2.1 JOINT DISTRIBUTION

Two Dimensional Random variables:

Let S be the sample space. Let $X = X(s)$ and $Y = Y(s)$ be two functions each assigning a real number to each outcome s ∈ S . Then (X, Y) is a two dimensional random variable.

Types of Random Variables:

(i) Discrete Random Variables

(ii) Continuous random variables

Discrete random Variables:

Two Dimensional Discrete Random variable

If the possible values (X, Y) are finite, then (X, Y) is called a two – dimensional discrete random variable and it can be represented by (x_i, y_j), $i = 1, 2, . . . , n; j = 1,2, \dots, m$

Two Dimensional Continuous Random variable

If (X, Y) can take all the values in a region R in the XY plane, then (X, Y) is called a two – dimensional continuous random variable.

Discrete Random variables:

In the study of two dimensional discrete random variables we have the following 5 important terms.

(i) Joint Probability function (or) Joint Probability mass function (PMF)

(ii) Joint Probability distribution

(iii) Marginal probability function of X

(iv) Marginal probability function of Y

(v)Conditional Probability function

Joint Probability Function (or) Joint Probability mass function

Let X, Y be a two dimensional discrete random variable for each possible outcome (X_i, Y_j). We associate a number $P(X_i, Y_j)$ representing $[X = x_i, Y = y_j]$ and satisfies the following conditions

i) $P[x_i, y_j] \geq 0$

ii) $\sum\sum p(x_i, y_j) = 1$

The function $P[x_i, y_j]$ is called joint probability mass function of x, y

Conditional distribution of *X* given *Y*

$$P[X = x_i / Y = y_j] = \frac{P[X = x_i \cap Y = y_j]}{P[Y = y_j]}$$

$$= \frac{P[X = x_i, Y = y_j]}{P[Y = y_j]}$$

$$P[Y = y_j / X = x_i] = \frac{P[Y = y_j \cap X = x_i]}{P[X = x_i]}$$

$$= \frac{P[X = x_i, Y = y_j]}{P[X = x_i]}$$

Test of independent:

$$P[X = x_i, Y = y_j] = P[X = x_i], P[Y = y_j]$$

Problems on Marginal distribution:

1.The joint probability marginal function of X , Y is given by $P(xy) = K(2x + 3y)$, $x = 0, 1, 2$, $y = 1, 2, 3$ find K. Find all the marginal distribution and conditional probability distribution . Also probability distribution X + Y.

Solution:

	1	2	3	$\sum x$
0	3K	6K	9K	18K
1	5K	8K	11K	24K
2	7K	10K	13K	30K
$\sum y$	15K	24K	33K	72K

We know that $\sum \sum P(x, y) = 1$

$$\Rightarrow 72K = 1 \Rightarrow K = \frac{1}{72}$$

Marginal distribution

X	0	1	2
P(X)	$\frac{18}{72}$	$\frac{24}{72}$	$\frac{30}{72}$

Y	1	2	3
P(Y)	$\frac{15}{72}$	$\frac{24}{72}$	$\frac{33}{72}$

Conditional distribution at x given y

$$P[X = 0/Y = 1] = \frac{P[X = 0, Y = 1]}{P[Y = 1]} = \frac{3/72}{15/72} = \frac{1}{5}$$

$$P[X = 2/Y = 2] = \frac{P[X = 2, Y = 2]}{P[Y = 2]} = \frac{6/72}{24/72} = \frac{1}{4}$$

$$P[X = 0/Y = 3] = \frac{P[X = 0, Y = 3]}{P[Y = 3]} = \frac{9/72}{33/72} = \frac{3}{11}$$

$$P[X = 1/Y = 1] = \frac{P[X = 1, Y = 2]}{P[Y = 1]} = \frac{5/72}{15/72} = \frac{1}{3}$$

$$P[X = 1/Y = 2] = \frac{P[X = 1, Y = 2]}{P[Y = 2]} = \frac{8/72}{24/72} = \frac{1}{3}$$

$$P[X = 1/Y = 3] = \frac{P[X = 1, Y = 3]}{P[Y = 3]} = \frac{11/72}{33/72} = \frac{1}{3}$$

$$P[X = 2/Y = 1] = \frac{P[X = 2, Y = 1]}{P[Y = 1]} = \frac{7/72}{15/72} = \frac{7}{15}$$

$$P[X = 2/Y = 2] = \frac{P[X = 2, Y = 2]}{P[Y = 2]} = \frac{10/72}{24/72} = \frac{5}{12}$$

$$P[X = 2/Y = 3] = \frac{P[X = 2, Y = 3]}{P[Y = 3]} = \frac{13/72}{33/72} = \frac{13}{33}$$

Conditional distribution at y given x

$$P[Y = 1/X = 0] = \frac{P[Y = 1, X = 0]}{P[X = 0]} = \frac{3/72}{18/72} = \frac{1}{6}$$

$$P[Y = 1/X = 1] = \frac{P[Y = 1, X = 1]}{P[X = 1]} = \frac{5/72}{24/72} = \frac{5}{24}$$

$$P[Y = 1/X = 2] = \frac{P[Y = 1, X = 2]}{P[X = 2]} = \frac{7/72}{30/72} = \frac{7}{30}$$

$$P[Y = 2/X = 0] = \frac{P[Y = 2, X = 0]}{P[X = 0]} = \frac{6/72}{18/72} = \frac{1}{3}$$

$$P[Y = 2/X = 1] = \frac{P[Y = 2, X = 1]}{P[X = 1]} = \frac{8/72}{24/72} = \frac{1}{3}$$

$$P[Y = 2/X = 2] = \frac{P[Y = 2, X = 2]}{P[X = 2]} = \frac{10/72}{30/72} = \frac{1}{3}$$

$$P[Y = 3/X = 0] = \frac{P[Y = 3, X = 0]}{P[X = 0]} = \frac{9/72}{18/72} = \frac{1}{2}$$

$$P[Y = 3/X = 1] = \frac{P[Y = 3, X = 1]}{P[X = 1]} = \frac{11/72}{24/72} = \frac{11}{24}$$

$$P[Y = 3/X = 2] = \frac{P[Y = 3, X = 2]}{P[X = 2]} = \frac{13/72}{30/72} = \frac{13}{30}$$

Distribution of $x + y$

1	P_{01}	$3/72$
2	$P_{02} + P_{11}$	$11/72$
3	$P_{03} + P_{12} + P_{21}$	$24/72$
4	$P_{13} + P_{22}$	$21/72$

5	P_{23}	$^{13}/_{72}$

2. The joint distribution of X and Y is given by $\boldsymbol{f(x,y) = \frac{x+y}{21}, x = 1,2,3\ ; y = 1,2.}$

Find the marginal distributions.

Solution:

Given $f(x,y) = \frac{x+y}{21}, x = 1,2,3\ ; y = 1,2$

$$\Rightarrow f(1,1) = \frac{2}{21}$$

$$\Rightarrow f(1,2) = \frac{3}{21}$$

$$\Rightarrow f(2,1) = \frac{3}{21},$$

$$\Rightarrow f(2,2) = \frac{4}{21},$$

$$\Rightarrow f(3,1) = \frac{4}{21},$$

$$\Rightarrow f(3,2) = \frac{5}{21}$$

The marginal distributions are given in the table.

		X			$P_Y(y) = P(Y=y)$
		1	2	3	
Y	1	$\frac{2}{21}$ P(1,1)	$\frac{3}{21}$ P(2,1)	$\frac{4}{21}$ P(3,1)	$\frac{9}{21}$
	2	$\frac{3}{21}$ P(1,2)	$\frac{4}{21}$ P(2,2)	$\frac{5}{21}$ P(3,2)	$\frac{12}{21}$ P(1,1)
$P_X(x) = P(X=x)$		$\frac{5}{21}$	$\frac{7}{21}$	$\frac{9}{21}$	1

The marginal distribution of X

$P_X(1) = P(X=1) = \frac{5}{21}$, $P_X(2) = P(X=2) = \frac{7}{21}$, $P_X(3) = P(X=3) = \frac{9}{21}$

The marginal distribution of Y

$P_Y(1) = P(Y=1) = \frac{9}{21}$, $P_Y(2) = P(Y=2) = \frac{12}{21}$

2. 3. Covariance

If X and Y are random variables, then covariance between X and Y is defined as

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

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

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

$$Cov\ (X, Y) = E(XY) - E(X).E(Y) \qquad ...(A)$$

If X and Y are independent then $E(XY) = E(X)E(Y) \qquad ...(B)$

Sub (B) in (A), we get $Cov\ (X, Y) = 0$

Therefore, if X and Y are independent then $Cov\ (X, Y) = 0$

Correlation:

If the change in one variable affects a change in the other variable, the variables are said to be correlated.

Two types of correlations are Positive correlation, Negative correlation

Positive Correlation:

If the two variables deviate in the same direction

Eg: Height and Weight of a group of persons, Income and Expenditure.

Negative Correlation:

If the two variables constantly deviate in opposite directions.

Eg: Price and Demand of a commodity, the correlation between volume and pressure of a perfect gas.

Measurement of Correlation:

We can measure the correlation between the two variables by using Karl – Pearson's coefficient of correlation.

Karl - Pearson' s coefficient of correlation

Correlation coefficient between two random variables X and Y, usually denoted by

$$r(X,Y) = \frac{COV(X,Y)}{\sigma_X.\sigma_Y}$$

Where $COV(X,Y) = \frac{1}{n}\sum XY - \bar{X}\bar{Y}$

$$\sigma_X = \sqrt{\frac{1}{n}\sum X^2 - \bar{X}^2}, \bar{X} = \frac{\sum X}{n}$$

$$\sigma_Y = \sqrt{\frac{1}{n}\sum Y^2 - \bar{Y}^2}, \bar{Y} = \frac{\sum Y}{n}$$

(n is the number of items in the given data)

Note:

1. Correlation coefficient may also be denoted by $\rho(X, Y) or\ \rho_{XY}$
2. If $\rho(X, Y) = 0$, We say that X and Y are uncorrelated.
3. Correlation coefficient does not exceed unity.

Note:

Types of correlation based on 'r'

Value of 'r'	Correlation is said to be
$r = 1$	Perfect and positive
$0 < r < 1$	Positive
$-1 < r < 0$	Negative
$r = 0$	Uncorrelated

Problems under Correlation

1. Calculate the correlation coefficient for the following heights (in inches) of father *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
$\sum(X) = 544$	$\sum(Y) = 552$	$\sum(XY) = 37560$	$\sum(X^2) = 37028$	$\sum(Y^2) = 38132$

$\bar{X} = \frac{544}{8} = 68$, $\bar{Y} = \frac{552}{8} = 69$

$\bar{X}\bar{Y} = 68 X 69 = 4692$

$\sigma_X = \sqrt{\frac{1}{n}\sum X^2 - \bar{X}^2} = 2.121$

$\sigma_Y = \sqrt{\frac{1}{n}\sum Y^2 - \bar{Y}^2} = 2.345$

$$r(X,Y) = \frac{COV(X,Y)}{\sigma_X.\sigma_Y} = \frac{\frac{1}{n}\sum XY - \bar{X}\bar{Y}}{\sigma_X.\sigma_Y} = \frac{\frac{1}{8}37560 - 4692}{2.121\ X\ 2.345} = 0.6031$$

Find the correlation coefficient between industrial production and export using the following data:

Production (X)	55	56	58	59	60	60	62
Export (Y)	**35**	**38**	**37**	**39**	**44**	**43**	**44**

Solution:

U	V	$U = X - 58$	$V = Y - 40$	UV	U^2	V^2
55	35	-3	-5	15	9	25
56	38	-2	-2	4	4	4
58	37	0	-3	0	0	9
59	39	1	-1	-1	1	1
60	44	2	4	8	4	16
60	43	2	3	6	4	9
62	44	4	4	16	16	16
		$\sum(U) = 4$	$\sum(V) = 0$	$\sum(UV) = 48$	$\sum(U^2) = 38$	$\sum(V^2) = 80$

Now $\bar{U} = \frac{\sum U}{n} = \frac{4}{7} = 0.5714$

$\bar{V} = \frac{\sum V}{n} = 0$

$Cov\,(U,V) = \frac{\sum UV}{n} - \bar{U}\bar{V} = 6.857$

$\sigma_X = \sqrt{\frac{1}{n}\sum U^2 - \bar{U}^2} = 2.2588$

$\sigma_Y = \sqrt{\frac{1}{n}\sum V^2 - \bar{V}^2} = 3.38$

$r\,(U,V) = \frac{COV(U,V)}{\sigma_U.\sigma_V} = \frac{\frac{1}{n}\sum UV - \bar{U}\bar{V}}{\sigma_U.\sigma_V} = 0.898$

3. Two R.V.'S X and Y have joint p.d.f of $f(x,y) = \begin{cases} \frac{xy}{96} & 0 < x < 4,\ 1 < y < 5 \\ 0 & elsewhere \end{cases}$

Find (i) $E(X)$ (ii) $E(Y)$ (iii) $E(XY)$ (iv) $E(2X + 3Y)$ (v) $Var(X)$ (vi) $Var(Y)$ (vii) $Cov(X,Y)$

Solution:

(i) $E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x,y)\,dxdy$

$= \int_1^5\int_0^4 x\,\frac{xy}{96}\,dxdy$

$$= \frac{1}{96}\int_1^5\int_0^4 x^2 y\,dxdy$$

$$= \frac{1}{96}\int_1^5 y\left(\frac{x^3}{3}\right)_0^4 dy$$

$$= \frac{64}{288}\int_1^5 y\,dy = \frac{2}{9}\left[\frac{y^2}{2}\right]_1^5 = \frac{24}{9}$$

(ii) $E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f(x,y)dxdy$

$$= \int_1^5\int_0^4 y\frac{xy}{96}dxdy$$

$$= \frac{1}{96}\int_1^5\int_0^4 xy^2 dxdy$$

$$= \frac{1}{96}\int_1^5 y^2\left(\frac{x^2}{2}\right)_0^4 dy$$

$$= \frac{1}{96\,(2)}\int_1^5 y^2\,(4^2 - 0)dy$$

$$= \frac{16}{192}\left[\frac{y^3}{3}\right]_1^5 = \frac{124}{36}$$

$$\Rightarrow E(Y) = \frac{31}{9}$$

(iii) $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)dxdy$

$$= \int_1^5\int_0^4 xy\left(\frac{xy}{96}\right)dxdy$$

$$= \frac{1}{96}\int_1^5\int_0^4 x^2y^2 dxdy$$

$$= \frac{1}{96}\int_1^5 y^2 \left(\frac{x^3}{3}\right)_0^4 dy$$

$$= \frac{1}{96\,(3)} \int_1^5 y^2 \,(4^3 - 0)dy$$

$$= \frac{64}{288} \int_1^5 y^2\, dy = \frac{2}{9}\left[\frac{y^3}{3}\right]_1^5 = \frac{248}{27}$$

$$\Rightarrow E(XY) = \frac{248}{27}$$

(iv) $E[2X + 3Y] = 2\,E[X] + 3\,E[Y]$

$$= 2\left(\frac{8}{3}\right) + 3\left(\frac{31}{9}\right)$$

$$= \frac{16+31}{3} = \frac{47}{3}$$

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

Now, $E(X^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2\, f(x,y)dxdy$

$$= \int_1^5 \int_0^4 x^2 \left(\frac{xy}{96}\right)dxdy$$

$$= \frac{1}{96}\int_1^5\int_0^4 x^3 y dxdy$$

$$= \frac{1}{96}\int_1^5 y\left(\frac{x^4}{4}\right)_0^4 dy$$

$$= \frac{1}{96\,(4)}\int_1^5 y\,(4^4 - 0)dy$$

$$= \frac{256}{384}\int_1^5 y\,dy = \frac{2}{3}\left[\frac{y2}{2}\right]_1^5 = \frac{24}{3} = 8$$

$$\Rightarrow E(X^2) = 8$$

$$Var\ (X\ ) = \mathrm{E}\ (X^2) - [E(X)]^2 = 8 - (\frac{8}{3})^2 = \frac{8}{9}$$

$$\sigma_X^2 = \frac{8}{9} \Rightarrow \sigma_X = \frac{\sqrt{8}}{3}$$

$$E(Y^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2\ f(x,y)dxdy$$

$$= \int_1^5 \int_0^4 y^2\ (\frac{xy}{96})dxdy$$

$$= \frac{1}{96}\int_1^5 \int_0^4 xy^3dxdy$$

$$= \frac{1}{96}\int_1^5 y^3(\frac{x^2}{2})_0^4\ dy$$

$$= \frac{1}{96\ (2)}\int_1^5 y^3\ (4^2 - 0)dy$$

$$= \frac{16}{192}\int_1^5 y^3\ dy$$

$$= \frac{1}{12}\left[\frac{y^4}{4}\right]_1^5 = \frac{624}{48} = 13$$

$$\Rightarrow E(Y^2) = 13$$

(vi) $Var\ (Y\ ) = \mathrm{E}\ (Y^2) - [E(Y)]^2$

$$= 13 - \left(\frac{31}{9}\right)^2$$

$$= \frac{92}{81}$$

$$\sigma_Y^2 = \frac{92}{81} \Rightarrow \sigma_Y = \frac{\sqrt{92}}{9}$$

(vii) $Cov\ (X, Y) = E\ (XY) - E(X)\ .E(Y)$

$$= \frac{248}{27} - \frac{248}{27} = 0$$

4. If the independent random variables X and Y have the variances 36 and 16 respectively, find the correlation coefficient between $X + Y$ and $X - Y$

Solution:

Given that $Var\ (X) = 36$, $Var\ (Y) = 16$. Since X and Y are independent,

$E(XY) = E(X)\ .E(Y)$

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

$Var\ (U) = Var\ (X + Y)$

$$= 1^2 Var\ (X) + 1^2 Var(Y)$$

$(\because Var\ (a\,X + b\,Y) = a^2\ Var(X) + b^2\ Var\ (Y))$

$$= 36 + 16 = 52$$

$\sigma_U = \sqrt{52}$

$Var\ (V) = Var\ (X - Y)$

$= 1^2 Var\ (X) + (-1)^2 Var(Y)$

$(\because Var\ (aX + bY) = a^2\ Var(X) + b^2\ Var\ (Y))$

$= 36 + 16 = 52$

$\sigma_V = \sqrt{52}$

$Cov\ (U, V) = E(UV) - E(U).E(V) \quad \ldots (1)$

$E(UV) = E\ [(X + Y)\ (X - Y)]$

$= E\ [X^2 - Y^2]$

$= E\ (X^2) - E(Y^2) \quad \ldots (2)$

$E(U) = E(X + Y) = E(X) + E(Y) \quad \ldots (3)$

$E(V) = E(X - Y) = E(X) - E(Y) \quad \ldots (4)$

Substituting (2), (3), (4) in (1) we get

$Cov(U, V) = E(X^2) - E(Y^2) - [E(X) + E(Y)][E(X) - E(Y)]$

$= E\ (X^2) - E(Y^2) - [E(X)]^2 + [E(Y)]^2 - E(X)E(Y) + E(X)E(Y)$

$= \{E(X^2) - [E(X)]^2\} - \{E(Y^2) - [E(Y)]^2\}$

$= Var(X) - Var(Y)$

$Cov\ (U,V) = 36 - 16 = 20$

Hence $\rho(U,V) = \frac{Cov(U,V)}{\sigma_U \sigma_V} = \frac{20}{\sqrt{52}\sqrt{52}} = \frac{20}{52} = \frac{5}{13}$

5. If the joint pdf of (X,Y) is given by $f(x,y) = x + y\ , 0 \leq x, y \leq 1, find$ ρ_{XY}

Solution:

Given $f(x,y) = x + y, 0 \leq x, y \leq 1$

Now, $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\ f(x,y)dxdy$

$= \int_1^1\int_0^1 xy\ (x + y)dxdy$

$= \int_1^1\int_0^1 (x^2y + xy^2)dxdy$

$= \int_1^1 [\frac{x^3y}{3} + \frac{x^2y^2}{2}]_0^1 dy$

$= \int_1^1 [\frac{y}{3} + \frac{y^2}{2}]\ dy$

$= [\frac{y^2}{6} + \frac{y^3}{6}]_0^1$

$= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

Marginal pdf of X is $f(x) = \int_0^1 f(x,y)dy$

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

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

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

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

Marginal pdf of Y is $f(y) = \int_0^1 f(x, y)dx$

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

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

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

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

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$= \int_0^1 x\left(x + \frac{1}{2}\right) dx$$

$$= \int_0^1 \left(x^2 + \frac{x}{2}\right) dx$$

$$= \left[\frac{x^3}{3} + \frac{x^2}{4}\right]_0^1$$

$$= \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$$

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy$$

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

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

$$= \left[\frac{y^3}{3} + \frac{y^2}{4}\right]_0^1$$

$$= \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

$$= \int_0^1 x^2 \left(x + \frac{1}{2}\right) dx$$

$$= \int_0^1 \left(x^3 + \frac{x^2}{2}\right) dx$$

$$= \left[\frac{x^4}{4} + \frac{x^3}{3}\right]_0^1$$

$$= \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$$

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 f(y) dy$$

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

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

$$= \left[\frac{y^4}{4} + \frac{y^3}{3}\right]_0^1$$

$$= \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$$

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

$$= \frac{5}{12} - \left[\frac{7}{12}\right]^2 = \frac{11}{144}$$

$$\sigma_X{}^2 = \frac{11}{144}$$

$$\sigma_X = \frac{\sqrt{11}}{12}$$

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

$$= \frac{5}{12} - \left[\frac{7}{12}\right]^2 = \frac{11}{144}$$

$$\sigma_Y{}^2 = \frac{11}{144}$$

$$\sigma_Y = \frac{\sqrt{11}}{12}$$

Correlation coefficient $\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_X \, \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \, \sigma_Y}$

$$= \frac{\frac{1}{3} - \frac{7}{12} \times \frac{7}{12}}{\frac{\sqrt{11}}{12} \times \frac{\sqrt{11}}{12}} = -\frac{1}{11}$$

Rank Correlation:

$$r = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n(n^2 - 1)}$$

Where $d_i = x_i - y_i$

This formula is called Spearman's formula for Rank correlation.

Problems on Rank correlation:

1. Find the rank correlation from the following data.

X	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Rank in y	**4**	**3**	**1**	**2**	**6**	**5**	**7**

Solution:

X	Y	$d_i = x_i - y_i$	${d_i}^2$
1	4	-3	9
2	3	-1	1
3	1	2	4
4	2	2	4
5	6	-1	1

6	5	1	1
7	7	0	0
		$\sum d_i = 0$	$\sum {d_i}^2 = 20$

Rank correlation coefficient

$$r = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n(n^2-1)}$$

$$r = 1 - \frac{6\times 20}{7(49-1)}$$

$$= 1 - \frac{120}{336} = 0.6429$$

2. The ranks of some 16 students in Mathematics and Physics are as follows. Calculate rank correlation coefficients for proficiency in Mathematics and Physics

Rank in Maths	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Rankin Physics	**1**	**10**	**3**	**4**	**5**	**7**	**2**	**6**	**8**	**11**	**15**	**9**	**14**	**12**	**16**	**13**

Solution:

X	Y	$d_i = x_i - y_i$	${d_i}^2$
1	1	0	0
2	10	-8	64
3	3	0	0
4	4	0	0
5	5	0	0
6	7	-1	1
7	2	5	25
8	6	2	4
9	8	1	1
10	11	-1	1
11	15	-4	16
12	9	3	9
13	14	-1	1
14	12	2	4
15	16	-1	1
16	13	3	9
		$\sum d_i = 0$	$\sum {d_i}^2 = 136$

Rank correlation coefficient

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

$$r = 1 - \frac{6 \times 136}{16(256-1)}$$

$$= 1 - \frac{816}{4080} = 0.8$$

Regression:

Regression is a mathematical measure of the average relationship between two or more

variables in terms of the original limits of the data.

Lines of Regression

If the variables in a bivariate distribution are related we will find that the points in the scattered diagram will cluster around some curve called the curve of regression. If the curve is a straight line, it is called the line of regression and there is said to be linear regression between the variables, otherwise regression is said to be curvilinear.

The line of regression of Y on X is given by

$$\boldsymbol{y - \bar{y} = r.\frac{\sigma_Y}{\sigma_X}(x - \bar{x})}$$

Where r is the correlation coefficient, σ_Y and σ_X are standard deviations.

The line of regression of Y on X is given by

$$x - \bar{x} = r\frac{\sigma_Y}{\sigma_X}(y - \bar{y})$$

Note:

Both the lines of regression passes through the mean $(\bar{x}, \bar{y})$

Angle between two lines of regression

If the equations of lines of regression of Y on X and X on Y are

$$y - \bar{y} = r.\frac{\sigma_Y}{\sigma_X}(x - \bar{x})$$

$$x - \bar{x} = r\frac{\sigma_Y}{\sigma_X}(y - \bar{y})$$

The angle θ between the two lines of regression is given by

$$tan\theta = \frac{1 - r^2}{r}\left(\frac{\sigma_Y \sigma_X}{\sigma_Y{}^2 + \sigma_X{}^2}\right)$$

Problems on Regression

1. From the following data, find (i) the two regression equations (ii) The coefficient of correlation between the marks in Economics and Statistics (iii) The most likely marks in statistics when marks in Economics are 30.

Marks in Economics	25	28	35	32	31	36	29	38	34	32
Marks in Statistics	43	46	49	41	36	32	31	30	33	39

Solution:

X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	-32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	-24
29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10

32	39	0	1	0	1	0
320	380	0	0	140	398	-93

Now $\bar{X} = \frac{\sum X}{n} = \frac{320}{10} = 32$

$\bar{Y} = \frac{\sum Y}{n} = \frac{380}{10} = 38$

Coefficient of regression of Y on X is $b_{YX} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2}$

$$= -\frac{93}{140} = -0.6643$$

Coefficient of regression of X on Y is $b_{XY} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(Y-\bar{Y})^2}$

$$= -\frac{93}{398} = -0.2337$$

Equation of the line of regression of X on Y is $x - \bar{x} = b_{XY}(y - \bar{y})$

$\Rightarrow x - 32 = -0.2337(y - 38)$

$\Rightarrow x = -0.2337y + 0.2337 \times 38 + 32$

$\Rightarrow x = -0.2337y + 40.8806$

Equation of the line of regression of Y on X is $y - \bar{y} = b_{YX}(x - \bar{x})$

$\Rightarrow y - 38 = -0.6642(y - 38)$

$\Rightarrow y = -0.6642x + 0.6642 \times 32 + 38$

$\Rightarrow y = -0.6642x + 59.2576$

Correlation of coefficient $r^2 = b_{YX} \times b_{XY}$

$= -0.6643 \times (-0.2337)$

$= 0.1552$

$r = \pm\sqrt{0.1552}$

$= \pm 0.394$

Now we have to find the most likely marks in statistics (Y) when marks in Economics (X) are 30. We use the line of regression of Y on X.

$\Rightarrow y = -0.6642x + 59.2576$

Put $x = 30$ we get

$\Rightarrow y = -0.6642 \times 30 + 59.2576$

$\Rightarrow y = 39.3286$

2. The two lines of regression are $8x - 10y + 66 = 0, 40x - 18y - 214 = 0$. The variance of X is 9. Find (i) the mean value of X and Y (ii) correlation coefficient between X and Y.

Solution:

Since both the lines of regression passes through the mean values $\bar{x}$ and $\bar{y}$ the point $(\bar{x}, \bar{y})$ must satisfy the two given regression lines.

$(1) \times 5 \Rightarrow 40\bar{x} - 50\bar{y} = -330$

$(2) \Rightarrow 40\bar{x} - 18\bar{y} = 214$

Subtracting (1) – (2) we get

$\Rightarrow 32\bar{y} = 544$

$\Rightarrow \bar{y} = 17$

Sub $\bar{y} = 17$ in (1) we get,

$(1) \Rightarrow 8\bar{x} - 10\bar{y} = -66$

$\Rightarrow 8\bar{x} - 10 \times 17 = -66$

$\Rightarrow 8\bar{x} = -66 + 170$

$\Rightarrow \bar{x} = 13$

Hence the mean value ae given by $\bar{x} = 13$ and $\bar{y} = 17$

(ii) Let us suppose that equation (A) is the equation of line of regression of Y on X and (B) is the equation of the line regression of X on Y, we get after rewriting (A) and (B)

$\Rightarrow 10y = 8x + 66$

$\Rightarrow y = \frac{8}{10}x + \frac{66}{10}$

$\Rightarrow b_{YX} = \frac{8}{10}$

$\Rightarrow 40x = 18y + 214$

$\Rightarrow x = \frac{18}{40}y + \frac{214}{40}$

$\Rightarrow b_{XY} = \frac{18}{40}$

Correlation of coefficient $r^2 = b_{YX} \times b_{XY}$

$$= \frac{8}{10} \times \frac{18}{40} = \frac{9}{25}$$

$$r = \pm\frac{3}{5}$$

$$= \pm 0.6$$

Since both the regression coefficients are positive, r must be positive.

Hence $r = 0.6$

Important note:

If we take equation (A) as the line of regression of X on Y we get,

$\Rightarrow 8x = 10y - 66$

$\Rightarrow x = \frac{10}{8}y - \frac{66}{8}$

$\Rightarrow b_{XY} = \frac{10}{8}$

$\Rightarrow 18y = 40x - 214$

$\Rightarrow y = \frac{40}{18}x - \frac{214}{18}$

$\Rightarrow b_{YX} = \frac{40}{18}$

Correlation of coefficient $r^2 = b_{YX} \times b_{XY}$

$$= \frac{10}{8} \times \frac{40}{8} = \frac{25}{9}$$

$$r = 2.78$$

But r^2 should always lies between 0 and 1. Hence our assumption that line (A) is line of regression of X on Y and the line (B) is line of regression of Y on X is wrong.

3. The two lines of regression are $4x - 5y + 33 = 0$, $20x - 9y - 107 = 0$. The variance of X is 25. Find (i) the mean value of X and Y (ii) correlation coefficient between X and Y.

Solution:

Since both the lines of regression passes through the mean values $\bar{x}$ and $\bar{y}$ the point $(\bar{x}, \bar{y})$ must satisfy the two given regression lines.

$(1) \Rightarrow 20\bar{x} - 9\bar{y} = 107$

$(2) \times 5 \Rightarrow 20\bar{x} - 25\bar{y} = -165$

Subtracting (1) – (2) we get

$\Rightarrow 16\bar{y} = 272$

$\Rightarrow \bar{y} = 17$

Sub $\bar{y} = 17$ in (1) we get,

$(2) \Rightarrow 4\bar{x} - 5\bar{y} = -33$

$\Rightarrow 4\bar{x} - 5 \times 17 = -33$

$\Rightarrow 4\bar{x} = -33 + 85$

$\Rightarrow \bar{x} = 13$

Hence the mean value ae given by $\bar{x} = 13$ and $\bar{y} = 17$

(ii) Let us suppose that equation (A) is the equation of line of regression of Y on X and (B) is the equation of the line regression of X on Y, we get after rewriting (A) and (B)

$\Rightarrow 5y = 4x + 33$

$\Rightarrow y = \frac{4}{5}x + \frac{33}{5}$

$\Rightarrow b_{YX} = \frac{4}{5}$

$\Rightarrow 20x = 9y + 107$

$\Rightarrow x = \frac{9}{20}y + \frac{107}{20}$

$\Rightarrow b_{XY} = \frac{9}{20}$

Correlation of coefficient $r^2 = b_{YX} \times b_{XY}$

$$= \frac{4}{5} \times \frac{9}{20} = \frac{3}{5}$$

$$r = \pm 0.6$$

4. Can $Y = 5 + 2.8X$ and $X = 3 - 0.5Y$ be the estimated regression equations of Y on X and X on Y respectively? Explain your answer.

Solution:

Given,

$\Rightarrow X = 3 - 0.5Y$

$\Rightarrow b_{XY} = -0.5$

$\Rightarrow Y = 5 + 2.8X$

$\Rightarrow b_{YX} = 2.8$

Correlation of coefficient $r^2 = b_{YX} \times b_{XY}$

$$= 2.8 \times (-0.5) = -1.4$$

$r = \sqrt{-1.4}$ which is imaginary quantity.

Here r cannot be imaginary.

Hence the given lines are not estimated as regression equations.

2.4 Transformation of Random Variables:

Let (X, Y) be a continuous two dimensional random variables with JPDF $f_{XY}(x, y)$. Transform X and Y to new random variables $U = h(x, y), V = g(x, y)$.

Then the joint PDF of (U, V) is given by

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

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

Procedure to find the Marginal pdf of U & V

(1) Take u as the random variable to which the PDF to be computed and take $v =$ y. (if not given)

(2) Express x and y in terms of u and v.

(3) Find $J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$

(4) Write the JPDF of (U, V), $f_{UV}(u, v) = |J|\, f_{XY}(x, y)$

(5) Substitute the values of J, x and y.

(6) Find the range of u and v using the range of x and y.

(7) The PDF of U is $f_U(u) = \int_{v=-\infty}^{v=\infty} f_{uv}(u, v)dv$

(8) The PDF of V is $f_V(v) = \int_{u=-\infty}^{u=\infty} f_{uv}(u, v)du$

Problem based on Transformation of Random Variables

1. If the JPDF $f(x, y)$ is given by $f_{XY}(x, y) = x + y; 0 \leq x, y \leq 1$, find PDF of $U = XY$.

Solution:

Given (X, Y) is a continuous $2D$ RV defined in $0 < x < 1$ and $0 < y < 1$.

Also Given $f_{xy}(x, y) = x + y$ $0 \leq x, y \leq 1$

we have to find the PDF of $u = xy$ .... (1)

let $v = y \Rightarrow y = u$.

$(1) \Rightarrow \text{u} = xv \Rightarrow x = \frac{u}{v}$

$\therefore x = \frac{u}{v}$ $\quad\quad y = v$

$\frac{\partial z}{\partial u} = \frac{1}{v}; \frac{\partial x}{\partial v} = \frac{-u}{v^2}; \frac{\partial y}{\partial u} = 0; \frac{\partial y}{\partial v} = 1$

$$J = \begin{vmatrix} \frac{1}{v} & \frac{-u}{v^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{v}$$

$$J = \frac{1}{v}$$

The JPDF of $(U,V) f_{uv}(u,v) = |J| f_{xy}(x,y)$

$$= |\frac{1}{v}| (x+y) = \frac{1}{v}(\frac{u}{v}+v)$$

$$= \frac{u}{v^2} + 1$$

$$f_{uv}(u,v) = \frac{u}{v^2} + 1$$

To find the range for u and v:

We have $0 \le x \le 1 \Rightarrow 0 \le \frac{u}{v} \le 1$

i.e $0 \le u \le v$

Also $0 \le y \le 1 \Rightarrow 0 \le v \le 1$

On combining the two limits, we get $0 \le u \le v \le 1$

$$\therefore f_{uv}(u,v) = \frac{u}{v^2} + 1,\ 0 \le u \le v \le 1$$

PDF of U is given by

$$f_U(u) = \int_{v=u}^{v=1} f_{uv}(u,v)dv \qquad 0 \le u \le v < 1$$

$$= \int_u^1 (\frac{u}{v^2} + 1)\, dv$$

$$= \int_u^1 (uv^{-2} + 1)dv$$

$$= [\frac{uv^{-1}}{-1} + v]_u^1$$

$= (\frac{u}{-1} + 1) + 1 - u$

$= -u + 1 + 1 - u$

$= 2 - 2u$

$f_U(u) = 2(1 - u)\ 0 < u < 1$

2. Let (X, Y) be a continuous two dimensional randow. with JPDF $f(x, y) = 4xye^{-(x^2+y^2)} x > 0, y > 0$. Find the PDF of $\sqrt{X^2 + Y^2}$

Solution:

Given (X, Y) is a continuous two dimensional random variables defined in $0 < x < \infty$ and

$0 < y < \infty$

Given $f(x, y) = 4xye^{-(x^2+y^2)},\ 0 < x < \infty, 0 < y < \infty$

let $u = \sqrt{x^2 + y^2} \ldots\ldots (1)$ Take $v = y \Rightarrow y = v$

$(1) \Rightarrow u^2 = x^2 + y^2$

$u^2 = x^2 + y^2 \qquad y = v$

$x^2 = u^2 - v^2 \Rightarrow \text{x}=\sqrt{u^2 - v^2}$

$x\sqrt{u^2 - v^2}$, $y = v$

$$\frac{\partial x}{\partial u} = \frac{1}{2}\frac{1}{\sqrt{u^2 - v^2}}(2u) = \frac{u}{\sqrt{u^2 - v^2}}; \frac{\partial y}{\partial u} = 0$$

$$\frac{\partial x}{\partial v} = \frac{1}{2}\frac{1}{\sqrt{u^2 - v^2}}(-2v) = \frac{-v}{\sqrt{u^2 - v^2}}; \frac{\partial y}{\partial v} = 1 = 1$$

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

$$J = \frac{u}{\sqrt{u^2-v^2}}$$

PDF of (U, V) is $f_{UV}(u, v) = |J| f_{XY}(x, y)$

$$= \frac{u}{\sqrt{u^2-v^2}} 4xye^{-(x^2+y^2)}$$

$$= \frac{u}{\sqrt{u^2-v^2}} 4\sqrt{u^2 - v^2}(v)e^{-u^2}$$

$$f_{UV}(u, v) = 4uve^{-u^2}$$

To find the range for u and v:

We have $x > 0$ We have $y > 0$

$\sqrt{u^2 - v^2} > 0$ $v > 0$

$u^2 - v^2 > 0 \qquad \Rightarrow \; 0 < v < \infty$

$u^2 > v^2 \Rightarrow u > v$

$\Rightarrow v < u$

On combining the two limits, we get $0 < v < u < \infty$

$$f_{UV}(u, v) = 4uve^{-u^2}, 0 < v < u < \infty$$

PDF of U is given by

$$\mathrm{f}_U(u) = \int_{v=0}^{v=u} f_{uv}(u, v)dv$$

$$= \int_0^u 4uve^{-u^2}dv$$

$$= 4ue^{-u^2}\int_0^u vdv$$

$$= 4ue^{-u^2}\left[\frac{v^2}{2}\right]_0^u$$

$$= 2u^3e^{-u^2} 0 < u < \infty$$

3. The JPDF to two dimensional random variables X and Y is given by,

$(x, y) = e^{-(x+y)}, x > 0, y > 0$. Find the PDF of $\frac{X+Y}{2}$

Solution:

Given (X, Y) is a continuous two dimensional random variable defined in

$0 < x < \infty$ and

$0 < y < \infty$. Also given $f(x, y) = e^{-(x+y)}; 0 < x < \infty, 0 < y < \infty$

let u $= \frac{x+y}{2}$ ... ... (1). Take $v = y \Rightarrow y = v$

$(1) \Rightarrow u = \frac{1}{2}(x + v)$

$2u = x + vx = 2u - v$

$\therefore x = 2u - v;$

$y = v \frac{\partial x}{d_u} = 2 \frac{\partial x}{\partial v} = -1; \frac{\partial y}{\partial u} = 0; \frac{\partial y}{\partial v} = 1$

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

the PDF of (U, V) is $f_{uv}(u, v) = |J| f_{XY}(x, y)$

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

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

$$= 2e^{-2u}$$

To find range for u and v :

We have $x > 0 \Rightarrow 2u - v > 0$

i. e., $2u > v \Rightarrow v < 2u$

Also $y > 0 \Rightarrow v > 0$

$\therefore v < 2u; v > 0$ $\qquad 0 < v < 2u < \infty$

On combining the two limits, we get $0 < v < 2u < \infty$

$\therefore f_{UV}(u, v) = 2e^{-2u}, 0 < v < 2u < \infty$

The PDF of *U* is

$$f_U(u) = \int_{v=0}^{v=2u} f_{uv}(u, v)dv$$

$$= \int_0^{2u} 2e^{-2u}dv$$

$$= 2e^{-2u} \int_0^{2u} dv$$

$$= 2e^{-2u}[v]_0^{2u}$$

$$= 2e^{-2u}(2u)$$

$$f_u(u) = 4ue^{-2u}; u > 0$$

UNIT STEP FUNCTION:

$$u(x) = 1 \text{ for } x > 0$$

$$u(x) = 0 \text{ for } x < 0$$

1. If X and Y are two independent random variables each normally distributed with mean $= 0$ and variance σ^2 , find the density function of $R = \sqrt{X^2 + Y^2}$ and $\phi = \tan^{-1}\left(\frac{Y}{X}\right)$

Solution:

Given that X follows $N(0, \sigma)$.

$$\therefore f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2\sigma^2}x^2}; -\infty < x < \infty$$

Also Y follows $N(0, \sigma)$.

$$\therefore f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2\sigma^2}y^2}; -\infty < y < \infty$$

Since X and Y are independent, $f_{XY}(x, y) = f_X(x) f_Y(y)$

$$= \frac{1}{\sigma^2 2\pi} e^{\frac{-1}{2\sigma^2}(x^2+y^2)}; -\infty < x < \infty, -\infty < y < \infty$$

We have $r = \sqrt{x^2 + y^2}; \theta = \tan^{-1}\left(\frac{y}{x}\right)$

$\Rightarrow x = r\cos\theta, y = r\sin\theta,$

$\Rightarrow J = r$

JPDF of (R, ϕ) is $f_{R\phi}(r, \theta) = |J|\, f_{XY}(x, y)$

$$= r\frac{1}{\sigma^2 2\pi} e^{\frac{-1}{2\sigma^2}(x^2+y^2)}$$

$$= \frac{r}{\sigma^2 2\pi} e^{\frac{-1}{2\sigma^2}r^2}$$

To find the range for r and θ :

We have $-\infty < x < \infty, -\infty < y < \infty$ t.e entire XY plane.

The entire XY plane is transformed into $x = \text{r}\cos\theta$, $y = \text{r}\sin\theta$

i.e the entire XY plane is transformed into $x^2 + y^2 = r^2$ (a circle of infinite radius)

Whole region is transformed into a circle of infinite radius.

$$\therefore 0 \leq r < \infty, 0 \leq \theta \leq 2\pi$$

$$\therefore f_{R\phi}(r,\theta) = \frac{r}{\sigma^2 2\pi} e^{\frac{-1}{2\sigma^2}r^2} \quad 0 \leq r < \infty, 0 \leq \theta \leq 2\pi$$

The PDF of R is

$$f_R(r) = \int_{r=0}^{\infty} f_{r\theta}(r,\theta) d\theta$$

$$= \int_0^{in} \frac{r}{\sigma^2 2\pi} e^{\frac{-1}{2\sigma^2}r^2} d\theta$$

$$= \frac{r}{\sigma^2 2\pi} e^{\frac{-1}{2\sigma^2}r^2} \int_0^{\infty} d\theta$$

$$= \frac{r}{\sigma^2 2\pi} e^{\frac{-1}{2\sigma^2} r^2} [\theta]_0^{2\pi}$$

$$f_R(r) = \frac{r}{\sigma^2} e^{\frac{-1}{2\sigma^2} r^2}; 0 \leq r < \infty$$

The PDF of ϕ is

$$f_\phi(\theta) = \int_{r=0}^{\infty} f_{r\theta}(r, \theta) dr$$

$$= \int_0^{\infty} \frac{r}{\sigma^2 2\pi} e^{\frac{-1}{2\sigma^2} r^2} dr$$

$$= \frac{1}{\sigma^2 2\pi} \int_0^{\infty} r e^{\frac{-1}{2\sigma^2} r^2} dr$$

Put $\frac{1}{2\sigma^2} r^2 = t$

$$\frac{1}{2\sigma^2} 2rdr = dt$$

$$rdr = \sigma^2 dt$$

There is no change on the limits

$$f_0(\theta) = \frac{1}{\sigma^2 2\pi} \int_0^{\infty} e^{-t} \sigma^2 dt$$

$$= \frac{1}{2\pi} \left[\frac{e^{-t}}{-1}\right]_0^{\infty}$$

$= \frac{1}{2\pi}(0+1)$

$f_\phi(\theta) = \frac{1}{2\pi} 0 \le \theta \le 2\pi$

2. The random variables X and Y each follows an exponert distribution with parameter 1 and are independent. Find the PDF of $U = X - 1$

Solution:

Given X and Y follows exponential distribution with parameter with $\lambda = 1$

$\therefore f_x(x) = \lambda e^{-\lambda x}; x > 0$

$= e^{-x}$

$f_y(y) = e^{-y}; y > 0$

Since X and y are independent,

$f_{XY}(x, y) = f_x(x) f_y(y)$

$= e^{-x} e^{-y}$

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

let $u = x - y$ ……………..(1) Take $v = y \Rightarrow y = v$

$(1) \Rightarrow u = x \; v \Rightarrow x = u + v$

$x = u + v \; ; y = v$

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

The JPDF of (U, V) is $f_{uv}(u, v) = |J| f_{XY}(x, y)$

$$= (1)e^{-(x+y)} = e^{-(u+v+v)}$$

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

To find the range for u and v :

We have $x > 0 \Rightarrow u + v > 0 \Rightarrow u > -v$

fie $y > 0 \Rightarrow v > 0$

$\therefore f_{uv}(u, v) = e^{-(u+2v)} u > -v, v > 0$

The PDF of U is

$f_u(u) = \int f(u, v) dv$

Since there are two slopes, the region is divided into two sub regions R_1 and R_2

ln R_1:

At $P_1, v = -u$; At Q_1, v

ln R_2 :

At $P_2, v = 0$; At $Q_2, v = \infty$

In R_1 :

$$f_U(u) = \int_{v=-4}^{\infty} f(u,v)dv$$

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

$$= \int_{-u}^{\infty} e^{-u}e^{-2v}dv$$

$$= e^{-u}\int_{-u}^{\infty} e^{-2v}dv$$

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

$$= e^{-u}\left[0 - \frac{e^{2u}}{-2}\right]$$

$$= \frac{e^{u}}{2}; u < 0$$

In R_2

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

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

$$= \int_{0}^{\infty} e^{-u}e^{-2v}dv$$

$= \int_0^{\infty} e^{-u} \left[0 - \frac{1}{-2}\right]$

$= \frac{e^{-u}}{2}; u > 0$

$= e^{-u} \int_0^{\infty} e^{-2v}\, dv$

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

$$f_U(u) = \begin{cases} \frac{e^{u}}{2} & u < 0 \\ \frac{e^{-u}}{2} & u > 0 \end{cases}$$

2.5 Central limit theorem:

Statement:

Let $x_1, x_2, \ldots, x_n$ are n independent identically distributed random variables with same mean μ and standard deviation σ and if $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, then the variate $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ has a distribution that approaches the standard normal distribution an $n \to \infty$ provided the MGF of x_i exist.

Proof:

MGF of z about origin is $M_X(t) = E(e^{tz})$

$$= E\left[e^{t\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)}\right]$$

$$= E\left[e^{\frac{\sqrt{n}t}{\sigma}(\bar{X}-\mu)}\right]$$

$$= E\left[e^{\frac{\bar{X}\sqrt{n}t}{\sigma}} e^{-\frac{\mu\sqrt{n}t}{\sigma}}\right]$$

$$= e^{-\frac{\mu\sqrt{n}t}{\sigma}} E\left[e^{\frac{\bar{X}\sqrt{n}t}{\sigma}}\right]$$

$$= e^{-\frac{\mu\sqrt{n}t}{\sigma}} E\left[e^{\frac{\sqrt{n}t}{\sigma}\frac{1}{n}(x_1 + x_2 + \ldots + x_n)}\right]$$

$$= e^{-\frac{\mu\sqrt{n}t}{\sigma}} E\left(e^{\frac{tx_1}{\sigma\sqrt{n}}}\right) E\left(e^{\frac{tx_2}{\sigma\sqrt{n}}}\right) \ldots E\left(e^{\frac{tx_n}{\sigma\sqrt{n}}}\right)$$

$$= e^{-\frac{\mu\sqrt{n}t}{\sigma}}\{M_X(\frac{t}{\sigma\sqrt{n}})\}^n$$

Taking log on both sides

$$logM_z(t) = loge^{-\frac{\mu\sqrt{n}t}{\sigma}} + log\{M_X(\frac{t}{\sigma\sqrt{n}})\}^n$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + nlogM_X(\frac{t}{\sigma\sqrt{n}})$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + nlogE(e^{\frac{tX}{\sigma\sqrt{n}}})$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + nlog[E(1 + \frac{\frac{tx}{\sigma\sqrt{n}}}{1!} + \frac{(\frac{tx}{\sigma\sqrt{n}})^2}{2!} + \ldots)] \quad \mu$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + nlog[E(1 + \frac{tx}{\sigma\sqrt{n}} + \frac{1}{2!}\frac{t^2x^2}{\sigma^2 n} + \ldots)]$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + nlog[1 + \frac{tx}{\sigma\sqrt{n}}E(x) + \frac{1}{2!}\frac{t^2x^2}{\sigma^2 n}E(x^2) + \ldots]$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + nlog[1 + \frac{tx}{\sigma\sqrt{n}}\mu_1' + \frac{1}{2!}\frac{t^2x^2}{\sigma^2 n}\mu_2' + \ldots]$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + n[(\frac{tx}{\sigma\sqrt{n}}\mu_1' + \frac{1}{2!}\frac{t^2x^2}{\sigma^2 n}\mu_2' + \ldots) - \frac{1}{2}(\frac{tx}{\sigma\sqrt{n}}\mu_1' + \frac{1}{2!}\frac{t^2x^2}{\sigma^2 n}\mu_2' + \ldots)^2 + \ldots]$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + \frac{\mu_1' t\sqrt{n}}{\sigma} + \frac{\mu_2' t^2}{2!\sigma} + \ldots - \frac{(\mu_1')^2 t^2}{2\sigma^2} + \text{terms containing "}n\text{" in the denominator}$$

Put $\mu = \mu_1'$

$$= \frac{-\mu_1' t\sqrt{n}}{\sigma} + \frac{\mu_1' t\sqrt{n}}{\sigma} + \frac{\mu_2' t^2}{2!\sigma} + \ldots - \frac{(\mu_1')^2 t^2}{2\sigma^2} + \text{ terms containing "} n \text{" in the denominator}$$

$$= \frac{t^2}{2\sigma^2}(\mu_2' - (\mu_1')^2) + \text{ terms containing "} n \text{" in the denominator}$$

$$= \frac{t^2}{2\sigma^2}\sigma^2 + \text{ terms containing "} n \text{" in the denominator}$$

$$log M_z(t) = \frac{t^2}{2} + \text{ terms containing "} n \text{" in the denominator}$$

Letting $n \to \infty$

$$log M_z(t) = \frac{t^2}{2}$$

$$\Rightarrow M_z(t) = e^{\frac{t^2}{2}} = \text{MGF of } N(0,1)$$

Hence z follows standard normal distribution as $n \to \infty$

Standard Normal Distribution:

Let $z = \frac{X-\mu}{\sigma}$, z follows normal distribution with mean 0 and variance 1, then z follows standard normal distribution.

Problems on Central limit theorem:

1. If $X_1, X_2, ..., X_n$ are Poisson variables with parameter $\lambda = 2$, use central limit theorem to estimate $P(120 < S_n < 160)$ where $S_n = X_1 + X_2 + ... + X_n$ and $n = 75$

Solution:

To find mean and variance

Given mean = 2

Variance = 2

(For Poisson distribution Mean = variance = λ

To find $n\mu$ and $n\sigma^2$

$n\mu = 75 \times 2 = 150$

$n\sigma^2 = 75 \times 2 = 150$

$\sigma\sqrt{n} = \sqrt{150}$

Application of central limit theorem

$S_n \sim N(n\mu, \sigma\sqrt{n})$

$\sim N(150, \sqrt{150})$

To find $P(120 < S_n < 160)$

Let $z = \frac{S_n - n\mu}{\sigma\sqrt{n}}$

$$= \frac{S_n - 150}{\sqrt{150}}$$

If $S_n = 120$

$$z = \frac{120-150}{\sqrt{150}} = -2.45$$

If $S_n = 160$

$$z = \frac{160-150}{\sqrt{150}} = 0.85$$

$$P(120 < S_n < 160) = P\left(\frac{S_n-150}{\sqrt{150}} \le z \le \frac{S_n+150}{\sqrt{150}}\right)$$

$$= P(-2.45 \le z \le 0.85)$$

$$= P(-2.45 \le z \le 0) + P(0 \le z \le 0.85)$$

$$= 0.4927 + 0.2939 = 0.7866$$

2. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variable variables with mean = 2 and variance = ¼ . Find $P(192 < X_1 + X_2 + \ldots + X_n < 210)$

Solution:

To find mean and variance

Given mean = 2

Variance = ¼ , $n = 4$

To find $n\mu$ and $n\sigma^2$

$n\mu = 100 \times 2 = 200$

$n\sigma^2 = 100 \times 1/4 = 25$

$\sigma\sqrt{n} = 5$

Application of central limit theorem

$S_n \sim N(n\mu, \sigma\sqrt{n})$

$\sim N(200, 5)$

To find $P(192 < S_n < 210)$

Let $z = \frac{S_n - n\mu}{\sigma\sqrt{n}}$

$= \frac{S_n - 200}{5}$

If $S_n = 192$

$z = \frac{192-200}{5} = -1.6$

If $S_n = 210$

$$z = \frac{210-200}{5} = 2$$

$$P(192 < S_n < 210) = P\left(\frac{S_n-200}{5} \leq z \leq \frac{S_n+200}{5}\right)$$

$$= P(-1.6 \leq z \leq 2)$$

$$= P(-1.6 \leq z \leq 0) + P(0 \leq z \leq 2)$$

$$= 0.4452 + 0.4772 = 0.9224$$

3. The resistors r_1, r_2, r_3 and r_4 are independent random variables and is uniform in the interval $(450, 550)$. Using the central limit theorem, find $P(1900 < r_1 + r_2 + r_3 + r_4 < 2100)$

Solution:

To find mean and variance

A random variable X is said to have uniform distribution on the interval (a, b) if its probability density function is given by

$$f(x) = \frac{1}{b-a}, a < x < b$$

Mean $= \frac{a+b}{2}$, Variance $= \frac{(b-a)^2}{12}$

Mean $= \frac{450+550}{2} = 500$

Variance $= \frac{(550-450)^2}{12} = 833.33, n = 4$

To find $n\mu$ and $n\sigma^2$

$n\mu = 4 \times 500 = 2000$

$n\sigma^2 = 4 \times 833.33 = 25$

$\sigma\sqrt{n} = 2\sqrt{833.33} = 57.73$

Application of central limit theorem

$S_n \sim N(n\mu, \sigma\sqrt{n})$

$\sim N(200, 57.73)$

To find $P(1900 < S_n < 2100)$

Let $z = \frac{S_n - n\mu}{\sigma\sqrt{n}}$

$= \frac{S_n - 2000}{57.73}$

If $S_n = 1900$

$$z = \frac{1900-2000}{57.73} = -1.73$$

If $S_n = 2100$

$$z = \frac{2100-2000}{57.73} = 1.73$$

$$P(1900 < S_n < 2100) = P\left(\frac{S_n - 2000}{57.73} \le z \le \frac{S_n + 2000}{57.73}\right)$$

$$= P(-1.73 \le z \le 1.73)$$

$$= P(-1.73 \le z \le 0) + P(0 \le z \le 1.73)$$

$$= 2 \times P(0 \le z \le 1.73)$$

$$= 2 \times 0.4582 = 0.9164$$

4. If $x_i, i = 1, 2, \dots, 50$ are independent random variables each having a Poisson distribution with parameter $\lambda = 0.03$ and $S_n = X_1 + X_2 + \dots + X_n$ evaluate $P(S_n \ge 3)$

Solution:

To find mean and variance

Given mean = 0.03

Variance = 0.03 , $n = 4$

To find $n\mu$ and $n\sigma^2$

$n\mu = 50 \times 0.03 = 1.5$

$n\sigma^2 = 50 \times 0.03 = 1.5$

$\sigma\sqrt{n} = \sqrt{1.5}$

Application of central limit theorem

$S_n \sim N(n\mu, \sigma\sqrt{n})$

$\sim N(1.5, \sqrt{1.5})$

To find $P(S_n \geq 3)$

Let $z = \frac{S_n - n\mu}{\sigma\sqrt{n}}$

$= \frac{S_n - 1.5}{\sqrt{1.5}}$

If $S_n = 3$

$z = \frac{3-1.5}{\sqrt{1.5}} = \sqrt{1.5}$

$P(S_n \geq 3) = P(z \geq \sqrt{1.5})$

$= P(z \geq 1.23)$

$= 0.5 - P(z < 1.23)$

$= 0.1112$

5. A coin is tossed 300 times. What is the probability that heads will appear more than 140 times and less than 150 times.

Solution:

To find mean and variance

Let p be the probability of getting head in a single trial.

$p = \frac{1}{2}, q = 1 - \frac{1}{2} = \frac{1}{2}$

Here n = 300

To find np and npq

mean $= np = 300 \times \frac{1}{2} = 150$

Variance $= npp = 300 \times \frac{1}{2} \times \frac{1}{2} = 75$

To find $P(140 < S_n < 150)$

Let $z = \frac{X-\mu}{\sigma}$

$= \frac{X-150}{\sqrt{75}}$

If $X = 140$

$z = \frac{140-150}{\sqrt{75}} = -1.15$

If $X = 150$

$z = \frac{150-150}{\sqrt{75}} = 0$

$P(140 < X < 50) = P\left(\frac{X-150}{\sqrt{75}} \leq z \leq \frac{X+150}{\sqrt{75}}\right)$

$= P(-1.15 \leq z \leq 0)$

$= P(0 \leq z \leq 1.15)$

$= 0.3749$