Module - 2

Curve Fitting

Let $(x_1, y_1), (x_2, y_2) - - (x_n, y_n)$ be n pairs of observations of related data. The general problem of finding a relation of the form $y = f(x)$ which fits best to the given data is called curve fitting.

Principle of Least squares:

Let the given data be $(x_1, y_1), (x_2, y_2), - - (x_n, y_n)$ ~~be a best fit~~ and $y = f(x)$ be a best fit curve. y_i is called the observed value of y corresponding to x_i and $f(x_i)$ is called the expected value of y corresponding to x_i.

E_i is called error or residual for y_i.

The principle of least squares states that, for a best fit curve, the sum of the squares of the residuals is a minimum, ie

$$E = \sum_{i=1}^{n} [y_i - f(x_i)]^2$$

Fitting a straight Line.

Let $Y = a + bx$ be a best fit str. line to the given data (x_i, y_i), $i = 1, 2, \dots n$

Then $E = \sum_{i=1}^{n} \{y_i - F(x_i)\}^2$

$= \sum_{i=1}^{n} [y_i - a - bx_i]^2$ and

it is a function of a and b

For E to be minimum,

$\frac{\partial E}{\partial a} = 0$ and $\frac{\partial E}{\partial b} = 0$.

$\frac{\partial E}{\partial a} = 0 \Rightarrow \sum_{i=1}^{n} 2(y_i - a - bx_i)(-1) = 0$

Hence $\sum y_i = na + b\sum x_i$ — ①

Also $\frac{\partial E}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} 2(y_i - a - bx_i)(-x_i)$.

Hence $\sum x_i y_i = a\sum x_i + b\sum x_i^2$ — ②

Equations ① and ② are called normal equations. Solving these 2 normal equations, we can find a and b.

Fitting a parabola of the form $y = a + bx + cx^2$.

Here $E = \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2)$

which is a function of a, b and c.

for E to be minimum, $\frac{\partial E}{\partial a} = 0$, and $\frac{\partial E}{\partial b} = 0$ and $\frac{\partial E}{\partial c} = 0$.

Now, $\frac{\partial E}{\partial a} = \sum 2(y_i - a - bx_i - cx_i^2)(-1) = 0$

ie $(\sum y_i = na - b\sum x_i - c\sum x_i^2)(-1)$

$\Rightarrow \sum y_i = na + b\sum x_i + c\sum x_i^2$ —— ①.

Similarly,

$\frac{\partial E}{\partial b} = \sum 2(y_i - a - bx_i - cx_i^2)(-x_i) = 0$

$\Rightarrow \sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3$ —— ②.

Similarly,

$\frac{\partial E}{\partial c} = \sum 2(y_i - a - bx_i - cx_i^2)(-x_i^2)$

$\Rightarrow \sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4$

—— ③

①, ② and ③ are normal eqtns and can be solved to find a, b and c.

Note:-

(1) Fitting a parabola of the form $y=a+bx^2$, we find $X=x_i^2$ and then $y=a+bX$.

$\therefore$ the normal equations are

$\Sigma y_i = na + b\Sigma X_i$ and

$\Sigma X_i y_i = a\Sigma X_i + b\Sigma X_i^2$

ie $\Sigma y_i = na + b\Sigma x_i^2$ and

$\Sigma x_i^2 y_i = a\Sigma x_i^2 + b\Sigma x_i^4$.

(2) If we want to fit a curve of the form $y=ae^{bx}$, we apply logarithm ie

$$\log y = \log ae^{bx}$$
$$= \log a + \log e^{bx}$$
$$= \log a + bx\log e. \quad (\text{since } \log e = 1)$$

$\Rightarrow \log y = \log a + bx$

Let $Y=\log y$ and $A=\log a$.

Then the above equation becomes,

$Y = A + bx$.

The normal eqtn's are,

$$\Sigma Y = nA + b\Sigma x$$

$$\Sigma xY = A\Sigma x + b\Sigma x^2$$

On solving, we get the values of A and b. $\therefore$ $\underline{a = e^A}$.

(3) If the values of x and y are large, then the calculation of Σx, Σx^2, Σx^3, Σxy etc becomes tedious. They canbe simplified by a suitable change of origin $X = x - h$ and $Y = y - k$. If we take $h = \bar{x}$ and $k = \bar{y}$, then $\Sigma X = 0$, ~~and~~ $\Sigma Y = 0$, ~~and~~ $\Sigma X^3 = 0$ etc and hence the solutions become easier.

<u>Qtns</u>

1. By the method of least squares, find the best fitting straight line to the data given below

x	5	10	15	20	25.
y	15	19	23	26	30.

?.

Let $y = a + bx$.

The normal equations are

$$\Sigma y = na + b\Sigma x$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

x	y	x^2	xy
5	15	25	75
10	19	100	190
15	23	225	345
20	26	400	520
25	30	625	750
75	113	1375	1880

$\Sigma x = 75$

$\Sigma y = 113$

$\Sigma x^2 = 1375$

$\Sigma xy = 1880$

The normal equations then becomes

$$113 = 5a + b(75)$$

$$1880 = a(75) + 1375b.$$

$\therefore \quad a = 11.5, \; b = 0.74.$

Hence the str8. line fitting the given data is $y = 11.5 + 0.74x$.

2 Fit a strt. line of the form $y = ax + b$ for (4)

x	1	2	3	4	5
y	5	7	9	10	11

The required normal equations are

$$\Sigma y = a \Sigma x + nb$$
$$\Sigma xy = a \Sigma x^2 + b \Sigma x$$

x	y	x^2	xy
1	5	1	5
2	7	4	14
3	9	9	27
4	10	16	40
5	11	25	55

$\Sigma x = 15$.
$\Sigma y = 42$.
$\Sigma x^2 = 55$.
$\Sigma xy = 141$

Hence the normal equations changes to,

$$42 = a(15) + 5b$$
$$141 = a(55) + 15b.$$

$\therefore\ a = 1.5$ and $b = 3.9$.

Hence the strt. line fitting the given data is $\underline{\underline{y = 1.5x + 3.9}}$

3. Convert the equation $y=ax+bx^2$ to a linear form, the corresponding normal equations to fit it?

$y=ax+bx^2$

Divide throughout with x.

$\frac{y}{x}=a+bx$ — ①

Take $\frac{y}{x}$ as Y. Then ① becomes,

$Y=a+bx$ which is linear.

The corresponding normal eqtns are $\Sigma Y = na+b\Sigma x$

$\Sigma xY = a\Sigma x + b\Sigma x^2$

4. Use the method of least squares to determine the formula $y=ax+bx^2$ for the following data

x	1	2	3	4	5
y	1.8	5.1	8.9	14.1	19.8

$$y = ax + bx^2$$

Divide throughout with 'x'.

$$\frac{y}{x} = a + bx \quad . \text{ Let } \quad Y = \frac{y}{x}.$$

Then $Y = a + bx$ — ①.

The corresponding normal equations are

$$\sum Y = na + b\sum x$$

$$\sum xY = a\sum x + b\sum x^2.$$

xY	x	y	Y	x^2
1.8	1	1.8	1.8	1
5.1	2	5.1	2.55	4
8.9	3	8.9	2.967	9
14.1	4	14.1	3.525	16
19.8	5.	19.8	3.96	25

$\sum x = 15$

$\sum xY = 49.7$

$\sum x^2 = 55$

$\sum Y = 14.802$

Hence the normal equations changes to

$$14.802 = 5a + b(15)$$

$$49.7 = a(15) + b(55)$$

On solving, we gets,

$$a = 1.3722, \quad b = 0.5294.$$

Hence the formula becomes.

$$y = 1.3722x + 0.5294x^2$$

(5) Convert the eqtn. $y = \frac{x}{a+bx}$ to a linear form and hence find the normal equations ?

$$y = \frac{x}{a+bx} \Rightarrow \frac{a+bx}{x} = \frac{1}{y}$$

$$\Rightarrow \frac{a}{x} + b = \frac{1}{y} \quad \text{--- (1)}$$

Let $\frac{1}{y} = Y$ and $\frac{1}{x} = X$; then (1) becomes, $aX + b = Y$.

The corresponding normal equations are

$$\sum Y = a\sum X + nb$$

$$\sum XY = a\sum X^2 + b\sum X$$

(6) Fit $y = ae^{bx}$ to the following data

x	0	2	4
y	5.012	10	31.62

?

$$y = ae^{bx}$$

$$\log_{10} y = \log_{10} a + bx \log_{10} e \quad \text{--- (1)}$$

Put $Y = \log y$, $\log_{10} a = A$

and $b \log_{10} e = B$.

Then (1) becomes, $Y = A + Bx$.

The corresponding normal equations are

$$\Sigma Y = nA + B\Sigma x \text{ and}$$

$$\Sigma xY = A\Sigma x + B\Sigma x^2.$$

x	y	$Y = \log y$	xY	x^2
0	5.012	0.7	0	0
2	10	1	2	4
4	31.62	1.499	5.999	16

$\Sigma x = 6$, $\Sigma xY = 7.999$.

$\Sigma x^2 = 20$, $\Sigma Y = 3.199$.

Then, the normal equations changes to

$$3.199 = 3A + 6B$$

$$7.999 = 6A + 20B.$$

$\Rightarrow A = 0.6661$, $B = 0.2$

$A = \log a = 0.6661$.

$\Rightarrow a = e^{0.6661} = 1.9466$

$B = b \log e = 0.2$.

$\Rightarrow b = \underline{0.461}$

1. Fit a second degree curve of the form $y = ax^2 + bx + c$ to the following data

x	1911	1912	1913	1914	1915
y	10	12	8	10	14

?

The calculations can be made simple

Take $X = x - 1913$

$y = aX^2 + bX + c.$

$\Sigma y = a\Sigma X^2 + b\Sigma X + nc.$

$\Sigma Xy = a\Sigma X^3 + b\Sigma X^2 + c\Sigma X.$

$\Sigma X^2 y = a\Sigma X^4 + b\Sigma X^3 + c\Sigma X^2.$

Xy.	x	y	$X = x - 1913$	X^2	X^3	X^4	X^2y
-20	1911	10	-2	4	-8	16	40
-12	1912	12	-1	1	-1	1	12
0	1913.	8	0	0	0	0	0
10	1914	10	1	1	1	1	10.
28	1915	14	2.	4	8	16.	56.

$\Sigma y = 54,\ \Sigma X = 0,\ \Sigma X^2 = 10,$

$\Sigma X^3 = 0,\ \Sigma X^4 = 34,$

$\Sigma Xy = 6,\ \Sigma X^2 y = 118.$

$54 = 10a + 0b + 5c.$

$6 = 0a + 10b + 0c$

$118 = 34a + 0b + 10c.$

$a = 0.7143,\ b = 0.6,\ c = 9.37143$

$y = 0.7143X^2 + 0.6X + 9.37143.$

$= 0.7143(x-1913)^2 + 0.6(x-1913) + 9.37143.$

$= 0.7143x^2 + 2614030.137 - 2732.9118x + 0.6x - 1147.8 + 9.37143$

$= 0.7143x^2 - 2732.3118x + 2612891.708$

2. Fit a second degree parabola to the following data.

x	1989	1990	1991	1992	1993
y	352	356	357	358	360

x	1994	1995	1996	1997
y	361	361	360	359.

?

$X = x - h = x - \bar{x} = x - 1993.$

$Y = y - k = y - \bar{y} = y - 358$

x	y	$X = x - 1993$	$Y = y - 358$	XY	X^2
1989	352.	-4	-6	24	16
1990	356	-3	-2	6	9
1991	357	-2.	-1	2	4
1992	358	-1	0	0	1
1993	360	0	2	0	0
1994	361	1	3	3	1
1995	361	2	3	6	4
1996	360	3	2	6	9
1997	359	4	1	4	16.

$\Sigma X = 0, \ \Sigma Y = 2,$

$\Sigma XY = 51, \ \Sigma X^2 = 60.$

$\Sigma X^2 Y = -69, \ \Sigma X^3 = 0, \ \Sigma X^4 = 708.$

$\Sigma x = 17937, \ \Sigma y = 3224$

x^2y	x^3	x^4
-96	-64	256
-18	-27	81
-4	-8	16
0	-1	1
0	0	0
3	1	1
12	8	16
18	27	81
16	64	256.

$y = a + bx + x^2 c$

$\Sigma y = na + b\Sigma x + c\Sigma x^2$

$\Sigma xy = \Sigma x \cdot a + b\Sigma x^2 + c\Sigma x^3$

$\Sigma x^2 y = \Sigma x^2 a + b\Sigma x^3 + c\Sigma x^4$.

$2 = 9a + b(0) + c(60)$

$51 = 0a + 60b + 0c$

$-69 = 60a + 708c + 0b$.

$\Rightarrow a = 2.004,\ b = 0.85,\ c = -0.2673$

Hence the required parabola is

$y - 358 = 2.004 + 0.85(x - 1993) + -0.2673(x - 1993)^2$.

$y = 358 + 2.004 + 0.85x - 169405 + -0.2673x^2 - 1061728.698 + 1065.4578x$

$$\Rightarrow y = -0.2673x^2 + 1066.3078x - 1063062.744$$

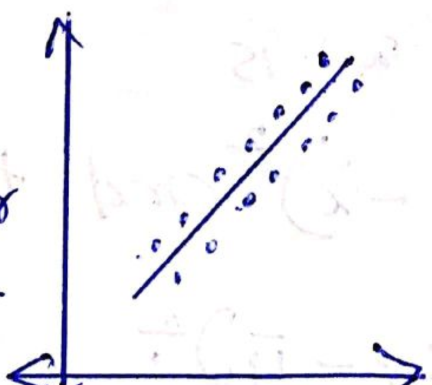
Correlation and Regression.

The word correlation means the study of relation b/w. the variables X and Y.

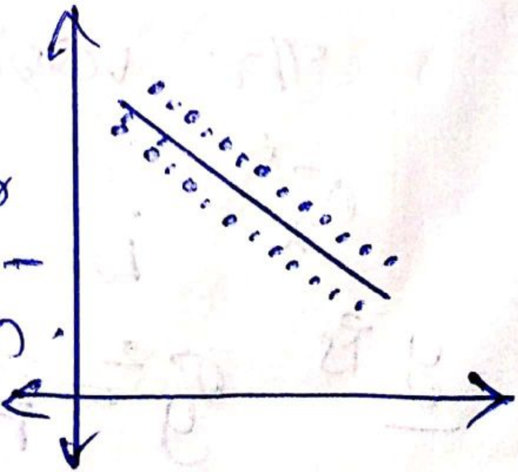
Scatter diagram.

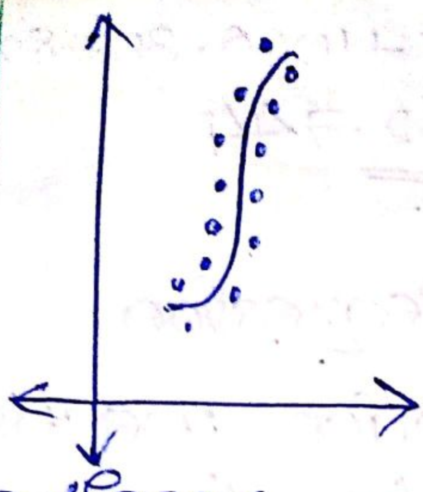
Let (x_i, y_j), $i \in \mathbb{N}$, be a distribution. Let the values of the variables x & y be plotted along the x-axis and y-axis. Then, corresponding to every ordered pair, there exists a point on the xy plane. The diagram showing all those points so obtained, in the form of dots in the xy plane is called dot or scatter diagram.

+ve linear correlation

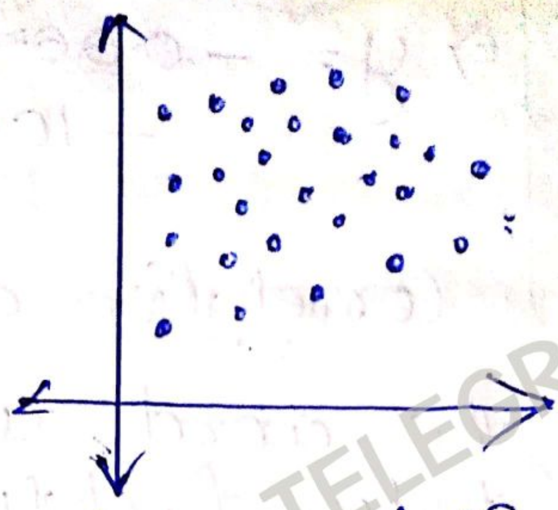


-ve linear correlation.





Non linear correlation.



No correlation.

Coefficient of Correlation

Karl pearson's coefficient of correlation is denoted by r or r_{xy}.

Let (x_i, y_i), $i \in N$, be a set of n observations. Let $\bar{x}$ and $\bar{y}$ be the means of x_i and y_i respectively.

ie ~~then~~ $\bar{x} = \frac{\Sigma x_i}{n}$, $\bar{y} = \frac{\Sigma y_i}{n}$.

The variance of x is

$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ and that of

y is $\sigma_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$.

The covariance of x and y is

$$\text{cov}(x,y) = \sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

Then $$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

ie $$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^2}\cdot\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y-\bar{y})^2}}$$

$$= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$= \frac{\sum_{i=1}^{n} xy - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}\,\sqrt{\sum_{i=1}^{n} y^2 - n\bar{y}^2}}$$

$$= \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\cdot\sqrt{n\sum y^2 - (\sum y)^2}}$$

$$r_{uv} = \frac{n\sum uv - \sum u \sum v}{\sqrt{n\sum u^2 - (\sum u)^2} \cdot \sqrt{n\sum v^2 - \cancel{n}(\sum v)^2}}$$

where $u = x - \bar{x}$ &

$v = y - \bar{y}$.

<u>Note</u>: If X and Y are independent, then ~~E(XY) = E(X)~~

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

$\therefore \; \underline{cov(X, Y) = 0}$.

<u>Questions</u>

1. Calculate the correlation coefficient from the following data.

x	78	89	97	69	59	79	68	57
y	125	137	156	112	107	138	123	108

x	y	$x \cdot y$	x^2	y^2
78	125	9750	6084	15625
89	137	12193	7921	18769
97	156	15132	9409	24336
69	112	7728	4761	12544
59	107	6313	3481	11449
79	138	10902	6241	19044
68	123	8364	4624	15129
57	108	6156	3249	11664
596	1006	76538	45770	128560

$$r_{xy} = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{8 \times 76538 - 596 \times 1006}{\sqrt{8 \times 45770 - (596)^2} \cdot \sqrt{8 \times 128560 - (1006)^2}}$$

$$= \frac{\cancel{12728}}{\cancel{\quad}} = \cancel{\quad}$$

$$= \frac{12728}{104.614 \times 128.2342} = \underline{\underline{0.949}}$$

<u>OR</u>

x	y	$u=x-74.5$	$v=y-\bar{y}$	uv
78	125	3.5	-0.8	-2.8
89	137	14.5	11.2	162.4
97	156	22.5	30.2	679.5
69	112	-5.5	-13.8	75.9
59	107	-15.5	-18.8	291.4
79	138	4.5	12.2	54.9
68	123	-6.5	-2.8	18.2
57	108	-17.5	-17.8	311.5

u^2	v^2
12.25	0.64
210.25	125.44
506.25	912.04
30.25	190.44
240.25	353.44
20.25	148.84
42.25	7.84
306.25	316.84

$\Sigma u = 0$

$\Sigma v = -0.4$

$\Sigma uv = 1591$

$\Sigma u^2 = 1368$

$\Sigma v^2 = 2055.52$

$$r_{uv} = \frac{n\Sigma uv - \Sigma u \cdot \Sigma v}{\sqrt{n\Sigma u^2 - (\Sigma u)^2} \cdot \sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$

$$= \frac{8 \times 1591 - 0}{\sqrt{8 \times 1368} \cdot \sqrt{8 \times 2055.52 - 0.4^2}}$$

$$= \frac{12728}{104.612 \times 128.2342}$$

$$= \underline{\underline{0.9419}}$$

2. show that $-1 \leq r \leq 1$?

Here we use schwartz's inequality.

$$(\Sigma AB)^2 \leq \Sigma A^2 \cdot \Sigma B^2$$

If $A = x_i - \bar{x}$ and $B = y_i - \bar{y}$, then

$$\left[\Sigma (x_i - \bar{x})(y_i - \bar{y})\right]^2 \leq \left[\Sigma (x_i - \bar{x})\right]^2 \cdot \left[\Sigma (y_i - \bar{y})\right]^2$$

ie $$\frac{\left[\Sigma (x_i - \bar{x})(y_i - \bar{y})\right]^2}{\left[\Sigma (x_i - \bar{x})\right]^2 \left[\Sigma (y_i - \bar{y})\right]^2} \leq 1.$$

$\Rightarrow$ ie $r \leq 1. \Rightarrow -1 \leq r \leq 1.$

Note:

If $r>0$, the correlation is +'ve.

If $r<0$, the correlation is -'ve.

If $r=0$, the correlation doesn't exist.

If $r=+1$, correlation is perfect +'ve

If $r=-1$, correlation is perfect -'ve.

3. Show that

$2r\sigma_x\sigma_y = \sigma_x^2+\sigma_y^2-\sigma_{x-y}^2$?

proof.

Let $z=x-y$ then $\bar{z}=\bar{x}-\bar{y}$.

$z-\bar{z}=(x-y)-(\bar{x}-\bar{y})$

$=(x-\bar{x})-(y-\bar{y})$ or

$(z-\bar{z})^2=(x-\bar{x})^2+(y-\bar{y})^2-2(x-\bar{x})(y-\bar{y})$.

Summing up for n terms, we get

$$\frac{\Sigma(z-\bar{z})^2}{n} = \frac{\Sigma(x-\bar{x})^2}{n} + \frac{\Sigma(y-\bar{y})^2}{n} - \frac{2\Sigma(x-\bar{x})(y-\bar{y})}{n}$$

$$\Rightarrow \sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$$

$$\left\{ \text{since } r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y} \right\}.$$

4. If a, b, c, d are constants and r is the coefficient of correlation ~~conversion~~ b/w. x and y, then prove that the correlation b/w coefficient b/w. ax+d and cx+d is r and -r, according as, a and c are of the same or opposite sign?

Proof Let $z = ax+b$ and $w = cy+d$.

then $\bar{z} = a\bar{x}+b$ & $\bar{w} = c\bar{y}+d$.

The correlation coefficient b/w. z and w, $r_{zw} = \dfrac{\sigma_{zw}}{\sigma_z \cdot \sigma_w}$ — ①.

$$\sigma_{zw} = \frac{\sum(z-\bar{z})(w-\bar{w})}{n}$$

$$= \frac{\sum ac(x-\bar{x})(y-\bar{y})}{n}$$

$$\sigma_z^2 = \sum(z-\bar{z})^2$$

$$= \frac{\sum a^2(x-\bar{x})^2}{n}$$

$$\sigma_w^2 = \sum(w-\bar{w})^2$$

$$= \frac{\sum c^2(y-\bar{y})^2}{n}$$

Hence ① becomes

$$r_{zw} = \frac{ac\sum(x-\bar{x})(y-\bar{y})}{|a|\sqrt{\sum(x-\bar{x})^2}\,.\,|c|\sqrt{\sum(y-\bar{y})^2}}$$

If a and c are of the same sign then, ac is +ve and hence $r_{zw} = r$ & if their signs are opposite, then ac is -ve s.t. $r_{zw} = -r$.

5. If the correlation coefficient b/w x and y is ρ and if $u = x\cos\alpha + y\sin\alpha$, $v = y\cos\alpha - x\sin\alpha$. show that u & v are not correlated if

$$\tan 2\alpha = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \ ?$$

If u and v are not correlated, then, $\sigma_{uv} = 0$.

ie $\Sigma(u-\bar{u})(v-\bar{v}) = 0$.

ie $\Sigma\left\{(x\cos\alpha + y\sin\alpha) - (\bar{x}\cos\alpha + \bar{y}\sin\alpha)\right\}$
$\left\{(y\cos\alpha - x\sin\alpha) - (\bar{y}\cos\alpha - \bar{x}\sin\alpha)\right\}$
$= 0$.

$$\Rightarrow \Sigma\left[\left((x-\bar{x})\cos\alpha+(y-\bar{y})\sin\alpha\right)\cdot\left((y-\bar{y})\cos\alpha-(x-\bar{x})\sin\alpha\right)\right]=0$$

$$\Rightarrow \Sigma\left\{(x-\bar{x})(y-\bar{y})\cos^2\alpha-(x-\bar{x})^2\sin\alpha\cos\alpha+(y-\bar{y})^2\sin\alpha\cos\alpha-(x-\bar{x})(y-\bar{y})\sin^2\alpha\right\}=0$$

$$\Rightarrow (\cos^2\alpha-\sin^2\alpha)\Sigma(x-\bar{x})(y-\bar{y})+\sin\alpha\cos\alpha\left[\Sigma(y-\bar{y})^2-\Sigma(x-\bar{x})^2\right]=0$$

$$\Rightarrow \cos 2\alpha\, n\sigma_{xy}+\frac{\sin 2\alpha}{2}\, n\sigma_y^2-n\sigma_x^2=0$$

$$\Rightarrow 2n\sigma_{xy}+\frac{\tan 2\alpha}{(1)}(n\sigma_y^2-n\sigma_x^2)=0$$

$$\Rightarrow \tan 2\alpha=\frac{2n\sigma_{xy}}{n(\sigma_x^2-\sigma_y^2)}=\frac{2\sigma_{xy}}{\sigma_x^2-\sigma_y^2}$$

$$\Rightarrow \tan=\frac{2\rho\,\sigma_x\sigma_y}{\sigma_x^2-\sigma_y^2}\qquad\left(\because \rho=\frac{\sigma_{xy}}{\sigma_x\sigma_y}\right).$$

Regression

It is the average relationship b/w. 2 or more variables in terms of the original limits of the data.

Lines of Regression

The line of regression of y on x is given by $y-\bar{y} = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$,

where r is the correlation coefficient σ_x and σ_y are standard deviations.

The line of regression of x on y is given by

$$x-\bar{x} = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

Regression Coefficient

Regression coefficient of y on x,

$$\frac{r\sigma_y}{\sigma_x} = b_{yx} \quad \text{— (1)}$$

Regression coefficient of x on y, $\frac{r\sigma_x}{\sigma_y} = b_{xy}$ — ②

From ① and ②, we get,

$$r\frac{\sigma_y}{\sigma_x} \cdot r\frac{\sigma_x}{\sigma_y} = b_{yx} \cdot b_{xy}$$

$$r^2 = b_{yx} \cdot b_{xy}$$

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

correlation coefficient,

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

The regression coefficient b_{yx} and b_{xy} can be easily obtained by using the above ~~following~~ formula.

$\because$ σ_x and σ_y are +'ve, r has the same sign as that of the regression coefficients.

Thus r is +'ve, if b_{yx} & b_{xy} are both +'ve or −'ve, if both b_{yx} & b_{xy} are −'ve.

Note :- (1) $b_{yx} = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$.

$$b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2}.$$

(2) Regression coefficients are independent of change of origin but not of scale.

ie $b_{yx} = b_{vu}$

$$= \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2}$$ and

$$b_{xy} = b_{uv}$$

$$= \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma v^2 - (\Sigma v)^2}$$, where

$u = x - \bar{x}$, $v = y - \bar{y}$.

Problems

(1) Heights of fathers & their sons are given in cm. Find the 2 lines of regression & calculate the expected average height of the

son when the height of the father is 154 cm ?

x h(father)	150	152	155	157	160	161	164	166
y h(son)	154	156	158	159	160	162	161	164

x	y	$u = x - \bar{x}$	$v = y - \bar{y}$	u^2	v^2	uv
150	154	-8.125	-5.25	66.01	27.56	42.65
152	156	-6.125	-3.25	37.51	10.56	19.90
155	158	-3.125	-1.25	9.76	1.56	3.90
157	159	-1.125	0.25	1.26	0.06	0.28
160	160	-1.875	0.75	3.51	0.56	1.40
161	162	2.875	2.75	8.26	7.56	7.90
164	161	5.875	1.75	34.51	3.06	10.28
166	164	7.875	4.75	62.01	22.56	37.40

$\Sigma u = -3.75$, $\Sigma u^2 = 222.8$

$\Sigma v = 0$, $\Sigma v^2 = 73.48$

$\Sigma uv = 123.71$

$$\bar{x} = \frac{1265}{8} = \underline{158.125}$$

$$\bar{y} = \frac{1274}{8} = \underline{\underline{159.25}}$$

$$b_{yx} = b_{vu} = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2}$$

$$= \frac{8 \times 123.71 - 0}{8 \times 222.8 - (3.75)^2} = \frac{989.68}{1768.33}$$

$$= \underline{\underline{0.55}}$$

$$b_{xy} = b_{uv} = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma v^2 - (\Sigma v)^2}$$

$$= \frac{989.68}{8 \times 73.48 - 0} = \frac{989.68}{587.84} = \underline{\underline{1.68}}$$

The line of regression of y on x is given by $(y-\bar{y}) = b_{yx}(x-\bar{x})$

$$y - 159.25 = 0.55(x - 158.125)$$

$$y = 159.25 + 0.55x - 86.96$$

$$y = \underline{\underline{0.55x + 72.29}}$$

The line of regression of x on y is given by $x-\bar{x} = b_{xy}(y-\bar{y})$.

$x - 158.125 = 1.68\,(y - 159.25)$.

$x = 158.125 + 1.68\,x - 159.25 + 1.68y$.

$\Rightarrow x = \underline{\underline{1.68y + 109.415}}$

when the height of the father (x) is 154 cm, the height of the son (y) will be

$y = 0.55(154) + 72.29$

$= \underline{\underline{156.99}}$

2. The 2 lines of regression are $8x-10y+66=0$, $40x-18y-214=0$.
the variance of x is 9.
Find (1) the mean values of x and y
(2) the S.D. of y
(3) coefficient of correlation b/w. x and y.

Since both lines of regression pass through the point $(\bar{x}, \bar{y})$, we have,

$$8\bar{x} - 10\bar{y} + 66 = 0 \quad \text{--- (1)}$$

$$40\bar{x} - 18\bar{y} + -214 = 0 \quad \text{--- (2)}$$

on solving we get,

$\bar{y} = 17$, $\bar{x} = 13$.

Given variance of $x = 9$

ie $\sigma_x^2 = 9 \Rightarrow \underline{\sigma_x = 3}$

The equations of lines of regression can be written as

$$8x - 10y + 66 = 0$$

$$10y = 8x + 66$$

$$\underline{y = 0.8x + 6.6}$$

and

$$40x - 18y = 214$$

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

$$\underline{x = 0.45y + 5.35}$$

The regression coefficient of y on x is $r\frac{\sigma_y}{\sigma_x} = 0.8$ — (4)

The regression coefficient of x on y is $r\frac{\sigma_x}{\sigma_y} = 0.45$ — (5)

$(4) \times (5) \Rightarrow r\frac{\sigma_y}{\sigma_x} \cdot r\frac{\sigma_x}{\sigma_y} = 0.8 * 0.45$

$r^2 = 0.36$

$\Rightarrow r = \pm 0.6$

$\Rightarrow \underline{\underline{r = 0.6}}$ (the +ve root is taken becoz. regression coefficients are always +ve).

$r\sigma_y = 0.8\sigma_x$.

$\sigma_y = \frac{0.8\sigma_x}{r} = \frac{0.8 \times 3}{0.6}$

$= \underline{\underline{4}}$

(3) The following data were available $\bar{x} = 970$, $\bar{y} = 18$, $\sigma_x = 38$, $\sigma_y = 2$. Correlation coefficient $r = 0.6$. Find the line of regression and obtain the value of x and $y = 20$?

We know that the line of regression of x on y is given by

$$x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (y - \bar{y}) \quad \text{--- (1)}$$

Given $\bar{x} = 970$, $\bar{y} = 18$, $\sigma_x = 38$, $\sigma_y = 2$.

$$(1) \Rightarrow x - 970 = 0.6 \times \frac{38}{2} (y - 18)$$

$$\underline{x = 11.4y + 764.8}$$

when $y = 20$,

$$x = 11.4(20) + 764.8$$

$$\therefore x = \underline{992.8}$$

A If 'θ' is the angle b/w the regression lines, show that

$$\tan\theta = \left(\frac{\pm\, r^2 - 1}{r}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)?$$

The 2 regression equation are $y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$ and — ①

$(x - \bar{x}) = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$ — ②

Let m_1 = slope of line (1)

= coefficient of x.

$= r\frac{\sigma_y}{\sigma_x}$

② can be written as

$y - \bar{y} = \frac{\sigma_y}{r\sigma_x}(x - \bar{x})$ and hence

its slope $= m_2 =$ coefficient of x

$= \frac{\sigma_y}{r\sigma_x}$.

The angle θ btw. the regression lines is given by

$$\tan\theta = \pm\frac{m_1 - m_2}{1+m_1m_2} = \pm\frac{\frac{r\sigma_y}{\sigma_x} - \frac{\sigma_y}{r\sigma_x}}{1+\left(r\frac{\sigma_y}{\sigma_x}\right)\left(\frac{\sigma_y}{\sigma_x r}\right)}$$

$$= \pm\frac{r^2\sigma_y\sigma_x - \sigma_x\sigma_y}{r\sigma_x^2} \cdot \frac{r\sigma_x^2}{r\sigma_x^2 + r\sigma_y^2}$$

$$= \pm\frac{(r^2-1)\sigma_y\sigma_x}{r(\sigma_x^2+\sigma_y^2)}.$$

(5) The regression equation of x and y is $3y-5x+108=0$. If the mean value of y is 44 and if the variance of x is $\frac{9}{16}$ th of the variance of y, find the mean value of x and the correlation coefficient?

Since the line of regression passes through $(\bar{x}, \bar{y})$, we get

$3\bar{y} - 5\bar{x} + 108 = 0$

Given $\bar{y} = 44 \Rightarrow \bar{x} = 48$

Hence mean value of x is 48.

Given, $3y - 5x + 108 = 0$

$5x = 3y + 108$

$x = \frac{3}{5}y + \frac{108}{5}$, which is the required line of regression of x on y.

$b_{xy} = \frac{3}{5}$

$\frac{r\sigma_x}{\sigma_y} = \frac{3}{5}$ —(1)

Given $\sigma_x^2 = \frac{9}{16}\sigma_y^2$.

$\Rightarrow \sigma_x = \frac{3}{4}\sigma_y$.

(1) $\Rightarrow r\frac{3}{4}\sigma_y / \sigma_y = \frac{3}{5}$

$$r = \frac{12}{15} = \underline{\underline{0.8}}$$

Hence the correlation coefficient is $\underline{\underline{0.8}}$

Rank correlation

The spearman's rank correlation coefficient is given by

$$\gamma = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \quad d_i = x_i - y_i .$$

Problems

1\. Calculate the rank correlation for

Rank in x : 1 2 3 4

Rank in y : 3 4 2 1

R_x	R_y	$d = R_x - R_y$	d^2
1	3	-2	4
2	4	-2	4
3	2	1	1
4	1	3	9

$\sum d^2 = 18$

$$r = 1 - \frac{6 \cdot \Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 18}{4(4^2-1)} = 1 - 1.8$$

$$= -0.8$$

2. Calculate the rank correlation coefficient for

x	10	15	12	17	13	16	24	14	22
y	30	42	45	46	33	34	40	35	39

x	y	R_x	R_y	$d = R_x - R_y$	d^2
10	30	9	9	0	0
15	42	5	3	2	4
12	45	8	2	6	36
17	46	3	1	2	4

13	33	7	8	-1	1
16	34	4	7	-3	9
24	40	1	4	-3	9
14	35	6	6	0	0
22	39	2	5	-3	9

$\Sigma d^2 = 72$

$$r = 1 - \frac{6 \times \sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 72}{9(81-1)}$$

$$= 1 - 0.6 = \underline{\underline{0.4}}$$

Repeated Ranks

In the rank correlation formula we add the factor $\frac{m(m^2-1)}{12}$ to Σd^2, where m is the no. of times an item is repeated. This correlation factor is to be added for each repeated value.

Problems

1. Obtain the rank correlation for the following data.

x	68	64	75	50	64	80	75
y	62	58	68	45	81	60	68
	40	55	64				
	48	50	70				

In the ~~value~~ x-series, the value 75 occurs twice, they would have been given the ranks 2 and 3.

∴ The common rank is

$$\frac{2+3}{2} = 2.5$$

The value 64 occurs thrice, they would have been given the ranks 5, 6, 7.

∴ The common rank can be

$$\frac{5+6+7}{3} = 6$$

Similarly, in the y-series, 68 occurs twice, they would have been given the ranks 3 and 4.

$\therefore$ its common rank can be

$$\frac{3+4}{2} = 3.5$$

x	y	R_x	R_y	$d = R_x - R_y$	d^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16

$$\Sigma d^2 = 72$$

correction factor for 75 ($m=2$)

$$= \frac{m(m^2-1)}{12} = \frac{2(4-1)}{12} = \frac{1}{2}$$

correction factor for 68 ($m=2$)

$$= \frac{m(m^2-1)}{12} = \frac{2(4-1)}{12} = \frac{1}{2}$$

correction factor for 64 ($m=3$)

$$= \frac{m(m^2-1)}{12} = \frac{3(9-1)}{12} = 2$$

$$r = 1 - \frac{6 \times \left(\sum d_i^2 + \text{correction factor}\right)}{n(n^2-1)}$$

$$= 1 - \frac{6 \times \left(72 + \frac{1}{2} + \frac{1}{2} + 2\right)}{10(10^2-1)}$$

$$= 0.5455$$

2. A sample of 12 fathers and their eldest sons have the following data, about their heights in inches.

Fathers:	65	63	67	64	68	62	70
Sons:	68	66	68	65	69	66	68
	66	68	67	69	71		
	65	71	67	68	70		

Calculate the rank correlation coefficient.

Ans

common rank for 65 $= \dfrac{11+12}{2}$

$= \underline{\underline{11.5}}$

Common rank for 68 $= \dfrac{4+5}{2} = \underline{\underline{4.5}}$

Common rank for 67 $= \dfrac{6+7}{2} = \underline{\underline{6.5}}$

Common rank for 68 $= \dfrac{4+5+6+7}{4}$

$= \underline{\underline{5.5}}$

Common rank for 66 $= \dfrac{9+10}{2} = \underline{\underline{9.5}}$

x	y	R_x	R_y	$d = R_x - R_y$	d^2
65	68	9	5.5	3.5	12.25
63	66	11	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	10	11.5	−1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	−3.5	12.25
66	65	8	11.5	−3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	−1.5	2.25
69	68	3	5.5	−2.5	6.25
71	70	1	2	−1	1

$\sum d^2 = 72.5$

Correction Factor

In X-series, 68 is repeated twice $\therefore CF = \frac{2(2^2-1)}{12} = \frac{1}{2}$

In X-series, 67 is repeated twice $\therefore CF(67) = \frac{2(2^2-1)}{12} = \frac{1}{2}$

In Y-series, 68 is repeated four times $\therefore CF(68) = \frac{4(4^2-1)}{12} = 5$.

In Y-series, 66 is repeated twice,
$\therefore CF(66) = \frac{2(2^2-1)}{12} = \frac{1}{2}$

In Y-series, 65 is repeated twice,
$\therefore CF(65) = \frac{2(2^2-1)}{12} = \frac{1}{2}$

$$r = 1 - \frac{6 \times (72.5 + 0.5 + 0.5 + 5 + 0.5 + 0.5)}{12(12^2-1)}$$

$$= 1 - 0.2779 = 0.72203$$

Joint Distribution or Bivariate Distribution.

Given 2 random variables X and Y. We want to determine the joint probability function (pdf) and the relation b/w the joint pdf and pdf of X & Y.

Definition – Discrete Case:

Let X and Y be two distance random variables on the same sample space S. Suppose that X can assume the values $x_1, x_2, \ldots x_n$ and Y can assume the values $y_1, y_2 \ldots y_m$ the probablity distribution function or joint pdf of X & Y is defined by $F(x_i, y_j) = P(X = x_i, Y = y_j)$ for $i \in \mathbb{N}, j \in \mathbb{N}, i \leq n, j \leq m$,

where (i) $f(x_i, y_j) \geq 0$ for $\forall$ i & j.

(ii) $\sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i, y_j) = 1$

Marginal Probablity Function of X

The Marginal Probablity Function of X is given by

$$P(X=x_i) = f_X(x_i)$$
$$= \sum_{y=1}^{m} f(x_i, y_j) \quad \forall\, i \in \mathbb{N} \;\&\; i \leq n.$$

value of j varies.

Marginal Probablity Function of Y

The marginal probablity function of Y is given by

$$P(Y=y_j) = f_Y(y_j)$$
$$= \sum_{i=1}^{n} f(x_i, y_j) \quad \forall\, j \in \mathbb{N}, \;\&\; j \leq m.$$

value of i varies.

The joint probablity distribution of X and Y can be represented in the form of a rectangular table as follows:

$f_X \downarrow$	$X \backslash Y$	y_1	y_2	$\dots$	y_m
$f_X(x_1)$	x_1	$F(x_1, y_1)$	$F(x_1, y_2)$		$f(x_1, y_m)$
$f_X(x_2)$	x_2	$F(x_2, y_1)$	$F(x_2, y_2)$		$f(x_2, y_m)$
	$\vdots$				
$f_X(x_n)$	x_n	$F(x_n, y_1)$	$F(x_n, y_2)$		$f(x_n, y_m)$
1	$f_Y \rightarrow$	$f_Y(y_1)$	$f_Y(y_2)$		$f_Y(y_m)$

The joint distribution function of X and Y is defined by

$$F(x, y) = P(X \le x, Y \le y)$$
$$= \sum_{u \le x} \sum_{v \le y} F(u, v).$$

In the above table, $f(x, y)$ is the sum of all elements for which $x_i \le x$ and $y_i \le y$.

Note

In the discrete case, if $P(X=x_i)=p_i$, $P(Y=y_j)=q_j$ and $P(X=x_i, Y=y_j)=p_{ij}$, then the total probability functions or masses are $p_i=\sum_j p_{ij}$ and $q_j=\sum_i p_{ij}$

X \ Y	y_1	y_2	...	y_m	f_X (row total)
x_1	p_{11}	p_{12}		p_{1m}	$P(X=x_1) = p_1$
x_2	p_{21}	p_{22}		p_{2m}	$P(X=x_2) = p_2$
⋮					
x_n	p_{n1}	p_{n2}		p_{nm}	$P(X=x_n) = p_n$
$f_Y \rightarrow$ (column totals)	$P(Y=y_1) = q_1$	$P(Y=y_2) = q_2$		$P(Y=y_m) = q_m$	Grand total $=1$

Definition - Continuous Case

The joint distribution of two continuous random variables X & Y is defined as

$$F_{XY}(x,y) = F(x,y) = P(X \le x, Y \le y)$$
$$= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(x,y)\,dx\,dy.$$

where $f_{XY}(x,y) \ge 0$ and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$$

or

Simply we can write as follows.

$$F(x,y) = P(X \le x, Y \le y)$$
$$= \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)\,dx\,dy.$$

Joint Probablity Density Function

The joint pdf of X & Y is given by

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Marginal probablity Distribution Functions or Density Functions of X and Y.

Marginal density of X is given by

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

Marginal density of Y is given by

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

* Note that $f_X(x) = \frac{\partial}{\partial x} F(x,\infty)$ &

$$f_Y(y) = \frac{\partial}{\partial y} F(\infty,y)$$

Definition:

Two random variables X and Y are independant if

$F(x,y) = F_X(x) \; F_Y(y)$

Clearly if X and Y are idependant, then $f(x,y) = f_X(x) f_Y(y)$:

Conditional Probablity Density Functions

The conditional probablity density of X given Y is

$$f(x/y) = P(X=x \mid Y=y)$$

$$= \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{P_{ij}}{P_{\cdot j}}$$

$$= \frac{f(x,y)}{f_Y(y)} \text{ and}$$

the conditional probablity density of Y given X is

$$f(y/x) = P(Y=y \mid X=x)$$

$$= \frac{P(X=x, Y=y)}{P(X=x)}$$

$$= \frac{P_{ij}}{P_i} = \frac{f(x,y)}{f_X(x)}$$

Note that if X and Y are independent then

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

$$= \underline{\underline{f_X(x)}}, \text{ marginal pdf of } X.$$

Similarly $\underline{\underline{f(y/x)}} = f_Y(y)$, marginal pdf of Y.

Problems of discrete random variable

1. The joint ~~pmf~~ pdf of X and Y are given by $F(x,y) = C(x+2y)$ where $x = 0, 1, 2$, and $y = 0, 1, 2, 3$.

(a) Find C

(b) Find $P(X \leq 1)$, $P(Y \leq 2)$,

$P(X \leq 1, Y \leq 2)$, $P(X \leq 1 / Y \leq 2)$,

$P(X+Y \leq 3)$ and $P(X^2+Y^2 \leq 4)$.

(c) Find the marginal pmf and conditional distribution.

The joint pmf (pdf) can be represented as

X ↓ \ Y →	0	1	2	3	f_X ↓ (row totals)
0	0	2C	4C	6C	12C
1	C	3C	5C	7C	16C
2	2C	4C	6C	8C	20C
f_Y (column totals) →	3C	9C	15C	21C	1

(a) $\sum_i \sum_j P_{ij} = 1$.

Add all rowtotals or column totals to equate to 1
$12C + 16C + 20C = 48C = 1$.

$\therefore 48C = 1 \Rightarrow C = \frac{1}{48}$

(b) $P(X \le 1) = P(X=0) + P(X=1)$

$= \sum_{j=0}^{3} P_{0j} + \sum_{j=1}^{3} P_{1j}$

$= 12C + 16C = 28C$.

$= \frac{28}{48} = 0.5834$

$P(Y \le 2) = P(Y=0) + P(Y=1) + P(Y=2)$

$= \sum_{i=0}^{2} P_{i0} + \sum_{i=0}^{2} P_{i1} + \sum_{i=0}^{2} P_{i2}$.

$= 3C + 9C + 15C = 27C$.

$= \frac{27}{48} = 0.5625$

$$P(X \le 1, Y \le 2) = \sum_{i=0}^{1} \sum_{j=0}^{2} P_{ij}$$

$$= P_{00} + P_{01} + P_{02} + P_{10} + P_{11} + P_{12}$$

$$= 0 + 2C + 4C + C + 3C + 5C$$

$$= 15C = \frac{15}{48} = \underline{\underline{0.3125}}$$

conditional probability $P(X \le 1 / Y \le 2) = \dfrac{P(X \le 1, Y \le 2)}{P(Y \le 2)}$

$$= \frac{15C}{27C} = \frac{15}{27}$$

$$= \underline{\underline{0.556}}$$

$$P(X+Y \le 3) = P_{00} + P_{01} + P_{02} + P_{03} + P_{10} + P_{11} + P_{12} + P_{20} + P_{21}.$$

$$= 0 + 2C + 4C + 6C + C + 3C + 5C + 2C + 4C.$$

$$= 27C = \underline{\underline{0.5625}}$$

$$P(X^2+Y^2 \le 4) = P_{00}+P_{01}+P_{02}+P_{10}+ P_{11}+P_{20}$$

$$= 0+2c+4c+c+3c+2c.$$

$$= 12c = \frac{12}{48} = 0.25$$

(c) The marginal prob. f_x and f_y are given by

X	0	1	2
f_x	$\frac{12}{48}$	$\frac{16}{48}$	$\frac{20}{48}$.

row totals

Y	0	1	2	3.
f_y	$\frac{3}{48}$	$\frac{9}{48}$	$\frac{15}{48}$	$\frac{21}{48}$.

column totals

Conditional Distribution.

$f(x,y)$ is calculated as each element of ~~the~~ a column is divided ~~into~~ by the corresponding column sum.

$$P(X=i \mid Y=j) = \frac{P(X=i, Y=j)}{P(Y=j)}$$

$$= \frac{P_{ij}}{\sum_i P_{ij}}.$$

$f(x/y)$.

X \ Y	0	1	2	3.
0	0	2/9	4/15	6/21
1	1/3	3/9	5/15	7/21
2.	2/3.	4/9	6/15	8/21.

Similarly, f(y/x) is calculated as each element of a row is divided by the corresponding row sum.

Since $P(Y=j/X=i) = \frac{P(X=i, Y=j)}{P(X=i)}$

$$= \frac{P_{ij}}{\sum_j P_{ij}}$$

X \ Y	0	1	2	3
0	0	2/12	4/12	6/12
1	1/16	3/16	5/16	7/16
2	2/20	4/20	6/20	8/20

Problems of continuous random variable

1. The joint density X and Y is given by $f(x,y) = \begin{cases} e^{-(x+y)}, \forall\ x \geq 0, y \geq 0 \\ 0 \quad , \text{otherwise.} \end{cases}$

Find $P(\frac{1}{2} < X < 2, 0 < Y < 4)$?

$P(\frac{1}{2} < X < 2, 0 < Y < 4)$

$$= \int_{1/2}^{2} \int_{0}^{4} f(x,y)\, dy\, dx$$

$$= \int_{1/2}^{2} \int_{0}^{4} e^{-(x+y)}\, dy\, dx.$$

$$= \int_{\frac{1}{2}}^{2} \int_{0}^{4} e^{-x} \left(\int_{0}^{4} e^{-y} dy \right) dx$$

$$= \int_{1/2}^{2} e^{-x} \left(\frac{e^{-y}}{-1} \right)_{0}^{4} dx$$

$$= (1-e^{-4}) \int_{1/2}^{2} e^{-x} dx.$$

$$= (1-e^{-4}) \left(\frac{e^{-x}}{-1} \right)_{\frac{1}{2}}^{2}$$

$$= (1-e^{-4})(e^{-1/2} - e^{-2})$$

(2) If $f(x,y) = \begin{cases} x^2 + \frac{xy}{3}, \forall\ 0<x<1, \\ \qquad\qquad 0<y<2. \\ 0 \quad , \text{otherwise}. \end{cases}$

is a joint pdf, find $P(X > \frac{1}{2})$, $P(X<Y)$, $P(Y< \frac{1}{2})$, $P(X<\frac{1}{2} / Y<\frac{1}{2})$, $P(X<\frac{1}{2}, Y>\frac{1}{2})$?

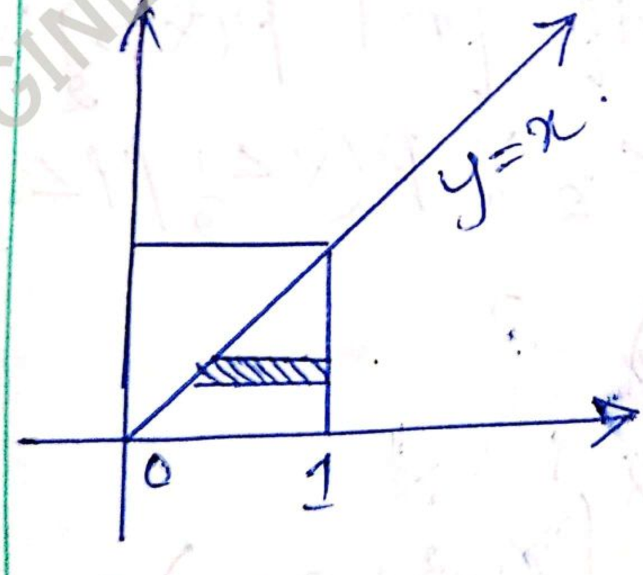
$$P(X > \frac{1}{2}) = \int_{0}^{2} \int_{\frac{1}{2}}^{1} \left(x^2 + \frac{xy}{3}\right) dx\, dy.$$

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

$$= \int_0^2 \left(\frac{1}{3} + \frac{1}{6} y - \frac{1}{24} - \frac{y}{24} \right) dy.$$

$$= \left(\frac{1}{3} y + \frac{1}{6} \frac{y^2}{2} - \frac{y}{24} - \frac{y^2}{48} \right)_0^2$$

$$= \frac{2}{3} + \frac{1}{3} - \frac{1}{12} - \frac{1}{12} = 1 - \frac{1}{6} = \frac{5}{6}$$



$P(X<Y) = \iint_R f(x,y)\,dx\,dy$

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

$$= \int_0^1 \left(\frac{x^3}{3} + \frac{y}{3} \cdot \frac{x^2}{2} \right)_y^1 dy.$$

$$= \int_0^1 \left(\frac{1}{3}+\frac{1}{6}y-\frac{y^3}{3}-\frac{y^3}{6}\right)dy$$

$$= \left(\frac{1}{3}y+\frac{y^2}{12}-\frac{y^4}{12}-\frac{y^4}{24}\right)_0^1$$

$$= \frac{8-1}{24} = \underline{\underline{\frac{7}{24}}}$$

$$P\left(Y<\frac{1}{2}\right) = \iint_R f(x,y)\,dx\,dy.$$

$$= \int_0^{1/2}\int_0^1 \left(x^2+\frac{xy}{3}\right)dx\,dy.$$

$$= \int_0^{1/2} \left(\frac{x^3}{3}+\frac{x^2}{2}\cdot\frac{y}{3}\right)_0^1 dy.$$

$$= \int_0^{1/2} \left(\frac{1}{3}+\frac{1}{6}y\right)dy.$$

$$= \left(\frac{1}{3}y+\frac{1}{6}\frac{y^2}{2}\right)_0^{\frac{1}{2}}$$

$$= \frac{1}{6} + \frac{1}{12} \times \frac{1}{4} = \frac{1}{6} + \frac{1}{48}$$

$$= \frac{8+1}{48} = \frac{9}{48} = \frac{3}{16}$$

$$P\left(X < \tfrac{1}{2} / Y < \tfrac{1}{2}\right) = \frac{P\left(X < \tfrac{1}{2}, Y < \tfrac{1}{2}\right)}{P\left(Y < \tfrac{1}{2}\right)}$$

$$P\left(0 < X < \tfrac{1}{2}, 0 < Y < \tfrac{1}{2}\right) =$$

$$\int_0^{1/2}\int_0^{1/2} \left(x^2 + \frac{xy}{3}\right) dx\,dy.$$

$$= \int_0^{1/2} \left(\frac{x^3}{3} + \frac{x^2}{2} \cdot \frac{y}{3}\right)_0^{1/2} dy.$$

$$= \int_0^{1/2} \left(\frac{1}{24} + \frac{y}{24}\right) dy.$$

$$= \left(y/24 + \frac{y^2}{48}\right)_0^{1/2}$$

$$= \frac{1}{48} + \frac{1}{192} = \frac{4+1}{192} = \frac{5}{192}$$

$$P\left(X < \tfrac{1}{2} \mid Y < \tfrac{1}{2}\right) = \frac{P\left(X < \frac{1}{2}, Y < \frac{1}{2}\right)}{P\left(Y < \frac{1}{2}\right)}$$

$$= \frac{\frac{5}{192}}{\frac{3}{16}} = \frac{5}{16} = 0.1389$$

$$P\left(X < \tfrac{1}{2}, Y > \tfrac{1}{2}\right) = \int_0^{1/2}\int_{1/2}^{2}\left(x^2 + \frac{xy}{3}\right) dy\, dx$$

$$= \int_0^{1/2}\left(x^2 y + \frac{xy^2}{6}\right)_{1/2}^{2} dx$$

$$= \cancel{\int_0^{1/2} 4y + \frac{1}{3}y^2 - \frac{1}{4}y - \frac{1}{12}y^2)}$$

$$= \int_0^{1/2}\left(2x^2 + \frac{2}{3}x - \frac{1}{2}x^2 - \frac{1}{24}x\right) dx$$

$$= \int_0^{1/2}\left(\frac{3}{2}x^2 + \frac{5}{8}x\right) dx$$

$$= \left(\frac{\cancel{3}x^3}{2\times \cancel{3}} + \frac{5x^2}{8\times 2} \right)_0^{1/2}$$

$$= \left(\frac{x^3}{2} + \frac{5x^2}{16} \right)_0^{1/2}$$

$= 0.0625 + 0.078125 = 9/64.$

$= \underline{0.140625}$

3. The joint pdf of X and Y is given by $f(x,y) = \begin{cases} cxy, & \forall\, 0 < x < 4,\ 1 < y < 5 \\ 0, & \text{otherwise.} \end{cases}$

Find (i) c (ii) $P(X \geq 3, Y \leq 2)$

(iii) $P(1 < X < 2,\ 2 < Y < 3)$

(iv) $P(X + Y < 3)$

(v) marginal distributions of X and Y.

Since $f(x,y)$ is a pdf,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1.$$

$$\int_{1}^{5}\int_{0}^{4} cxy\,dx\,dy = 1.$$

$$\int_{1}^{5} \left(\frac{cx^2}{2}y\right)_0^4 dy = 1.$$

$$\int_{1}^{5} \left(c\frac{16}{2}y\right) dy = 1$$

$$8\int_{1}^{5} cy\,dy = 1$$

$$8c\left(\frac{y^2}{2}\right)_1^5 = 1$$

$$\cancel{8c\left(\frac{25}{2}-1\right)=1+\frac{1}{2}}. \qquad 8c\left(\frac{25}{2}-\frac{1}{2}\right)=1.$$

$$\cancel{\Rightarrow c = \frac{2}{8\times 25} = 0.01 = \frac{1}{100}}$$

$8c \times 12 = 1.$

$$c = \frac{1}{8 \times 12} = 0.0104 = \frac{1}{96}$$

(ii) $P(X \geq 3, Y \leq 2)$

$$= \int_1^2 \int_3^4 \frac{1}{96} xy \, dx\, dy.$$

$$= \int_1^2 \left(\frac{x^2 y}{96 \times 2}\right)_3^4 dy.$$

$$= \int_1^2 y \left[\frac{16}{96 \times 2} - \frac{9}{96 \times 2}\right] dy.$$

$$= \int_1^2 \frac{7}{192} y \, dy = \left(\frac{7y^2}{192 \times 2}\right)_1^2$$

$$= \frac{7 \times 4}{192 \times 2} - \frac{7 \times 1}{192 \times 2} =$$

$$= \frac{28 - 7}{192 \times 2} = \frac{21}{192 \times 2} = 0.0547 = \frac{7}{128}$$

(iii) $P(1 < x < 2, 2 < y < 3)$

$$= \int_2^3 \int_1^2 \frac{1}{96} xy \, dx dy$$

$$= \int_2^3 y \left(\frac{x^2}{2 \times 96} \right)_1^2 dy$$

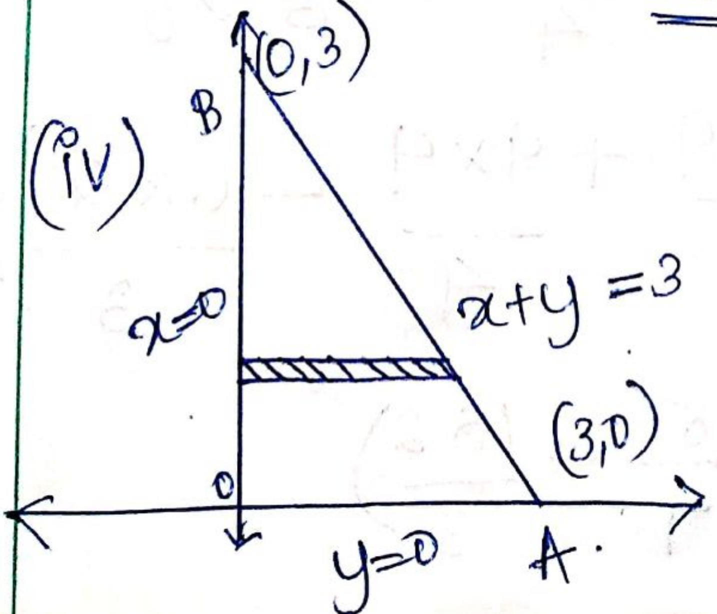
$$= \int_2^3 \left(\frac{4}{2 \times 96} - \frac{1}{2 \times 96} \right) y dy$$

$$= \frac{3}{2 \times 96} \left(\frac{y^2}{2} \right)_2^3$$

$$= \frac{3}{2 \times 96} \left(\frac{9}{2} - \frac{4}{2} \right) = \cancel{0.013} \cancel{\frac{5}{384}}$$

$$= 0.0391 = \frac{5}{128}$$

(iv)



$$P(x+y<3) = \iint_a f(x,y)\,dx\,dy$$

$$= \int_0^3 \int_0^{3-y} \frac{1}{96} xy\,dx\,dy$$

$$= \int_0^3 \frac{1}{96}\, y \left(\frac{x^2}{2}\right)_0^{3-y} dy$$

$$= \int_0^3 \frac{y}{192} (3-y)^2\,dy.$$

$$= \int_0^3 y/_{192} (9+y^2-6y)\,dy$$

$$= \frac{1}{192}\int_0^3 (9y + y^3 - 6y^2)\,dy$$

$$= \frac{1}{192}\left(\frac{9y^2}{2} + \frac{y^4}{4} - \frac{6y^3}{3}\right)_0^3$$

$$= \frac{1}{192}\left(\frac{9\times 9}{2} + \frac{9\times 9}{4} - \frac{6\times 27}{3}\right)$$

$$= \frac{1}{192}\left(\frac{81}{2} + \frac{81}{4} - \frac{162}{3}\right)$$

$$= \frac{1}{192} \times \frac{27}{4} = \frac{9}{256} = 0.01352$$

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

$$= \int_1^5 \frac{1}{96} xy\,dy = \left(\frac{xy^2}{96\times 2}\right)_1^5$$

$$= x\left(\frac{25-1}{96\times 2}\right) = \frac{x}{8}, \; 0<x<4$$

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

$$= \int_0^4 \frac{1}{96} xy\,dx$$

$$= \left(\frac{x^2 y}{96\times 2}\right)_0^4 = \left(\frac{16-0}{96\times 2}\right) y$$

$$= \frac{16}{96\times 2} y = \frac{y}{12}, \; 1<y<5$$

④ If the joint pdf. of X and Y is given by $F(x,y) = (1-e^{-x})(1-e^{-y})$

$\forall\ x > 0, y > 0$.

(i) Find the joint pdf of X and Y.

(ii) Find the marginal densities of X and Y.

(iii) $P(1 < x < 3, 1 < Y < 2)$.

(i) Given $F(x,y) = (1-e^{-x})(1-e^{-y})$

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

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x\ y)$$

$$= \frac{\partial^2}{\partial x \partial y}\left(1 - e^{-y} - e^{-x} + e^{-(x+y)}\right)$$

$$= \frac{\partial}{\partial x}\left[e^{-y} - e^{-(x+y)}\right]$$

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

$$f(x,y) = \begin{cases} e^{-(x+y)}, & \forall x>0, y>0. \\ 0, & \text{otherwise.} \end{cases}$$

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

$$= \int_0^{\infty} e^{-(x+y)}\,dy$$

$$= \left(-e^{-(x+y)}\right)_0^{\infty} = \underline{\underline{e^{-x}}}$$

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

$$= \int_0^{\infty} e^{-(x+y)}\,dx$$

$$= \left(-e^{-(x+y)}\right)_0^{\infty} = \underline{\underline{e^{-y}}}$$

$$f(x)\cdot f(y) = e^{-x}\cdot e^{-y} = e^{-x-y}$$

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

$\therefore$ X and Y are independant.

(iii) $P(1 < x < 3, 1 < y < 2)$

$$= \int_1^2 \int_1^3 e^{-(x+y)} \, dx \, dy$$

$$= \int_1^2 \left(-e^{-(x+y)}\right)_1^3 dy$$

$$= \int_1^2 \left[-e^{-(3+y)} + e^{-(1+y)}\right] dy$$

$$= \left[e^{-(3+y)} - e^{-(1+y)}\right]_1^2$$

$$= e^{-5} - e^{-3} - e^{-4} + e^{-2}$$

$$= 0.07397$$

(5) Examine whether the variables x and y are independent, whose joint density is

$$f(x,y) = x e^{-x(y+1)}, \forall \cdot x > 0, y < \infty.$$

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

$$= \int_0^{\infty} x e^{-x(y+1)} \, dy$$

A

$$= \int_0^\infty x e^{-x(y+1)} dy$$

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

$$= \underline{\underline{e^{-x}}}$$

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

$$= \int_0^\infty x e^{-x(y+1)} dx$$

$$= \left(x\frac{e^{-x(y+1)}}{-(y+1)}\right)_0^\infty - \left(\frac{e^{-x(y+1)}}{(y+1)^2}\right)_0^\infty$$

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

$$f(x)\cdot f(y) = e^{-x}\cdot\frac{1}{(y+1)^2} = \underline{\underline{\frac{e^{-x}}{(y+1)^2}}}$$

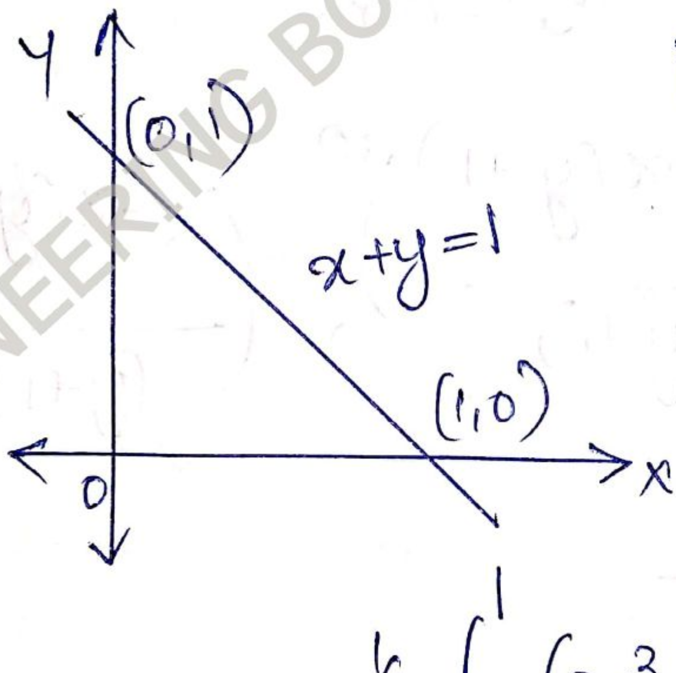
Since $f(x)\cdot f(y) \neq f(x,y)$

X & Y <u>aren't</u> independant.

(6) The joint density function of 2 random variables X and Y is $f(x,y) = \begin{cases} \frac{1}{3}(3x^2 + xy), & \forall\ 0 \le x \le 1,\ 0 \le y \le 2 \\ 0, & \text{otherwise}. \end{cases}$

Find $P(X+Y \ge 1)$?

$$P(X+Y \ge 1) = 1 - P(X+Y < 1)$$



$$P(X+Y<1) = \int_0^1 \int_0^{1-y} f(x,y)\,dx\,dy$$

$$= \int_0^1 \int_0^{1-y} \left(\frac{3x^2 + xy}{3}\right) dx\,dy$$

$$= \frac{1}{3} \int_0^1 \left(x^3 + \frac{x^2 y}{2}\right)_0^{1-y} dy.$$

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

$$= \frac{1}{3}\int_0^1 \left(\frac{-y^3}{2} + \frac{2y^2}{1} - \frac{5y}{2} + 1\right) dy$$

$$= \frac{1}{3}\left(\frac{-y^4}{4\times 2} + \frac{2y^3}{3} - \frac{5y^2}{4} + y\right)_0^1$$

$$= \frac{1}{3}\left(\frac{-1}{8} + \frac{2}{3} - \frac{5}{4} + 1\right) = \frac{19}{144} = \underline{\underline{0.132}}$$

$$P(X+Y \geq 1) = 1 - \frac{19}{144} = \frac{125}{144} = \underline{\underline{0.8681}}$$

(7) The joint density of X and Y is given by $f(x,y) = \begin{cases} 8xy, & \forall\, 0 \leq x \leq 1,\ 0 \leq y \leq x. \\ 0, & \text{otherwise}. \end{cases}$

Find the conditional densities of X and Y ?.

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

$$= \int_0^x 8xy\, dy = \left(8x\frac{y^2}{2}\right)_0^x$$

$$= \left(4xy^2\right)_0^x$$

$= 4x \cdot x^2 - 0 = \underline{\underline{4x^3}} \quad \forall\ 0 \le x \le 1.$

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

$$= \int_0^1 8xy\,dx = \left(\frac{8x^2}{2}y\right)_0^1$$

$$= (4x^2 y)_0^1 = \underline{\underline{4y}} \quad \forall\ 0 \le y \le x.$$

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{8xy}{4y}$$

$$= \underline{\underline{2x}} \ ; \forall\ 0 \le x \le 1.$$

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{8xy}{4x^3}$$

$$= \underline{\underline{\frac{2y}{x^2}}} \quad \forall\ 0 \le y \le x.$$

(8) The joint density of X and Y are given by the

$$f(x) = \begin{cases} \frac{x}{4}(1+3y^2), & 0 < x < 2,\ 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find (i) $f_X(x)$, (ii) $f_Y(y)$

(iii) $f(x/y)$, (iv) $P\left(\frac{1}{4} < x < \frac{1}{2} \,/\, y = \frac{1}{3}\right)$.

(i) $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 \frac{x}{4}(1+3y^2)\,dy$

$= \frac{x}{4}\left(y+y^3\right)_0^1 = \frac{x}{4}\cdot 2$

$= \frac{x}{2}, \ \forall\ \underline{0 < x < 2}$

(ii) $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^2 \frac{x}{4}(1+3y^2)\,dx$

$= \frac{1+3y^2}{4}\left(\frac{x^2}{2}\right)_0^2 = \left(\frac{1+3y^2}{8}\right)(4)$

$= \frac{1+3y^2}{2}, \ \forall\ 0 < y < 1.$

(iii) $f(x/y) = \frac{f(x,y)}{f(y)} = \frac{\frac{x}{4}(1+3y^2)}{\frac{(1+3y^2)}{2}}$

$= \left(\frac{x}{4}\right) \div \left(\frac{1}{2}\right) = x/2, \ \forall\ \underline{0 < x < 2}$

(iv) $P\left(\frac{1}{4} < x < \frac{1}{2} / y = \frac{1}{3}\right)$

$$= \frac{P\left(\frac{1}{4} < x < \frac{1}{2}, Y = \frac{1}{3}\right)}{P\left(Y = \frac{1}{3}\right)}$$

$$P\left(1/4 < X < 1/2, Y = 1/3\right) = \int_{1/4}^{1/2} \frac{x}{4}(1+3y^2)\,dx\,dy$$

$$= \int_{1/4}^{1/2} \frac{x(1+3y^2)}{4}\,dx\,dy$$

$$= \left(\frac{1+3y^2}{4}\right)\left(\frac{x^2}{2}\right)_{1/4}^{1/2} = \left(\frac{1+3y^2}{4}\right)\left(\frac{3}{32}\right)$$

$\left\{\begin{array}{l}\text{put} \\ y = 1/3\end{array}\right\}$ $$= \left(\frac{1 + 3 \times (1/3)^2}{4}\right)(3/32)$$

$$= \underline{\underline{1/32}}$$

$$P\left(Y = \frac{1}{3}\right) = f_Y(y)\Big|_{y = 1/3}$$

$$= \frac{1 + (1/3)^2 \times 3}{2} = \frac{2}{3}$$

$$P\left(\frac{1}{4} < X < \frac{1}{2} \Big/ Y = \frac{1}{3}\right) = \frac{P\left(\frac{1}{4} < X < \frac{1}{2}, Y = \frac{1}{3}\right)}{P(Y = \frac{1}{3})}$$

$$= \frac{1/32}{2/3} = \frac{3}{64} = 0.0469$$

(9) The joint probability density function of X and Y is

$$f(x,y) = \begin{cases} \frac{8}{9}xy, & 1 \le x \le y \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

(i) Find the marginal density functions of X and Y.

(ii) Find the conditional density function of Y given X = x.

(i) $$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_x^2 \frac{8xy}{9}\,dy.$$

$$= \left(\frac{8x}{9}\frac{y^2}{2}\right)_x^2 = \frac{4x}{9}(4-x^2),$$

$$\forall\ 1 \le x \le 2.$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{1}^{y} \frac{8}{9} xy\,dx$$

$$= \left(\frac{8x^2}{2} \cdot \frac{y}{9}\right)_1^y$$

$$= \frac{4y}{9}(y^2-1),\ \forall\ 1 \le y \le 2$$

The conditional density function of Y given $X=x$ is

$$f(y/x) = \frac{f(xy)}{f(x)} = \frac{\frac{8}{9}xy}{\frac{4x}{9}(4-x^2)}$$

$$= \frac{2y}{4-x^2},\quad x \le y \le 2$$

(10) If $f(x,y) = \begin{cases} 2e^{-(2x+y)}, & x \ge 0,\ y \ge 0 \\ 0, & \text{otherwise.} \end{cases}$ is

a joint pdf, find the distribution function?

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

$$= \int_{0}^{x}\int_{0}^{y} 2e^{-(2x+y)}\,dy\,dx$$

$$= 2\int_{0}^{x} e^{-2x}\left(\frac{e^{-y}}{-1}\right)_{0}^{y} dx$$

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

$$= 2(1-e^{-y})\left(\frac{e^{-2x}}{-2}\right)_{0}^{x}.$$

$$= -(1-e^{-y})(e^{-2x}-1)$$

$$= (1-e^{-y})(1-e^{-2x})$$

$$\therefore F(x,y) = \begin{cases} (1-e^{-y})(1-e^{-2x}), & \forall\ x \geq 0,\ y \geq 0. \\ 0, & \text{otherwise.} \end{cases}$$