f_X(x) = e^{-x},\ x \ge 0,\ f_Y(y) = e^{-y},\ y \ge 0$$

Also $f_{XY}(x,y) = f_X(x) f_y(y)$ since X and Y are independent

$$= e^{-x}e^{-y}$$
$$= e^{-(x+y)};\ x \ge 0,\ y \ge 0$$

Consider the transformations $u = x - y$ and $v = y$

$\Rightarrow x = u + v,\ y = v$



$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

$$f_{UV}(u,v) = f_{XY}(x,y)|J| = e^{-x}e^{-y} = e^{-(u+v)}e^{-v}$$
$$= e^{-(u+2v)},\ u + v \ge 0,\ v \ge 0$$

In Region I when $u < 0$

$$f(u) = \int_{-u}^{\infty} f(u,v)\,dv = \int_{-u}^{\infty} e^{-u}.e^{-2v}\,dv$$
$$= e^{-u}\left[\frac{e^{-2v}}{-2}\right]_{-u}^{\infty}$$
$$= \frac{e^{-u}}{-2}\left[0 - e^{2u}\right] = \frac{e^{u}}{2}$$

In Region II when $u > 0$

$$f(u) = \int_{0}^{\infty} f(u,v)\,dv$$
$$= \int_{0}^{\infty} e^{-(u+2v)}\,dv = \frac{e^{-u}}{2}$$

$$\therefore f(u) = \begin{cases} \dfrac{e^{u}}{2}, & u < 0 \\ \dfrac{e^{-u}}{2}, & u > 0 \end{cases}$$

28. a) The joint pdf of X and Y is given by $f(x,y)=e^{-(x+y)}, x>0, y>0$. Find the pdf of $U=\dfrac{X+Y}{2}$.

b) If X and Y are independent random variables each following $N(0,2)$, find the pdf of $Z=2X+3Y$. If X and Y are independent rectangular variates on $(0,1)$ find the distribution of $\dfrac{X}{Y}$.

Solution:

a). Consider the transformation $u=\dfrac{x+y}{2}$ $\& v=y$

$$\Rightarrow x=2u-v \text{ and } y=v$$

$$J=\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}=\begin{vmatrix}2 & -1\\ 0 & 1\end{vmatrix}=2$$

$$f_{UV}(u,v)=f_{XY}(x,y)|J|$$

$$=e^{-(x+y)}2=2e^{-(x+y)}=2e^{-(2u-v+v)}$$

$$=2e^{-2u},\ 2u-v\geq 0,\ v\geq 0$$

$$f_{UV}(u,v)=2e^{-2u},\ u\geq 0,\ 0\leq v\leq \frac{u}{2}$$

$$f(u)=\int_0^{\frac{u}{2}} f_{UV}(u,v)\,dv=\int_0^{\frac{u}{2}} 2e^{-2u}dv$$

$$=\left[2e^{-2u}v\right]_0^{\frac{u}{2}}$$

$$f(u)=\begin{cases}2\dfrac{u}{2}e^{-2u}, & u\geq 0\\ 0 & , otherwise\end{cases}$$

b).(i) Consider the transformations $w=y$,

i.e. $z=2x+3y$ and $w=y$

i.e. $x=\dfrac{1}{2}(z-3w), y=w$

$$|J|=\frac{\partial(x,y)}{\partial(z,w)}=\begin{vmatrix}\frac{\partial x}{\partial z} & \frac{\partial x}{\partial w}\\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w}\end{vmatrix}=\begin{vmatrix}\frac{1}{2} & -\frac{3}{2}\\ 0 & 1\end{vmatrix}=\frac{1}{2}.$$

Given that X and Y are independent random variables following $N(0,2)$

$$\therefore f_{XY}(x,y)=\frac{1}{8\pi}e^{\frac{-\left(x^2+y^2\right)}{8}},-\infty<x,y<\infty$$

The joint pdf of (z,w) is given by

$$f_{ZW}(z,w)=|J|\,f_{XY}(x,y)$$

$$=\frac{1}{2}\cdot\frac{1}{8\pi}e^{\frac{-\left[\frac{1}{4}(z-3w)^2+w^2\right]}{8}}$$

$$=\frac{1}{16\pi}e^{-\frac{1}{32}\left[(z-3w)^2+4w^2\right]},-\infty<z,w<\infty\ .$$

The pdf of z is the marginal pdf obtained by interchanging $f_{ZW}(z,w)$ w.r.to w over the range of w.

$$\therefore f_Z(z)=\frac{1}{16\pi}\int_{-\infty}^{\infty}\left(e^{-\frac{1}{32}\left(z^2-6wz+13w^2\right)}\right)dw$$

$$=\frac{1}{16\pi}e^{-\frac{z^2}{32}}\int_{-\infty}^{\infty}\left(e^{-\frac{13}{32}\left(w^2-\frac{6wz}{13}+\left(\frac{3z}{13}\right)^2-\left(\frac{3z}{13}\right)^2\right)}\right)dw$$

$$=\frac{1}{16\pi}e^{-\frac{z^2}{32}+\frac{9z^2}{13\times32}}\int_{-\infty}^{\infty}\left(e^{-\frac{13}{32}\left(w-\frac{3z}{13}\right)^2}\right)dw$$

$$=\frac{1}{16\pi}e^{-\frac{z^2}{8\times13}}\int_{-\infty}^{\infty}e^{-\frac{13}{32}t^2}dt$$

$$r=\frac{13}{32}t^2\Rightarrow dr=\frac{13}{16}tdt\Rightarrow\frac{16}{13t}dr=dt\Rightarrow\sqrt{\frac{r32}{13}}dr=dt$$

$$\frac{16}{13}\sqrt{\frac{13}{r32}}dr=dt\Rightarrow\frac{4}{\sqrt{13}\times\sqrt{2}}r^{-\frac{1}{2}}dr=dt$$

$$=\frac{2}{16\pi}\frac{4}{\sqrt{13}\times\sqrt{2}}e^{-\frac{z^2}{8\times13}}\int_0^{\infty}e^{-r}r^{-\frac{1}{2}}dr$$

$$=\frac{1}{2\pi\sqrt{13}\times\sqrt{2}}e^{-\frac{z^2}{8\times13}}\int_0^{\infty}e^{-r}r^{-\frac{1}{2}}dr$$

$$=\frac{1}{2\pi\sqrt{13}\times\sqrt{2}}e^{-\frac{z^2}{8\times13}}\sqrt{\pi}=\frac{1}{2\sqrt{13}\sqrt{2\pi}}e^{-\frac{z^2}{2\left(2\sqrt{13}\right)^2}}$$

i.e. $Z\sim N\left(0,2\sqrt{13}\right)$

b).(ii) Given that X and Y are uniform Variants over $(0,1)$

$$\therefore f_X(x)=\begin{cases}1, & 0<x<1\\ 0, & otherwise\end{cases}\text{ and }f_Y(y)=\begin{cases}1, & 0<y<1\\ 0, & otherwise\end{cases}$$

Since X and Y are independent,

$$f_{XY}(x,y) = f_X(x) f_y(y) \begin{cases} 1, & 0 < x, y < 1 \\ 0, & otherwise \end{cases}$$

Consider the transformation $u = \dfrac{x}{y}$ and $v = y$

i.e. $x = uv$ and $y = v$

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & 0 \\ u & 1 \end{vmatrix} = v$$

$$\therefore f_{UV}(u,v) = f_{XY}(x,y)|J|$$
$$= v,\ 0 < u < \infty,\ 0 < v < \infty$$

The range for u and v are identified as follows.

$0 < x < 1$ and $0 < y < 1$.

$\Rightarrow 0 < uv < 1$ and $0 < v < 1$

$\Rightarrow uv > 0,\ uv < 1,\ v > 0$ and $v < 1$

$\Rightarrow uv > 0$ and $v > 0 \Rightarrow u > 0$

Now $f(u) = \int f_{UV}(u,v)\,dv$

The range for v differs in two regions

$$f(u) = \int_0^1 f_{UV}(u,v)\,dv$$

$$= \int_0^1 v\,dv = \left[\frac{v^2}{2}\right]_0^1 = \frac{1}{2},\ 0 < u < 1$$

$$f(u) = \int_0^{\frac{1}{u}} f_{UV}(u,v)\,dv$$

$$= \int_0^{\frac{1}{u}} v\,dv = \left[\frac{v^2}{2}\right]_0^{\frac{1}{u}} = \frac{1}{2u^2},\ 1 \le u \le \infty$$

$$\therefore f(u) = \begin{cases} \dfrac{1}{2} &, \ 0 \le u \le 1 \\ \dfrac{1}{2u^2}, & u > 1 \end{cases}$$