UNIT - IV

CURVE FITTING

Part-I

Method of least squares

- Fitting a st. line (linear curves)
- Fitting a parabola (or) (Non-linear curves) (or) (second degree)
- Fitting general curves
  - Exponential
  - Power curves

→ Curve fitting

Several equation of different types can be obtained to express the given data approximately, but the problem is to find equation of the curve of best fit which may be most suitable for predicting the unknown values.

→ The process of finding such an equation of best fit is known as curve fitting.

→ The method we use to fit a curve is known as method of least squares

Fitting a st line (or) Linear curve by the method of least squares

Working rule

Step 1: let $y = a + bx$ —①

be a st line to be best fitted by the M.L.S where a, b are parameters.

Step 2: We have normal equations to fit a st. line

$$na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

Step 3:

x_i	y_i	$x_i y_i$	x_i^2
x_1	y_1		
x_2	y_2		
⋮	⋮		
x_n	y_n		
Σx_i	Σy_i	$\Sigma x_i y_i$	Σx_i^2

Step 4: Substitute all the values from the table in eqn ②
and solving we get a and b values

Step 5: substitute a and b values in eqn ①

$$S = \sum_{i=1}^{n} [y_i - (a+bx_i)]^2$$

$$S = [y_1 - (a+bx_1)]^2 + [y_2 - (a+bx_2)]^2 + \cdots + [y_n - (a+bx_n)]^2$$

$$\frac{\partial S}{\partial a} = 0 = 2[y_1 - (a+bx_1)](-1) + 2[y_2 - (a+bx_2)](-1) + 2[y_n - (a+bx_n)](-1)$$

$$(y_1 + y_2 + \cdots + y_n) = a + a + \cdots + a(n) + bx_1 + bx_2 + \cdots + bx_n$$

$$\Sigma y_i = na + b\Sigma x_i$$

1. Fitting a st.line (or) linear curve by the method of least squares to the following data

Sol: Given data

x	1	2	3	4	5
y	14	17	40	55	68

Sol: Here $n=5$

Table

Tabular form

x_i	y_i	x_iy_i	x_i^2
1	14	14	1
2	17	34	4
3	40	120	9
4	55	220	16
5	68	340	25

$\Sigma x_i = 15$ $\Sigma y_i = 194$ $\Sigma x_iy_i = 728$ $\Sigma x_i^2 = 55$

$\Sigma y_i = na + b\Sigma x_i$

$5a + 15b = 194$ -①

$a\Sigma x_i + b\Sigma x_i^2 = \Sigma x_iy_i$

$15a + 55b = 728$ -②

solving the above equation

$a = -5$ and $b = 14.6$

sub in ①

$\boxed{y = -5 + (14.6)x}$ which is req, st line to be fitted

Note:

let $y = ax + b$ -①

be a st line to be best fitted by the method of least squares

$nb + a\Sigma x_i = \Sigma y_i$

$b\Sigma x_i + a\Sigma x_i^2 = \Sigma x_iy_i$

where $i = 1$ to n

2. Fit a st. line or linear curve to the following data

x	0	5	10	15	20	25
y	12	15	17	22	24	30

Sol Here $n=6$

Tabular form.

x_i	y_i	x_iy_i	x_i^2
0	12	0	0
5	15	75	25
10	17	170	100
15	22	330	225
20	24	480	400
25	30	750	625

$\Sigma x_i = 75$ $\Sigma y_i = 120$ $\Sigma x_iy_i = 1805$ $\Sigma x_i^2 = 1375$

$\Sigma y_i = na + b\Sigma x_i$

$6a + 75b = 120$ —①

$a\Sigma x_i + b\Sigma x_i^2 = \Sigma x_iy_i$

$75a + 1375b = 1805 \rightarrow \div 5$

$15a + 275b = 361$ —②

solving above equations.

eq①×5 and eq②×2

$30a + 375b = 60$

$30a + 550b = 722$

$-175b = -122$

$\boxed{b = 0.6971}$

sub b in eq①

$6a + 75(0.6971) = 120$

$6a = 120 - 52.2825$

$\boxed{a = 11.28}$

$\boxed{y = 11.28 + (0.6971)x}$ which is the

req st line to best fitted

→ Fitting a second degree curve (or) parabola (or) non-linear curve by the method of least squares.

working rule:

step 1: let $y = a + bx + cx^2$ — ① be a parabola to be best fitted by the M.L.S

where a, b, c are parameters.

step 2: Normal equation to fit a parabola

$$\left.\begin{aligned} na + b\Sigma x_i + c\Sigma x_i^2 &= \Sigma y_i \\ a\Sigma x_i + b\Sigma x_i^2 + c\Sigma x_i^3 &= \Sigma x_i y_i \\ a\Sigma x_i^2 + b\Sigma x_i^3 + c\Sigma x_i^4 &= \Sigma x_i^2 y_i \end{aligned}\right\} — ②$$

where i = 1 to n and n = no. of observation

step 3:

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
x_1	y_1	x_1^2	x_1^3	x_1^4	$x_1 y_1$	$x_1^2 y_1$
x_2	y_2	x_2^2	x_2^3	x_2^4	$x_2 y_2$	$x_2^2 y_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_n	y_n	x_n^2	x_n^3	x_n^4	$x_n y_n$	$x_n^2 y_n$
$\Sigma x_i =$	$\Sigma y_i =$	$\Sigma x_i^2 =$	$\Sigma x_i^3 =$	$\Sigma x_i^4 =$	$\Sigma x_i y_i =$	$\Sigma x_i^2 y_i =$

step 4: sub all the Σ value from the table in eqn ② and solving we get a, b, c values

step 5: solve a, b, c in eqn ① we get required curves.

Scanned by CamScanner

1. Fit a second degree curve $y = a+bx+cx^2$ from the following data by M.L.S

x	0	1	2
y	1	6	17

Sol: step 1: let $y = a+bx+cx^2$ — ① be a parabola to be best fitted by the M.L.S

where a, b, c are parameters here $n=3$

step 2: Normal equation to fit a parabola

$$\left.\begin{aligned} na + b\Sigma x_i + c\Sigma x_i^2 &= \Sigma y_i \\ a\Sigma x_i + b\Sigma x_i^2 + c\Sigma x_i^3 &= \Sigma x_i y_i \\ a\Sigma x_i^2 + b\Sigma x_i^3 + c\Sigma x_i^4 &= \Sigma x_i^2 y_i \end{aligned}\right\} \quad -②$$

where $i = 1$ to n and n = no. of observation.

step 3:

Tabular form:

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	6	1	1	1	6	6
2	17	4	8	16	34	68

$\Sigma x_i = 3$, $\Sigma y_i = 24$ $\Sigma x^2 = 5$ $\Sigma x^3 = 9$ $\Sigma x^4 = 17$ $\Sigma xy = 40$ $\Sigma x^2y = 74$

sub all values in ②

$$3a + 3b + 5c = 24$$

$$3a + 5b + 9c = 40$$

$$5a + 9b + 17c = 74$$

Solving the above eqns we get

$a=1$, $b=2$ and $c=3$.

sub a, b, c in ①

$\boxed{y=1+2x+3x^2}$ is the required 2nd degree curves.

2. Fit a parabola of the form $y = ax^2+bx+c$ from the following

x	1	2	3	4	5
y	10	12	8	10	14

sv step 1:- let $y = a+bx+cx^2$ —① be a parabola to be best fitted by the M.L.S. where a, b, c are parameters

step 2:- Normal equation to fit a parabola

$$\left.\begin{aligned} na + b\Sigma x_i + c\Sigma x_i^2 &= \Sigma y_i \\ a\Sigma x_i + b\Sigma x_i^2 + c\Sigma x_i^3 &= \Sigma x_i y_i \\ a\Sigma x_i^2 + b\Sigma x_i^3 + c\Sigma x_i^4 &= \Sigma x_i^2 y_i \end{aligned}\right\} -②$$

where $i=1$ to n and $n=$ no. of observation

step 3:-

x	y	x^2	x^3	x^4	xy	x^2y
1	10	1	1	1	10	10
2	12	4	8	32	24	48
3	8	9	27	81	24	72
4	10	16	64	256	40	160
5	14	25	125	625	70	350

$\Sigma x=15$ $\Sigma y=54$ $\Sigma x^2=55$ $\Sigma x^3=225$ $\Sigma x^4=979$ $\Sigma xy=168$ $\Sigma x^2y=646$.

sub all the values in normal equation

$5c + 15b + 55a = 54$

$15c + 55b + 225a = 168$

$55c + 225b + 979a = 640$

Solving the above eqn

$c = 14$, $b = 3.68$ and $a = 0.71$

sub. a,b,c values in eq ①.

$$y = (0.71)x^2 - (3.68)x + 14$$

is the req parabola to be best filled by the method of least squares.

→ Fitting an exponential curve of the form $y = ae^{bx}$ by method of least squares.

Step 1: let $y = ae^{bx}$ — ①

be an exponential curve to be fitted by the MLS where a, b are parameters

Step 2: Apply '$\log_e$' on b.s of ①

$$\log_e y = \log_e (ae^{bx})$$

$$\log_e y = \log_e a + \log_e e^{bx}$$

$$\boxed{\log_e y = \log_e a + bx}$$

$$Y = A + BX \text{ — ②} \qquad \because \boxed{\log_e e^{bx} = bx}$$

which is a st. line

where $Y = \log_e y$, $A = \log_e a$ $\Rightarrow \boxed{a = e^A}$

$B = b$, $X = x$

Step 3: Normal eqns of ②

$$\left.\begin{aligned} nA + B\Sigma X_i &= \Sigma Y_i \\ A\Sigma X_i + B\Sigma X_i^2 &= \Sigma X_i Y_i \end{aligned}\right\} \text{— ③}$$

Scanned by CamScanner

step 4: Tabular form

$X=x$	y	$Y=\log_e^y$	X^2	XY

$\Sigma X=$ $\Sigma Y=$ $\Sigma X^2=$ $\Sigma XY=$

step 5: sub all the values in ③ and solve

step 6: solve a, b in ① we get req. curve

1. Fit the curve of the following $y=ae^{bx}$ form.

x	1	5	7	9	12
y	10	15	12	15	21

sol

Step 1: let $y=ae^{bx}$ – ①

be an exponential curve to be filled by the MLS where a, b are parameters

step 2: Apply $\log_e$ on bs of ①

$$\log_e y = \log(ae^{bx})$$

$$\log_e y = \log_e a + \log_e e^{bx}$$

$$\boxed{\log_e y = \log_e a + bx}$$ $\because \boxed{\log_e e^{bx} = bx}$

$$Y = A + BX \quad -②$$

which is a st line

where $Y = \log_e y$

$A = \log_e a \Rightarrow \boxed{a = e^A}$

$B = b, X = x$

step 3: Normal equations of ②

$$\left.\begin{aligned} nA + B\Sigma x_i &= \Sigma Y_i \\ A\Sigma x_i + B\Sigma x_i^2 &= \Sigma x_i Y_i \end{aligned}\right\} -③$$

Step 4:

$X=x$	y	$Y=\log_e^y$	x^2	xY
1	10	2.3025	1	2.3025
5	15	2.7080	25	13.54
7	12	2.4849	49	17.38
9	15	2.7080	81	24.37
12	21	3.0445	144	36.52
$\Sigma x=34$		$\Sigma Y=13.24$	$\Sigma x^2=300$	$\Sigma xY=94.12$

Step 5: sub all the values in ③ and solve

$5A+34B=13.24$

$34A+300B=94.12$

on solving

$A=2.24,\ B=b=0.059$

$a=e^A$

$a=e^{2.24}$

$a=9.39$ and $b=0.059$

sub in eq ①

$y=(9.39)e^{(0.059)x}$

→ Fitting a power curve of the type $y=ax^b$ (or) $y=ab^x$

working rule

Step 1: let $y=ax^b$ — ①

a, b are parameters.

Step 2: Apply $\log_{10}$ on b.s

$\log_{10}y=\log_{10}a+\log_{10}x^b$

$\log_{10} y = \log_{10} a + b \log_{10} x$

$\boxed{Y = A + BX}$ which is a st-line ②

$\log_{10} y = Y$ | $A = \log_{10} a$ | $B = b$ | $X = \log_{10} x$

step 3: Normal equations of ②

$$\left.\begin{aligned} nA + B\Sigma X &= \Sigma Y \\ A\Sigma X + B\Sigma X^2 &= \Sigma XY \end{aligned}\right\} - ③$$

step ④:

x	$\log_{10} x = X$	y	$Y = \log_{10} y$	X^2	XY
$\Sigma x =$	$\Sigma X =$	$\Sigma y =$	$\Sigma Y =$	$\Sigma X^2 =$	$\Sigma XY =$

step 5: sub all values in ③

and solving we get

A and B subs in eq ②

from which a and b

sub in ①.

H/w

1. Fit a curve of the type $y = ax^b$

x	1	5	10	15
y	12	15	17	20

Correlation and Regression:

Correlation: It is an statistical measure that indicates the degree to which two or more variables fluctuate together.

Types of correlations:

1) Positive correlation: when the increase in one variable is followed by corresponding increase in other variable then the correlation is said to be positive correlation. ex: height ↑ weight

2) Negative correlation: when the increase in one variable is followed by corresponding decrease in other variable then the correlation is said to negative correlation.

ex: Price ↑ demand ↓ ; as speed ↑ time ↓ .

3) Simple and multiple correlation

→ If only 2 variables are considered for correlation analysis then it is simple correlation

→ When more than 2 variables are considered for correlation analysis then it is multiple correlation.

4) Linear and non-linear correlation:

→ Linear correlation is the ~~relate~~ ratio of change b/w the two variables either in same direction or opposite and the graphical representation of the variable is a straight line.

→ If it is not a linear correlation then it is non-linear correlation

Correlation coefficient (or) karl Pearson's Correlation coefficient

→ Correlation coeff is denoted as r and defined as

$$r = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

where $x_i = x_i - \bar{x}$ $\bar{x}$ = mean of x

$y_i = y_i - \bar{y}$

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{y} = \frac{\sum y_i}{n}$$

1) Find the correlation coefficient for the following

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

Sol Given data

Here $n = 10$

$$\bar{x} = \frac{\sum x_i}{n} \qquad \bar{y} = \frac{\sum y_i}{n}$$

$$\bar{x} = \frac{635}{10} \qquad \bar{y} = \frac{610}{10}$$

$$\boxed{\bar{x} = 63.5} \qquad \boxed{\bar{y} = 61}$$

we have karl pearson's coefficient of correlation

$$r = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}}$$

where $x_i = x_i - \bar{x}$

$y_i = y_i - \bar{y}$

Tabular form

x_i	$X_i = x_i - \bar{x}$, $\bar{x} = 63.5$	X_i^2	y_i	$Y_i = y_i - \bar{y}$, $\bar{y} = 61$	Y_i^2	$X_i Y_i$
68	4.5	20.25	62	1	1	4.5
64	0.5	0.25	58	-3	9	-1.5
75	11.5	132.25	68	7	49	30.5
50	-13.5	182.25	45	-16	256	216
64	0.5	0.25	81	20	400	10
80	16.5	272.25	60	-1	1	-16.5
75	11.5	132.25	68	7	49	30.5
40	-23.5	552.25	48	-13	169	305.5
55	-8.5	72.25	50	-11	121	93.5
64	0.5	0.25	70	9	81	4.5

$\boxed{\Sigma X_i^2 = 1364.5}$ $\boxed{\Sigma Y_i^2 = 1136}$ $\Sigma X_i Y_i = 934.85$

$$r = \frac{\cancel{734.88}\ 777.01}{\sqrt{(1364.5)(1136)}}$$

$$\boxed{r = 0.624}$$

→ Rank - correlation coefficient (or) Spearman's rank correlated Coefficient

Type - I	Type - II	Type - III
Ranks are given directly	Ranks are not given	Ranks are not given but repeated

→ This method is based on Rank and is useful in dealing with qualitative analysis such as morality, character, beards, intelligence etc.. Rank correlation coefficient (ρ) denoted and defined as

$$\rho = 1 - \frac{6\left(\sum_{i=1}^{n} d_i^2\right)}{(n^3 - n)}$$

where

ρ = Rank correlation coefficient

d_i = difference of the two ranks

n = no. of paired observations

→ If $\rho = 1$, there is a complete agreement in the order of the rank and the direction of the rank is same.

→ If $\rho \neq 1$, there is a complete disagreement in the order of the rank and the direction of rank is opposite.

→ When the ranks are not given but data is available then we must give the rank in such a way by taking the value as 1 or the lowest values as 1 in both the series.

Rank - correlation coefficient (for equal parts):

$$\rho = 1 - \frac{6\left\{ \sum_{i=1}^{n} d_i^2 + \frac{(m^3-m)}{12} + \frac{(m^3-m)}{12} + \cdots \right\}}{(n^3-n)}$$

where,

m = no. of items whose ranks are common

Here $\frac{1}{12}(m^3-m)$ is correction factor

1) Following are the ranks of 10 students in two subjects Statistics and mathematics. calculate the rank correlation coefficient.

Statistics	1	2	3	4	5	6	7	8	9	10
Maths	2	4	1	5	3	9	7	10	6	8

Sol Given dat

Here $n=10$

∴ Rank correlation coefficient

$$\rho = 1 - \frac{6\left[\sum_{i=1}^{n} d_i^2\right]}{(n^3-n)} \quad -(1)$$

where,

ρ = Rank correlation coefficient

d_i = difference of the two ranks

n = The no. of paired observations

Scanned by CamScanner

Tabular form

Statistics x_i	Maths y_i	$d_i = x_i - y_i$	d_i^2
1	2	-1	1
2	4	-2	4
3	1	2	4
4	5	-1	1
5	3	2	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	3	9
10	8	2	4

$\Sigma d_i^2 = 40$

sub $\Sigma d_i^2 = 40$, $n=10$ in eq ①

$$\rho = 1 - \frac{6[40]}{(10^3-10)}$$

$$= 1 - \frac{6 \times 40}{10(10^2-1)}$$

$$= 1 - \frac{6 \times 40}{10(100-1)}$$

$$= 1 - \frac{24}{99}$$

$$\boxed{\rho = 0.75}$$

2) A random sample of 5 college students is selected and their grades in mathematics & statistics are found to be

Maths	85	60	73	40	90
statistics	93	75	65	50	80

calculate the rank correlation coefficient

Sol Given data

Here $n=5$

we have spearman's rank correlation coefficient for non-repeated

$$\rho = 1 - \frac{\left[\sum_{i=1}^{n} d_i^2\right]}{(n^3-n)}$$

Maths x	Ranks of Maths x_i	Statistics	Rank of statistics y_i	$d_i = x_i - y_i$	d_i^2
85	2	93	1	1	1
60	4	75	3	1	1
73	3	65	4	-1	1
40	5	50	5	0	0
90	1	80	2	-1	1

$\sum d_i^2 = 4$

sub $\sum d_i^2 = 4$ and $n=5$ in the formula

$$\rho = 1 - \frac{6[4]}{5^3-5}$$

$$\rho = 1 - \frac{24}{125-5}$$

$$\rho = 1 - \frac{24}{120}$$

$$\rho = 1 - 0.2$$

$$\boxed{\rho = 0.8}$$

3) From the following data calculate the rank correlation coefficient

x	48	33	40	9	15	16	65	24	16	57
y	13	13	24	6	15	4	20	9	6	19

Sol Given data

Here $n=10$

For repate ranks when we have Rank correlation coefficient

$$\rho = 1-6\left[\frac{\sum_{i=1}^{n} d_i^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \cdots}{n^3-n}\right]$$

where d_i = diff of two ranks

n = no. of obv

m = no. of obv whose ranks are common.

x	Rank of x	y	Rank of y	$d_i = x_i - y_i$	d_i^2
48	3	13	$\frac{2+6}{2}=5.5$	-2.5	6.25
33	5	13	5.5	-0.5	0.25
40	4	24	1	3	9
9	10	6	$\frac{8+9}{2}=8.5$	1.5	2.25
16	$\frac{7+8+9}{3}=8$	15	4	4	16
16	8	4	10	-2	4
65	1	20	2	-1	1
24	6	9	7	1	1
16	8	6	8.5	-0.5	0.25
57	2	19	3	-1	1

$\Sigma d_i^2 = 41$

Correlation factor for x-series

in x-series 16 repeats '3' times $m=3$

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

in y-series 13 repeats 2 times

$$CF = \frac{1}{12}(2^3-2) = 0.5$$

For 6 $CF = 0.5$

Now sub $\Sigma d_i^2 = 41$, CF values 2, 0.5, 0.5 and $n=10$ in observation table.

$$\rho = 1-6\left[\frac{41+2+0.5+0.5}{(10^3-10)}\right]$$

$$\rho = 1-\frac{6(44)}{10(99)}$$

$$\boxed{\rho = 0.73}$$

4) From the following data calculate the rank correlation coefficient.

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	49	50	70

Sol: Given data

$n = 10$

For repeated ranks rank correlation coeff

$$\rho = 1-6\left[\frac{\sum_{i=1}^{n} d_i^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \cdots}{(n^3-n)}\right]$$

where d_i = diff of two ranks

n = no. of obv

m = no. of obv whose ranks are common.

x	Rank of x	y	Rank of y	$d_i = x_i - y_i$	d_i^2
68	4	62	5	-1	1
64	$\frac{5+6}{2} = 5.5$	58	7	-1.5	2.25
75	$\frac{2+3}{2} = 2.5$	68	3.5	-1	1
50	9	45	10	-1	1
64	5.5	81	1	4.5	20.25
80	1	60	6	-5	25
75	2.5	68	3.5	-1	1
40	10	49	9	1	1
55	8	50	8	0	0
64	5.5	70	2	3.5	12.25

Lines of Regression / Regression lines :

The statistical method which help us to estimate the unknown value of one variable from the known value of related variable is called regression

Types of Regression lines

① Regression line y on x

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

where $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$

$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$, n is no. of obv

$r\frac{\sigma_y}{\sigma_x} = b_{yx}$ = Regression coefficient

② Regression line x on y

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

$r\frac{\sigma_x}{\sigma_y} = b_{xy}$ = Regression coe

r = Correlation coeff

$\sigma_x = S.D = \sqrt{\frac{(x_i-\bar{x})^2}{n}}$

$\sigma_y = S.D = \sqrt{\frac{(y_i-\bar{y})^2}{n}}$

Note ① The product of the slope of the two regression lines is given by $b_{xy}b_{yx} = \left(r\frac{\sigma_x}{\sigma_y}\right)\left(\frac{\sigma_y}{\sigma_x}\right)r$

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

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

② The regression lines pass through their means i.e. $(\bar{x}, \bar{y})$

Prob Find the regression lines y on x & x on y from the following data

x	1	2	3	4	5
y	2	5	3	8	7

Sol Given data ↑

$n=5$

$\bar{x} = \frac{\sum x_i}{n} = \frac{15}{5} = 3$

$\bar{y} = \frac{\sum y_i}{n} = \frac{25}{5} = 5$

Correlation coefficient is given by $r = \sum_{i=1}^{n} \frac{X_i Y_i}{\sqrt{\sum X_i^2 \sum Y_i^2}}$ — ①.

$X_i = x_i - \bar{x}$, $Y_i = y_i - \bar{y}$

Tabular form

x_i	$X_i = x_i - \bar{x}$	X_i^2	y_i	$Y_i = y_i - \bar{y}$	Y_i^2	$X_i Y_i$
1	-2	4	2	-3	9	6
2	-1	1	5	0	0	0
3	0	0	3	-2	4	0
4	1	1	8	3	9	3
5	2	4	7	2	4	4
		$\sum X_i^2 = 10$			$\sum Y_i^2 = 26$	$\sum X_i Y_i = 13$

subs in ①

$r = \frac{13}{\sqrt{10 \times 26}}$

$\boxed{r = 0.806}$

Regression line y on x

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

where $r = 0.806$

$\sigma_x = SD = \sqrt{\frac{(\sum x_i - \bar{x})^2}{n}}$

$\sigma_x = \sqrt{\frac{\sum X_i^2}{n}} = \sqrt{\frac{10}{5}} = \sqrt{2}$

$\boxed{\sigma_x = 1.414}$

$\sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} = \sqrt{\frac{\sum Y_i^2}{n}} = \sqrt{\frac{26}{5}}$

$\boxed{\sigma_y = 2.28}$

sub in ②

$y - 5 = (0.806)\left(\frac{2.28}{1.414}\right)(x$

$y - 5 = (1.3)(x - 3)$

$y = (1.3)x - 3(1.3) + 5$

$\boxed{y = (1.3)x + 1.1}$

Scanned by CamScanner

Reg line x on y

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

$$(x-3) = (2.806)\left(\frac{1.414}{2.28}\right)(y-5)$$

$$\boxed{x = (0.5)y + 0.5}$$

Imp Qe If the two regression lines are $8x-10y+66=0$ and $40x-18y=214$. find $\bar{x}$ and $\bar{y}$

Sol: i.e $8\bar{x}-10\bar{y}+66=0$.

$40\bar{x}-18\bar{y}-214=0$.

on solving the above we get

$\bar{x}=13$ and $\bar{y}=17$

Prob The regression lines y on x and x on y are $4x-5y+33=0$ and $20x-9y=107$ respectively. Calculate $\bar{x}$ and $\bar{y}$ and also find correlation coeff r.

Sol: Reg lines passing through the means $(\bar{x},\bar{y})$

$4\bar{x}-5\bar{y}+33=0$.

$20\bar{x}-9\bar{y}-107=0$.

on solving we get $\bar{x}=13$, & $\bar{y}=17$

w.k.T

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

∵ Reg y on x

$4x-5y+33=0$.

$5y = 4x+33$

$$y = \left(\frac{4}{5}\right)x + \left(\frac{33}{5}\right)$$

$$\boxed{b_{yx} = \frac{4}{5}}$$

Scanned by CamScanner

Reg line x on y

$20x - 9y = 107$

$20x = 9y + 107$

$x = \left(\frac{9}{20}\right)y + \frac{107}{20}$

$\boxed{b_{xy} = \frac{9}{20}}$

sub

$r = \sqrt{\left(\frac{4}{5}\right)\left(\frac{9}{20}\right)}$

$\boxed{r = 0.6}$

Hw

→ calculate the coeffcient correlation of the regression line from the following data

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

Also find 'y' values when $x = 6.2$

2m

→ Prove that $r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{(x-y)}^2}{2\sigma_x \sigma_y}$

(or)

Establish relation.

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

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

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

$(z - \bar{z})^2 = [(x - \bar{x}) - (y - \bar{y})]^2$

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

summing up the 'n' terms

$$\sum_{i=1}^{n} (z - \bar{z})^2 = \sum_{i=1}^{n} (x - \bar{x})^2 + \sum_{i=1}^{n} (y - \bar{y})^2 - 2\sum_{i=1}^{n} (x - \bar{x})(y - \bar{y})$$

$$\sum_{i=1}^{n} \frac{(z - \bar{z})^2}{n} = \sum_{i=1}^{n} \frac{(x - \bar{x})^2}{n} + \sum_{i=1}^{n} \frac{(y - \bar{y})^2}{n} - 2\sum_{i=1}^{n} \frac{(x - \bar{x})(y - \bar{y})}{n}$$

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

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

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{(x-y)}^2}{2\sigma_x\sigma_y}$$

$$\boxed{\therefore\ r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y}, \quad \sigma^2 = \frac{\Sigma(x_i-\bar{x})^2}{n}}$$

$$\rightarrow \quad \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n} = r\sigma_x\sigma_y$$

5m

→ If 'θ' is the angle between the two regression lines. Show that $\tan\theta = \frac{(1-r^2)}{r}\left[\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right]$. Explain the significance when $r=0$ and $r=\pm 1$.

Sol We have Regression lines

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

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

$\Rightarrow y = \left(r\frac{\sigma_y}{\sigma_x}\right)x + \bar{y} - \left(r\frac{\sigma_y}{\sigma_x}\right)\bar{x}$ —(1)

Here slope of (1) is $\boxed{m_1 = r\frac{\sigma_y}{\sigma_x}}$

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

(or)

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

$y = \left(\frac{1}{r}\frac{\sigma_y}{\sigma_x}\right)x + \bar{y} - \left(\frac{1}{r}\frac{\sigma_y}{\sigma_x}\right)\bar{x}$ —(2)

slope of eqn (2) is $m_2 = \left(\frac{1}{r}\frac{\sigma_y}{\sigma_x}\right)$ $\boxed{y = m_2x + c}$

Let 'θ' be the angle between two regression lines then

$$\tan\theta = \frac{m_2 - m_1}{(1+m_1m_2)}$$

sub all the values

$$\tan\theta = \frac{\left(\frac{1}{r}\frac{\sigma y}{\sigma x} - r\frac{\sigma y}{\sigma x}\right)}{1+\left(r\frac{\sigma y}{\sigma x}\right)\left(\frac{1}{r}\frac{\sigma y}{\sigma x}\right)}$$

$$\Rightarrow \tan\theta = \frac{\frac{\sigma y}{\sigma x}\left(\frac{1}{r}-r\right)}{1+\frac{\sigma y^2}{\sigma x^2}}$$

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

$$\tan\theta = \frac{\frac{(1-r^2)}{r}}{\sigma x^2+\sigma y^2} \quad \frac{\sigma y}{\cancel{\sigma x}} \times \sigma x^{\cancel{2}}$$

$$\tan\theta = \frac{(1-r^2)\sigma x\,\sigma y}{r(\sigma x^2+\sigma y^2)}$$

Significance

1) When $r=0$, $\tan\theta = \infty$

then $\boxed{\theta = \pi/2}$

The regression lines are perpendicular to each other.

(i.e variables are independent).

Scanned by CamScanner

② when $r = \pm 1$

$\tan\theta = 0$.

$\theta = 0$ (or) π

The Regression lines coincide.

i.e There is a perfect correlation b/w the variables

(***) For two random variables x and y with the same mean, the lines of regression of y on x and x on y are $y = ax + b$, $x = \alpha y + \beta$ respectively. Then S.T $\frac{b}{\beta} = \frac{(1-a)}{(1-\alpha)}$. Also find the common mean

Sol: Given that Reg line y on x

$y = ax + b$ — ①

here Regression coeff = $b_{yx} = r\frac{\sigma_y}{\sigma_x} = a$

Reg line x on y

$x = \alpha y + \beta$ — ②

Reg coeff = $b_{xy} = r\frac{\sigma_x}{\sigma_y} = \alpha$

Let 'm' be the common mean $\boxed{\bar{x} = \bar{y} = m}$

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

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

$y = ax + m(1-a)$ — ③.

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

$x = \alpha y + m(1-\alpha)$ — ④

From eq ① and ③ $b = m(1-a)$ — ⑤

From eq ② and ④ $\beta = m(1-\alpha)$ — ⑥

$$\frac{⑤}{⑥} = \frac{b}{\beta} = \frac{\cancel{\gamma h}(1-a)}{\cancel{\gamma h}(1-\alpha)}$$

$$\boxed{\frac{b}{\beta} = \frac{1-a}{1-\alpha}}$$

$\therefore$ Regression line passing through the mean $(\bar{x}, \bar{y})$

$\therefore \bar{x} = \bar{y} = m$

$y = ax + b \Rightarrow m = am + b$ — (I)

$x = \alpha y + \beta \Rightarrow m = \alpha m + \beta$ — (II)

(I) & (II) $am + b = \alpha m + \beta$

$m(a - \alpha) = \beta - b$

$$\boxed{m = \frac{\beta - b}{a - \alpha}}$$

Scanned by CamScanner

26/7/19
Friday

UNIT - I

PROBABILITY

Random Experiment: Experiment which is conducted under the same condition whose results are known but cannot be predicted is called random experiment.

eg: Tossing a coin, rolling a die etc

Sample space: A set that consists of all possible outcomes of a random experiment is called a sample space and it is denoted by 'S'.

eg: Tossing a coin then $S = \{H, T\}$

Mutually exclusive events:

Events are said to be mutually exclusive if happenin[g] of any one of the event preculates all the other events

i.e $P(A \cap B) = 0$.

Equally likely events:

Events are said to be equally likely if all of them hav[e] equal chance of occurence

Exhaustive events:

A set of events which include all possible events is known as Exhaustive events

i.e $E_1 \cup E_2 \cup \cdots \cup E_n = S$.

Probability:

If the random experiment results 'n' then mutually exclusive, exhaustive, and equally likely events and 'm' of them are favourable for the occurence of an event 'E' then the

Probability of an Event E denoted and defined as

$$P(E) = \frac{m}{n} = \frac{\text{No. of favourable cases}}{\text{Tot no. of cases}}$$

(1) $P(E) + P(E^c) = 1$

$$\boxed{P(E^c) = 1 - P(E)}$$

(2) $0 \le P(E) \le 1$

Additive law of probability

For any two non-disjoint events say A & B

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Note: If A & B are disjoint events (M.E.E)

then $P(A \cap B) = 0$.

$$\Rightarrow \boxed{P(A \cup B) = P(A) + P(B)}$$

(3) **Conditional Probability:**

Let A and B be any two events of a sample space is the Probability of occurrence of event B such that the event A has already occured is denoted by

$$\boxed{P(B/A) = \frac{P(A \cap B)}{P(A)}} \text{ where } P(A) > 0.$$

Note: ||ly we can calculate

$$\boxed{P(A/B) = \frac{P(A \cap B)}{P(B)}} \text{ where } P(B) > 0.$$

Scanned by CamScanner

Multiplication Theorem of Probability

let A and B be two events of sample space S the probability of occurence of both the events simultaneously denoted and defined as

$$P(A\cap B) = P(B/A)\, P(A)$$

where $P(A)>0$,

(OR)

$P(B)>0$

$$P(A\cap B) = P(A/B)\cdot P(B)$$

Note: If A and B are independent events then $P(A\cap B) = P(A)\cdot P(B)$.

Given $P(A)=\frac{1}{4}$, $P(B)=\frac{1}{3}$, $P(A\cup B)=\frac{1}{2}$

Evaluate (i) $P(A/B)$ (ii) $P(B/A)$ (iii) $P(A\cap \bar{B})$ and $P(A/\bar{B})$

So Given data

By additive law:

$$P(A\cup B) = P(A) + P(B) - P(A\cap B)$$

$$\frac{1}{2} = \frac{1}{4}+\frac{1}{3} - P(A\cap B)$$

$$\frac{1}{2} = \frac{3+4}{12} - P(A\cap B)$$

$$P(A\cap B) = \frac{7}{12} - \frac{1}{2} - \frac{7-6}{12} = \frac{1}{12}$$

$$\boxed{P(A\cap B) = \frac{1}{12}}$$

i) $P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{1/12}{1/3}$

$$P(A/B) = \frac{1}{\cancel{12}_4} \times \frac{\cancel{3}}{1}$$

$$\boxed{P(A/B) = \frac{1}{4}}$$

Scanned by CamScanner

ii) $P(B/A) = \frac{P(A\cap B)}{P(A)} = \frac{1/12}{1/4}$

$$\boxed{P(B/A) = \frac{1}{3}}$$

iii) $P(A\cap \bar{B}) = P(A) - P(A\cap B)$

$= \frac{1}{4} - \frac{1}{12}$

$= \frac{3-1}{12} = \frac{2}{12}$

$$\boxed{P(A\cap \bar{B}) = 1/6}$$

iv) $P(A|\bar{B}) = \frac{P(A\cap \bar{B})}{P(\bar{B})}$

$= \frac{P(A\cap \bar{B})}{1-P(B)}$

$= \frac{1/6}{1-1/3} = \frac{1/6}{2/3}$

$= \frac{1}{\cancel{6}_2} \times \frac{\cancel{3}}{2}$

$$\boxed{P(A|\bar{B}) = \frac{1}{4}}$$

8) A problem in Statistics is given to three students A, B & C whose chances of solving it, are $\frac{1}{2}, \frac{1}{3}$ & $\frac{1}{4}$ respectively what is the probability that the problem will be solved

Sol $P(A) = \frac{1}{2}$

$P(B) = \frac{1}{3}$

$P(C) = \frac{1}{4}$

then $P(\bar{A}) = 1 - P(A)$

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

$= \frac{1}{2}$

$P(\bar{B}) = 1 - P(B)$

$= 1 - \frac{1}{3} = \frac{2}{3}$

$P(\bar{C}) = 1 - P(C)$

$= 1 - \frac{1}{4} = 3/4$

P (at least one solve the problem) $= 1 - P(\bar{A}) \cdot P(\bar{B})\, P(\bar{C})$

$= 1 - \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4}$

$= 1 - \frac{1}{4} = 3/4$

3) A dice is thrown twice and the sum of the member appearing is noted be '8'. What is the probability that the number 5 has appeared at least once, when sum is '8'.

So $n(S) = 6^2 = 36$

$S = \begin{Bmatrix} (1,1); (1,2) -- (1,6) \\ (2,1), (2,2) -- (2,6) -- (6,1); (6,2) -- (6,6) \end{Bmatrix}$

let A: The no 5 appears at least once

B: The sum of the numbers is '8'.

$A = \begin{Bmatrix} (1,5); (2,5); (3,5); (4,5); (5,5); (6,5) \\ (5,1); (5,2); (5,3); (5,4); (5,6) \end{Bmatrix}$

$n(A) = 11$

$B = \{(5,3); (3,5); (4,4); (2,6); (6,2)\}$

$n(B) = 5$

$A \cap B = \{(5,3); (3,5)\}$

$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{5/36} = \frac{2}{5}$

$\boxed{P(A/B) = \frac{2}{5}}$

4) A and B thrown alternatively with a pair of di
A wins if he throw 6 before B throws 7 and B wins if he throws 7 before A throws 6. If A

begins find his chances of winning

Sol $n(S)=36$

The sum $6=\{(1,5)\ (5,1)\ (4,4)\ ,(4,2),\ (3,3)\}$

The sum $7=\{(2,5)\ (5,2)\ (1,6)\ ,\ (6,1)\ ,(3,4),\ (4,3)\}$

$P(A)=\frac{5}{36}$ and $P(\bar{A})=1-P(A)=1-\frac{5}{36}$

$$\boxed{P(\bar{A})=\frac{31}{36}}$$

$P(B)=\frac{6}{36}=\frac{1}{6}$ and $P(\bar{B})=1-P(B)=1-\frac{1}{6}$

$$\boxed{P(\bar{B})=\frac{5}{6}}$$

$P(A'$ wins the game$)=P(A)+P(\bar{A})\,P(\bar{B})\cdot P(A)+P(\bar{A})\,P(\bar{B})\,P(\bar{A})\,P(\bar{B})$

$P(A)+ -$

$$=\left(\frac{5}{36}\right)+\left(\frac{31}{36}\right)+\left(\frac{5}{6}\right)\left(\frac{5}{36}\right)+\left(\frac{31}{36}\right)\left(\frac{5}{6}\right)\left(\frac{31}{36}\right)\left(\frac{5}{36}\right)\left(\frac{5}{36}\right)+$$

$$=\frac{5}{36}\left[1+\left(\frac{31}{36}\times\frac{5}{6}\right)+\left(\frac{31}{36}\times\frac{5}{6}\right)^2+--\right]_{G.P}$$

$$=\frac{5}{36}\left[\frac{1}{1-\left(\frac{31}{36}\times\frac{5}{6}\right)}\right]$$

$$=\frac{5}{36}\left[\frac{36\times6}{36\times6-31\times5}\right]$$

$$=\frac{30}{61}$$

5) A speaks truth in 75% and B in 80% of the cases. In what % of cases are they likely to contradict each other in stating the same faced

Sol Given that $P(A)=75\%$ $\quad P(B)=80\%$

$P(A)=\frac{75}{100}=\frac{3}{4}$ $\quad P(B)=\frac{80}{100}=\frac{4}{5}$

$P(\bar{A})=\frac{1}{4}$ $\quad P(\bar{B})=\frac{1}{5}$

P(A & B contradict to each other for the same statement)

$= P(A)P(\bar{B}) + P(\bar{A})P(B)$

$= \frac{3}{4} \times \frac{1}{5} + \frac{1}{4} \times \frac{4}{5}$

$= \frac{3}{20} + \frac{4}{20} = \frac{7}{20} \times 100$

$= 35\%$

The % of contradict to each other is 35%.

Theorem of total probability:

Statement: Let $E_1, E_2, E_3 \cdots E_n$ are 'n' mutually exclusive and exhaustive events of a sample space 'S' such that $P(E_i) > 0$ for all i

∴ Let 'A' be any event in 'S'. Then the total probability of A is given by

$$P(A) = \sum_{i=1}^{n} P(A \cap E_i) = \sum_{i=1}^{n} P(E_i)\, P(A/E_i)$$

Proof: $E_1, E_2, E_3 \cdots E_n$ are mutually exclusive events and also $E_1 \cup E_2 \cup E_3 \cdots \cup E_n = S$

Consider $A = A \cap S$

$A = A \cap [E_1 \cup E_2 \cup E_3 \cup \cdots \cup E_n]$ [∵ By distributive law]

$A = (A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_n)$

$P(A) = P[(A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_n)]$

$P(A) = P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_n)$

$P(A) = P(E_1)P(A/E_1) + P(E_2)P(A/E_2) + \cdots P(E_n)P(A/E_n)$

$$\Rightarrow \boxed{P(A) = \sum_{i=1}^{n} P(E_i)\, P(A/E_i)}$$

$\because$ $A \cap E_1, A \cap E_2 \ldots$ are MEE

$\therefore$ $P(A \cup B) = P(A) + P(B)$

$\because$ By conditional probability $P(A/B) = \dfrac{P(A \cap B)}{P(B)}$

Prob: There are 3 boxes I, II, III. Box-I contains 4 red, 5 blue, and 6 white balls. Box-II contains 3 Red, 4 blue, 5 white balls. Box-III contains 5 red, 10 blue, 5 white balls. One box is choosen at random and one ball is drawn from it. What is the probability that red ball is drawn

Box-I — 4R, 5B, 6W

$P(B\text{-}I) = 1/3$

$P(R/B_1) = 4/15$

$P(R/B_2) = 3/12$

$P(R/B_3) = 5/20$

$P(B\text{-}II) = \frac{1}{3}$ Box-II — 3R, 4B, 5W

$P(B\text{-}III) = \frac{1}{3}$ Box-III — 5R, 10B, 5W

By total probability

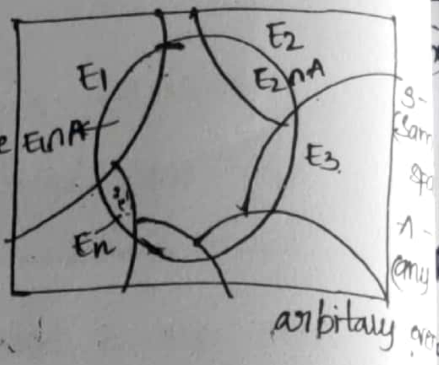
$$P(R) = \sum_{i=1}^{n} P(B_i)\, P(R/B_i)$$

$$P(R) = P(B_1)\, P(R/B_1) + P(B_2)\, P(R/B_2) + P(B_3)\, P(R/B_3)$$

$$P(R) = \frac{1}{3}\,\frac{4}{15} + \frac{1}{3} \times \frac{3}{12} + \frac{1}{3} \times \frac{5}{20}$$

$$P(R) = \frac{23}{90}$$

(**) - State and prove Baye's theorem



<u>Statement</u>: Let $E_1, E_2, E_3, \ldots, E_n$ are 'n' ME and exhaustive events of a sample space 'S' such that $P(E_i) > 0$ (for all i).

If 'A' is any arbitory event in a sample space 'S' and $P(A) > 0$

Then $P(E_i/A) = \dfrac{P(E_i)\,P(A/E_i)}{\sum_{i=1}^{n} P(E_i)P(A/E_i)}$ where $P(A) = \sum_{i=1}^{n} P(E)\,P(A/E_i)$

<u>Proof</u>: $\because E_1, E_2, \ldots E_2$ mutually exclusive events

$$E_i \cap E_j = \phi$$

$$E_1 \cup E_2 \cup \cdots \cup E_n = S \;[\because \text{exhaustive}]$$

$$A = A \cap S$$

$$A = A \cap (E_1 \cup E_2 \cup \cdots \cup E_n) \;[\text{By Distributive law}]$$

$$A = (A \cap E_1) \cup (A \cap E_2) \cup \cdots (A \cap E_n)$$

$$P(A) = P[(A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup P(A \cap E_n)] \;[\because A \cap E_1, A \cap E_2 \text{ are MEE}]$$

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \cdots P(A \cap E_n) \;[\because \text{By condition } P(A/B) = P(A \cap B)/P(B)]$$

$$P(A) = P(E_1)P(A/E_1) + P(E_2)P(A/E_2) + \cdots + P(E_n)\,P(A/E_n)$$

$$P(A) = \sum_{i=1}^{n} P(E_i)\,P(A/E_i) \longrightarrow ①$$

For any E_i

$$P(E_i/A) = \frac{P(E_i \cap A)}{P(A)}$$

$$\Rightarrow \boxed{P(E_i/A) = \frac{P(E_i)P(A/E_i)}{\sum_{i=1}^{n} P(E_i)\,P(A/E_i)}} \quad \text{by ①}$$

Q: An Urn I contains 3W & 4R balls and Urn II contains 5W & 6R balls. One ball is drawn at random from one of the urn and is found to be white. Find the Probability that it was drawn from Urn I.

Let E_1: Urn-I is chosen

E_2: Urn-II is chosen

$P(E_1) = 1/2$ & $P(E_2) = 1/2$

P(a white ball is drawn from U-I) = $P(W|E_1) = \frac{3}{7}$

P(a white ball is drawn from U-II) = $P(W|E_2) = 5/11$

$\frac{1}{2}$ Urn I — 3W, 4R $\quad P(W|E_1) = \frac{3}{7}$

$\frac{1}{2}$ Urn II — 5W, 6R $\quad P(W|E_2) = \frac{5}{11}$

By Bay theorem

$$P(E_1/W) = \frac{P(E_1)P(W|E_1)}{P(E_1)P(W|E_1) + P(E_2)P(W|E_2)}$$

$$P(E_1/W) = \frac{\frac{1}{2}\times\frac{3}{7}}{\frac{1}{2}\times\frac{3}{7} + \frac{1}{2}\times\frac{5}{11}} = \frac{\frac{3}{14}}{\frac{3}{4} + \frac{5}{22}}$$

$$\Rightarrow \boxed{P(E_1/W) = \frac{33}{68}}$$

Prob: In a college, 25% boys and 10% of girls are studying Maths. The girls constitute 60% of student body

i) What is the probability that Maths is being studied

ii) If a student is selected at random and is found to be studying Math. Find the probability that student is a girl ?

student a boy ?

Sol: P(Girls) = $P(G) = 60\% = \frac{60}{100}$

$$\boxed{P(G) = \frac{3}{5}}$$

P(Boy) - $P(B) = 40\% = \frac{40}{100}$

$$\boxed{P(B) = \frac{2}{5}}$$

Also given $P(M/B) = \frac{25}{100} = \frac{1}{4}$

$P(M/G) = \frac{10}{100} = \frac{1}{10}$

i) P(Math is being studied) = $P(M) = P(B)\,P(M/B) + P(G)\cdot P(M/G)$

$$P(M) = \frac{2}{5}\times\frac{1}{4} + \frac{3}{5}\times\frac{1}{10}.$$

$$P(M) = \frac{2}{20} + \frac{3}{50}$$

$$\boxed{P(M) = \frac{4}{25}} \quad \text{— (1)}$$ By Baye's theorem

(ii) $P(G|M) = \dfrac{P(G)\,P(M|G)}{P(G)\,P(M/G) + P(B)\,P(M/B)}$

$$P(G|M) = \frac{\frac{3}{5}\times\frac{1}{10}}{P(M)}$$

$$= \frac{\frac{3}{50}}{\frac{4}{25}}$$

$$= \frac{3}{\cancel{50}_2}\times\frac{\cancel{25}}{4}$$

$$\boxed{P(G/M) = \frac{3}{8}}$$

$$P(B/M) = \frac{P(B)\ P(M/B)}{P(M)}$$

$$= \frac{\frac{2}{5} \times \frac{1}{4}}{\frac{4}{25}}$$

$$= \frac{\cancel{2}}{\cancel{20}_4} \times \frac{\cancel{25}^5}{\cancel{4}_2}$$

$$\boxed{P(B/M) = \frac{5}{8}}$$

2) In a bolt factory machines A, B, C manufactures 25%, 35% and 60% of their total output respectively 5%, 4% & 2% are known to be defective. A bolt is drawn at random and found to be defective. Find the probability that is produced | manufactured by machine B.

Sol: $P(A) = \frac{25}{100} = \frac{1}{4}$

$P(B) = \frac{35}{100} = \frac{7}{20}$

$P(C) = \frac{60}{100} = \frac{3}{5}$

let 'D' be the defective product

$P(D/A) = \frac{5}{100} = \frac{1}{20}$

$P(D/B) = \frac{4}{100} = \frac{1}{25}$

$P(D/C) = \frac{2}{100} = \frac{1}{50}$

Probabity that is manufactured by machine B be $P(B/D)$

$$P(B/D) = \frac{P(B)\ P(D/B)}{P(A)\ P(D/A) + P(B)\ P(D/B) + P(C)\ P(D/C)}$$

$$P(B/D) = \frac{\frac{7}{20} \times \frac{1}{25}}{\frac{1}{4} \times \frac{1}{20} + \frac{7}{20} \times \frac{1}{25} + \frac{3}{5} \times \frac{1}{50}} \Rightarrow \frac{4}{11}$$

1) 5 men out of 100 and 25 women out of 1000 are Colorblind. A person is chosen at random. What is the probability that person is a male? (Assume that males and females are equal)

2) Box A contains 2 white and 4 black balls and Box B contains 5 white and 7 black balls. A ball is transfered from A to B box and a ball is drawn from box B. Find the probability that the ball is white

Sol:

A → 2W, 4B

B → 5W, 7B

$$P(W) = \frac{2}{\cancel{6}_3} \times \frac{\cancel{6}}{13} + \frac{\cancel{4}^2}{\cancel{6}_3} \times \frac{7}{13}$$

$$= \frac{2}{13} + \frac{14}{39}$$

$$= \frac{20}{39}$$

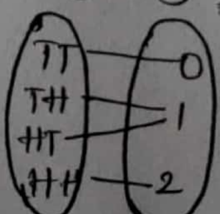
→ Random Variables

Discrete Random Variables | Continuous Random Variables

Random variable:

A random variable x is a real value whose domain is the sample points of the random experiment

Ex: Tossing a coin twice and getting a head a random variable



Continuous Random variable: A random variable x on a sample space 'S' is said to be continuous if it takes all possible values in the interval

ex: The height of the student lies between 5 and 6 feets

i.e $X(S) = \{x \mid 5 \le x \le 6\}$

Probability Mass function:

Let 'x' be a discrete random variable then the function probability is said to be probability mass function.

(i) $P(x_i) > 0$

(ii) $\sum_{i=1}^{n} P(x_i) = 1$

→ **Probability Density function (P.D.F):**

Let x be a continuous random variable where $-\infty < x > \infty$ then the function

$f(x) = P(X \geq x)$ is said to be PDF

if ① $f(x) \geq 0$

② $\int_{-\infty}^{\infty} f(x)\, dx = 1$

Discrete RV	Continuous RV
1) Mean $= \mu = E(X)$ $\mu = \sum_{i=1}^{n} x_i P(x_i)$ 2) $Var(X) = \sigma^2$ $\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 P(x_i)$ (or) $\sigma^2 = E(x^2) - (E(x))^2$ $SD = \sqrt{\sigma^2}$	1) Mean $= \mu = E(X)$ $\mu = \int_{-\infty}^{\infty} x f(x)\, dx$ 2) $Var(X) = \sigma^2$ $\sigma^2 = \int_{-\infty}^{\infty} (x_i - \mu)^2 f(x)\, dx$

Note: ① $E(x^r) = \sum_{i=1}^{n} x^r p(x_r)$

② $P(a \le x \le b) = \int_a^b f(x)\,dx$

③ $E(x^r) = \int_{-\infty}^{\infty} x^r f(x)\,dx$

Prob: A Random variable x has the following probability distributions.

$X = x_i$	0	1	2	3	4
$P(x_i)$	c	$2c$	$2c$	c^2	$5c^2$

i) Find the value of c

ii) Evaluate $P(x<3)$ & $P(0<x<4)$

iii) Determine the distribution factor of x.

Sol i) $\sum_{i=1}^{n} P(x=x_i) = 1$

$6c^2 + 5c = 1$

$6c^2 + 5c - 1 = 0$

$c = 1/6, -1$

$\boxed{c = 1/6}$ $\boxed{\because P(x \ge 0)}$

ii) $P(x<3) = P(x=0) + P(x=1) + P(x=2)$

$= c + 2c + 2c$

$= 5c$

$= 5(1/6)$

$= \frac{5}{6}$

$P(0<x<4) = P(x=1) + P(x=2) + P(x=3)$

$= 2c + 2c + c^2$

$= 4c + c^2$

$= \frac{4}{6 \times 6} + \frac{1}{36} = \frac{25}{36}$

iii)

$x = x_i$	0	1	2	3	4
$P(x_i)$	$1/6$	$2/6$	$2/6$	36	$5/36$
$F(x)$	$1/6$	$2/6$	$5/6$	$31/36$	$36/36 = 1$

1) If x is a discrete random variable and k is a constant then PE $(x+k) = E(x)+k$.

Sol: Req to prove PE $(x+k) = E(x)+k$

By def $E(x) = \sum_{i=1}^{n} x_i P(x_i)$ — ①

Then $E(x+k) = \sum_{i=1}^{n} (x_i+k) P(x_i)$

$= \sum_{i=1}^{n} x_i P(x_i) + k \sum_{i=1}^{n} P(x_i)$

$= E(x) + k(1)$

$\boxed{E(x+k) = E(x)+k}$ $(\because \sum_{i=1}^{n} P(x_i) = 1)$

where k is constant

Note

1) $E(ax+k) = aE(x)+k$

2) $E(x+y) = E(x) + E(y)$

3) $E(xy) = E(x)E(y)$

Scanned by CamScanner

(8) A RVX has the following probability function

x	0	1	2	3	4	5	6	7
$p(x)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

Determine 1) k

2) Mean 3) Varience and SD 4) $P(X<6)$ $P(X\geq 6)$

5) Find distibution function.

Sol:- 1) Given data

$$\because \sum_{i=1}^{n} P(x_i) = 1$$

$$\Rightarrow 0+k+2k+3k+2k+k^2+2k^2+7k^2+k=1$$

$$\Rightarrow 10k^2+9k-1=0$$

$$\Rightarrow 10k^2+9k-1=0$$

$$\Rightarrow k=\frac{1}{10}, -1$$

$$\therefore \boxed{k=1/10} \qquad k\neq -1$$

~~2) Mean~~ 4) $P(X<6) = P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)$

$$= 0+k+2k+2k+3k+k^2$$

$$= k^2+8k$$

$$= \left(\tfrac{1}{10}\right)^2 + 8\left(\tfrac{1}{10}\right)$$

$$\boxed{P(X<6) = 0.81}$$

$$P(X\geq 6) = 1-P(X<5)$$

$$= 1-0.81$$

$$\boxed{P(X\geq 6) = 0.19}$$

) Distribution function

x	$F(x) = P(X \le x_i)$
0	0
1	k
2	$3k$
3	$5k$
4	$8k$
5	$8k+k^2$
6	$8k+3k^2$
7	$9k+10k^2$

$= 9\left(\frac{1}{10}\right) + 10\left(\frac{1}{10}\right)^2$

$= \frac{9}{10} + \frac{1}{10} = 1$

2) Mean $E(X) = \sum_{i=1}^{n} x_i P(x_i)$

$\mu = 0 + k + 4k + 6k + 12k + 5k^2 + 12k^2 + 49k^2 + 7k$

$\mu = 66k^2 + 30k$

$\mu = 66\left(\frac{1}{10}\right)^2 + 30\left(\frac{1}{10}\right)$

$\boxed{\mu = 3.66}$

3) $Var(X) = \sigma^2 = E(X^2) - (E(X))^2$ — ①

Now $E(X^2) = \sum_{i=1}^{n} x_i^2 P(x_i)$

$E(X^2) = 0 + k + 8k + 18k + 48k + 25k^2 + 72k^2 + 343k^2 + 49k$

$= 440k^2 + 124k$

$= 440\left(\frac{1}{100}\right) + 124\left(\frac{1}{10}\right)$

$\boxed{E|x|^2 = 16.8}$

Prob: A player tosses 3 fair coins, he gains ₹500 if 3 head appears, ₹300 if 2 heads appears, ₹100 if 1 head occur on the other hand: He loses ₹1500 if 3 tails occur. find the expected gain of the player.

Sol let x denote the gain of the player

then Range of $x = \{-1500, 100, 300, 500\}$

$n(S) = 2^3 = 8$

sample space:

$S = \{HHH, HHT, HTH, TTT, TTH, THT, THH\}$

P(all 3 heads (getting ₹500)) = 1/8

P (1 heads) (getting ₹100) = 3/8

P (all 3 tails (looses ₹1500) = 1/8

P(2 heads) (getting ₹300) = 3/8

Distribution factor

$x = x_i$	1500	100	300	500
$P(x_i)$	1/8	3/8	3/8	1/8

expected gain of player

$$E(x) = \sum_{i=1}^{n} x_i P(x_i)$$

$$= -1500 \times 1/8 + 100 \times 3/8 + 300 \times 3/8 + 500 \times 1/8$$

$$= -1/8 [-1500 + 300 + 900 + 500]$$

$$= -1/8 [200]$$

$$= 25$$

Scanned by CamScanner

2) A sample of 4 items is selected at random from a box containing 12 items of which 5 are defective. Find the expected no. of defective items.

Sol: Let 'x' denoted the no. of defective items

then $x = \{0, 1, 2, 3, 4\}$

$P(x=0) = \frac{7C_4}{12C_4} = 7/99$

$P(x=1) = \frac{5C_1 \times 7C_3}{12C_4} = 35/99$

$P(x=2) = \frac{5C_2 \times 7C_2}{12C_4} = 42/99$

$P(x=3) = \frac{5C_3 \times 7C_3}{12C_4} = 14/99$

$P(x=4) = \frac{5C_4}{12C_4} = 1/99$

Distribution factor

$x = x_i$	0	1	2	3	4
$P(x_i)$	7/99	35/99	42/99	14/99	1/99

Expected value:

$E(x) = \sum_{i=1}^{n} x_i P(x_i)$

$= 0 + \frac{35}{99} + \frac{2(42)}{99} + \frac{3(14)}{99} + \frac{4(1)}{99}$

$= \frac{1}{99}(35 + 84 + 42 + 4)$

$E(x) = \frac{165}{99}$ //

Scanned by CamScanner

H/w

① Find the mean and variance of the uniform probability distribution given by $f(x) = \frac{1}{n}$ for $x = 1, 2 \ldots$

So

x	1	2	3	4	- - -	n
$f(x)$	$1/n$	$1/n$	$1/n$	$1/n$	- - -	$1/n$

Prob A dice is tossed thrice the success is getting one or six on toss. Find the mean & variance success

Defn Distribution function / cummulative Distribution function

The distribution function $F(x)$ of the discrete random variable is defined by $f(x) = P(X \leq x) = \sum_{j=1}^{x} P(x_j)$ where x is any integer.

Distribution function / CDF:

The distribution function of CRV X is denoted and defined as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

⊛ Note: $F(x) = \int_{-\infty}^{x} F(x)dx$

Then PdF is given

$$f(x) = \frac{d}{dx}[F(x)]$$

2m If x is a CRV, then PT $E(x+a) = E(x)+a$. where a is constant.

So $\therefore E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$ — ① by def

Now $E(x+a) = \int_{-\infty}^{\infty} (x+a) f(x) dx$

$= \int_{-\infty}^{\infty} x f(x) dx + a \int_{-\infty}^{+\infty} f(x) dx$

$= E(x) + a(1)$

$\boxed{E(x+a) = E(x) + a}$ $\because \boxed{\int_{-\infty}^{\infty} f(x) dx = 1}$

(*) If x is a CRV & k is constant then PT

(i) $V(X+k) = V(x)$

(ii) $V(kX) = k^2 V(X)$

$\because V(x) = \sigma^2 = E(X^2) - [E(x)]^2$

$= \int_{-\infty}^{\infty} x^2 f(x) dx - [\int_{-\infty}^{\infty} x f(x) dx]^2$

(i) Then $v(x+k) = [\int_{-\infty}^{\infty} (x+k)^2 f(x) dx - [\int_{-\infty}^{\infty} (x+k) f(x) dx]^2$

$= \int_{-\infty}^{\infty} [x^2 + 2kx + k^2] f(x) dx - [\int_{-\infty}^{\infty} x f(x) dx + k \int_{-\infty}^{\infty} f(x) dx]^2$

$= \int_{-\infty}^{\infty} x^2 f(x) dx + 2k \int_{-\infty}^{\infty} x f(x) dx + k^2 \int_{-\infty}^{\infty} f(x) dx - [\int_{-\infty}^{\infty} x f(x) dx + k \int_{-\infty}^{\infty} f(x) dx]^2$

$= E(x^2) + 2kE(x) + k^2(1) - [E(x) + k(1)]^2$

$V(x+k) = E(x^2) + 2kE(x) + k^2 - [E(x)]^2 - 2kE(x) - k^2$

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

$\boxed{V(x+k) = V(x)}$

(ii) $V(x) = \sigma^2 = E(X^2) - [E(x)]^2$

$= \int_{-\infty}^{\infty} x^2 f(x) dx - [\int_{-\infty}^{\infty} x f(x) dx]^2$

Then

$V(kx) = \int_{-\infty}^{\infty} (kx)^2 f(x) dx - [\int_{-\infty}^{\infty} (kx) f(x) dx]^2$

Scanned by CamScanner

$$= k^2\left[\int_{-\infty}^{\infty} x^2 f(x)\, dx\right] - \left[k \int_{-\infty}^{\infty} x f(x)\, dx\right]^2$$

$$= k^2[E(x^2)] - [kE(x)]^2$$

$$= k^2[E(x^2) - (E(x))^2]$$

$$= k^2 V(X).$$

Prob The PDF of a CRV X is given by $f(x) = kx(2-x)$, $0 \le x \le 2$ find the value of k, mean & variance (x)

Sol Given $f(x) = kx(2-x)$, $0 \le x \le 2$

$\because$ Total probability $= 1$

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

$$\int_0^2 f(x)\, dx = 1$$

$$\int_0^2 kx(2-x)\, dx = 1$$

$$k \int_0^2 [2x - x^2]\, dx = 1$$

$$k\left[2\,x^2/2 - x^3/3\right]_0^2 = 1$$

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

$$k[4 - 8/3] = 1$$

$$k\left[\frac{12-8}{3}\right] = 1$$

$$4k = 3$$

$$k = 3/4$$

$$\text{Mean} = E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$= \int_0^2 x\, kx(2-x)\, dx$$

$$= k\left[\int_0^2 (2x^2 - x^3)\, dx\right.$$

$= k\left[2\cdot x^3/3 - x^4/4\right]_0^2$

$= k\left[2\cdot 2^3/3 - 2^4/4\right]$

$= k\left[16/3 - 16/4\right]$

$= k\left[\frac{64-48}{12}\right] = \frac{3}{4}\left[\frac{4}{3}\right] = 1$

$\boxed{E(X) = \mu = 1}$

$\therefore V(X) = E(X^2) - [E(X)]^2$

$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - [1]^2$

$= \int_0^2 x^2 kx(2-x)\,dx - 1.$

$= k\left[\int_0^2 (2x^3 - x^4)\,dx - 1\right.$

$= k\left[2x^4/4 - x^5/5\right]_0^2 = \frac{6}{5} - 1 = \frac{1}{5}$

2) A CRV x has the PDF $f(x) = \begin{cases} a + bx, & 0 \le x \le 1 \\ 0, & \text{else where} \end{cases}$

If the mean of the distribution is $\frac{1}{3}$. Find a & b.

Sol Given PDF

$f(x) = \begin{cases} a + bx, & 0 \le x \le 1 \\ 0, & \text{else where} \end{cases}$

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

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

$\int_0^1 (a + bx)\,dx = 1$

$a(x)_0^1 + b(x^2/2)_0^1 = 1$

$a + b/2(1-0) = 1$

$a + b/2 = 1$

$2a+b=2$ —①

Mean $= \frac{1}{3}$

$E(X) = \frac{1}{3}$

$\int_{-\infty}^{\infty} x f(x)\, dx = \frac{1}{3}$

$\int_0^1 x(a+bx)\, dx = \frac{1}{3}$

$\int_0^1 (ax + bx^2)\, dx = 1/3.$

$\left[ax^2/2 + \frac{bx^3}{3} \right]_0^1 = \frac{1}{3}$

$a/2 + b/3 = 1/3$

$3a+2b = 6/3$

$3a+2b=2$ —②

By solving ① & ②.

$a=2,\ b=-2$ //

Q) A RVX has the density function $f(x) = \frac{k}{x^2+1}$, $-\infty < x < \infty$ find (i) k (ii) The distribution function, c

Sol Given $f(x) = \frac{k}{x^2+1}$, $-\infty < x < \infty$

$\because$ Total probability $= 1$

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

$\int_{-\infty}^{\infty} \frac{k}{x^2+1}\, dx = 1$

$k \int_{-\infty}^{\infty} \frac{1}{x^2+1}\, dx = 1$

$k\,(\tan^{-1} x)_{-\infty}^{\infty} = 1$

$k\,[\tan^{-1}\infty - \tan^{-1}(-\infty)] = 1$

$k\,(\pi/2 + (\frac{\pi}{2})) = 1$

$k\pi = 1$

$\boxed{k = 1/\pi}$

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

$F(x) = \int_{-\infty}^{x} \frac{k}{x^2+1}\,dx$

$F(x) = \frac{1}{\pi} \int_{-\infty}^{x} \frac{1}{x^2+1}\,dx$

$= \frac{1}{\pi}\left[\tan^{-1}(x)\right]_{-\infty}^{x}$

$= \frac{1}{\pi}\left[\tan^{-1}(x) + \tan^{-1}(\infty)\right]$

$F(x) = \frac{1}{\pi}\left[\tan^{-1}x + \pi/2\right]$

$\boxed{F(x) = \frac{1}{\pi}\cot^{-1}(x)}$ //

⑥ If PDF $f(x) = \begin{cases} 0, & x<0 \\ ax, & 0 \le x < 2 \\ (4-x)a; & 2 \le x \le 4 \\ 0, & x>4 \end{cases}$

Find (i) a (ii) $P(x > 2.5)$

Sol (i) Given $f(x) = \begin{cases} 0, & x<0 \\ ax, & 0 \le x < 2 \\ (4-x)a; & 2 \le x \le 4 \\ 0, & x>4 \end{cases}$

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

$\Rightarrow \int_{0}^{2} ax\,dx + \int_{2}^{4} (4-x)a\,dx = 1$

$\Rightarrow \frac{a}{2}(x^2)_0^2 + a\left(4x - \frac{x^2}{2}\right)_2^4$

$= \frac{a}{2}(4) + a[[4(4) - \frac{16}{2}] - (4(2) - \frac{4}{2})]$

$= 2a + a[32 - \frac{16}{2} - (\frac{16-4}{2})]$

$= 2a + a[\frac{16}{2} - \frac{12}{2}]$

$= 2a + 2a = 1$

$= 4a = 1$

$\Rightarrow \boxed{a = \frac{1}{4}}$ //

(ii) $P(x > 2.5) = \int_2^4 f(x)dx$

$= \int_{2.5}^4 (4-x)a\,dx$

$= a[4(x)_{2.5}^4 - \frac{1}{2}(x^2)_{2.5}^4]$

$= \frac{1}{4}[4(4-2.5) - \frac{1}{2}(16-6.25)]$

$= \frac{1}{4}[4(1.5) - \frac{1}{2}(9.75)]$

$= \frac{1}{4}(6 - 4.875)$

$= \frac{1.125}{4}$

$= 0.281 = \frac{9}{32}$

Q) If a Discrete function

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{3}, & 0 \le x < 1 \\ 1/3, & 1 \le x < 2 \\ x/6, & 2 \le x < 6 \\ 1, & x \ge 6 \end{cases}$$

Then find a Pdf $f(x)$

So Given CDF

$PdF = \frac{d}{dx}[F(x)]$

$$f(x) = \begin{cases} 0, & x<0 \\ 1/3, & 0 \le x < 1 \\ 0, & 1 \le x < 2 \\ 1/6, & 2 \le x < 6 \\ 0, & x \ge 6 \end{cases}$$

H/w ① A CRV has the pdf $f(x) = \begin{cases} kxe^{-\lambda x}, & \text{for } x \ge 0, \lambda > 0 \\ 0, & \text{otherwise} \end{cases}$

Determine (i) k (ii) Mean (iii) Variance (x).

② Find Mean, mode & median of a CRV X

If $f(x) = \begin{cases} \frac{1}{2} \sin x, & \text{for } 0 \le x \le \pi \\ 0, & \text{elsewhere.} \end{cases}$

UNIT - II

DISCRETE PROBABILITY DISTRIBUTION

→ **Binomial Distribution**

Def: A random variable x has a binomial distribution if it assumes only non-negative value. and its probability density function is given by

$$P(X=r) = P(r) = \begin{cases} n_{c_r} p^r q^{n-r}, & \text{when } r=0,1,2,\ldots n \\ 0, & \text{otherwise} \end{cases}$$

where p - Probability of success
q - probability of failure

Note: $P(X=x) = P(x) = b(x, n, p)$

$$= \begin{cases} n_{c_x} p^x q^{n-x}, & x=0,1,2 \ldots n \\ 0, & \text{else} \end{cases}$$

$\boxed{p+q=1}$

$\boxed{q=1-p}$

→ Constants of Binomial distribution:

① Mean of Binomial distribution:

let BD $P(X=r) = P(r) = \begin{cases} n_{c_r} p^r q^{n-r}, & r=0,1,2 \ldots n \\ 0, & \text{otherwise} \end{cases}$ —① $\boxed{p+q=1}$

mean $= \mu$

$\mu = E(X)$

$$\mu = \sum_{r=0}^{n} r P(r)$$

$$\mu = \sum_{r=0}^{n} r\, n_{c_r} p^r q^{n-r}$$

$$\mu = 0 + 1 \cdot n_{c_1} p^1 q^{n-1} + 2 n_{c_2} p^2 q^{n-2} + 3 n_{c_3} p^3 q^{n-3} + \ldots + n\, n_{c_n} \ldots$$

$$\therefore n_{c_r} = \frac{n!}{(n-r)!\, r!}$$

$$\mu = 1 \cdot npq^{n-1} + \frac{2\,n!}{(n-2)!\,2!} p^2 q^{n-2} + \frac{3\,n!}{(n-3)!\,3!} p^3 q^{n-3} + \ldots + np^n$$

$$\mu = npq^{n-1} + \frac{2n(n-1)(n-2)!}{(n-2)!\,2!} p^2 q^{n-2} + \frac{3n(n-1)(n-2)(n-3)!}{(n-3)!\,3\times 2} p^3 q^{n-3} + \dots + np^n$$

$$\mu = np[q^{n-1} + (n-1)pq^{n-2} + (n-1)(n-2)\,pq^{n-3} + \dots p^{n-1}]$$

$$\mu = np[q+p]^{n-1}$$

$$\mu = np(1)^{n-1} \quad \boxed{\because p+q=1}$$

$$\boxed{\mu = np}$$

mean of BD. $\mu = np$

Variance of a Binomial Distribution

$$V(X) = E(X^2) - [E(X)]^2 \qquad P(X=r) = p(r) = \begin{cases} n_{C_r} p^r q^{n-r}, & r = 0 \text{ to } n \\ 0, & 0 \text{ then } p+q=1 \end{cases}$$

$$V(X) = \sum_{r=0}^{n} r^2 p(r) - \mu^2$$

$$V(X) = \sum_{r=0}^{n} (r^2 + r - r)\, p(r) - \mu^2$$

$$V(X) = \sum_{r=0}^{n} r(r-1)p(r) + \sum_{r=0}^{n} rp(r) - \mu^2$$

$$V(X) = \sum_{r=0}^{n} r(r-1)\, p(r) + \mu - \mu^2$$

Variance of a Binomial Distribution

$$\therefore \text{BD is } p(r) = \begin{cases} n_{C_r} p^r q^{n-r} & r = 0, 1, 2, \dots n \\ 0, & \text{otherwise} \end{cases}$$

$\because p+q=1$

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

$$V(X) = \sum_{r=0}^{n} r^2 p(r) - \mu^2$$

$$V(X) = \sum_{r=0}^{n} (r^2 - r + r)\, p(r) - \mu^2$$

$$V(X) = \sum_{r=0}^{n} (r^2 - r)p(r) + \sum_{r=0}^{n} rp(r) - \mu^2$$

$$= \sum_{r=0}^{n} r(r-1)\, p(r) + \mu - \mu^2$$

$$V(X) = \sum_{r=0}^{n} r(r-1)\, {}^nC_r\, p^r q^{n-r} + \mu - \mu^2$$

$$= 0 + 0 + 2(2-1)\, {}^nC_2\, p^2 q^{n-2} + 3(3-1)\, {}^nC_3\, p^3 q^{n-3} + \cdots$$

$$+ n(n-1)\, {}^nC_n\, p^n q^0 + \mu - \mu^2$$

$$V(X) = 2 \cdot \frac{n!}{(n-2)!\,2!} p^2 q^{n-2} + \frac{6\, n!}{(n-3)!\,3!} p^3 q^{n-3} + \cdots + n(n-1) p^n + \mu - \mu^2$$

$$V(X) = \frac{n(n-1)(n-2)!}{(n-2)!} p^2 q^{n-2} + \frac{n(n-1)(n-2)(n-3)!}{(n-3)!} p^3 q^{n-3} + \cdots + n(n-1) p^n + \mu$$

$$V(X) = n(n-1)p^2 \left[q^{n-2} + (n-2) p\, q^{n-3} + \cdots + p^{n-2}\right] + \mu - \mu^2$$

By Binomial theorem

$$V(X) = n(n-1)p^2 [q+p]^{n-2} + \mu - \mu^2$$

$$V(X) = n(n-1)p^2 [1]^{n-2} + \mu - \mu^2$$

$$V(X) = n(n-1)p^2 (1) + \mu - \mu^2$$

$$V(X) = n^2p^2 - np^2 + np - (np)^2 \qquad \boxed{\because \text{mean} = \mu = np}$$

$$V(X) = np(1-p) \qquad \boxed{\because p+q=1}$$

$$\boxed{V(X) = npq} \qquad \boxed{{}^nC_r = \frac{n!}{(n-r)!\,r!}}$$

Hence variance of a binomial distribution $= npq$

<u>Mode of the Binomial Distribution</u>

Mode of the Binomial distribution is the value of x at which $p(x)$ has maximum value.

$$\text{Mode} = \begin{cases} \text{integral part } (n+1)p, & \text{If } (n+1)p \text{ is not integer} \\ (n+1)p \text{ and } (n+1)p-1, & \text{If } (n+1)p \text{ is an integer} \end{cases}$$

<u>Note</u>: <u>Recurrence Relation for the Binomial distribution</u>

$$p(r+1) = \frac{(n-r)p}{(r+1)q}\, p(r)$$

Binomial Frequency Distribution

- The possible no. of successes and their frequency distribution is called a Binomial frequency distribution.

1) A fair coin is tossed six times. Find the probability of getting four heads.

Sol: p = probability of getting head

$p = 1/2$, $q = 1 - 1/2 = 1/2$.

$n = 6$, $r = 4$

$\because p(r) = n_{C_r} p^r q^{n-r}$ where $r = 0, 1, 2 \ldots n$

$P(\text{four head}) = p(4) = 6_{C_4} (1/2)^4 (1/2)^{6-4}$

$= 6_{C_4} (1/2)^6$

$$p(4) = \frac{\cancel{6}^{3} \times 5 \times \cancel{4} \times \cancel{3}}{\cancel{4} \times \cancel{3} \times \cancel{2} \times 1} \times \left(\frac{1}{2^6}\right)$$

$p(4) = \frac{15}{64}$ //

2) Determine the probability of getting the sum 6 exactly 3 times in 7 thrown with a pair of fair dice.

Sol: $\because$ A pair of fair dice is thrown

$n(S) = 6^2 = 36$

The no. of favorable cases to occur sum 6

are $= \{(1,5)\ (5,1)\ (2,4)\ (4,2)\ (3,3)\} = 5$

$P(\text{getting a sum } 6) = \frac{5}{36}$

$p = 5/36$ $\qquad \because q = 1 - p$

$\therefore n = 7$ $\qquad \boxed{q = 1 - \frac{5}{36}}$

$r = 3$ $\qquad \boxed{q = 31/36}$

$P(r) = {}^{n}C_{r}\, p^{r} q^{n-r}$, $n = 0, 1, 2, 3 \ldots n$

$= {}^{7}C_{3}\, p^{3} q^{4}$

$= \frac{7 \times 6 \times 5}{3 \times 2 \times 1} p^{3} q^{4}$

$= 35 \left(\frac{5}{36}\right)^{3} \left(\frac{31}{36}\right)^{4}$

$= \frac{35 \times 5^{3} \times (31)^{4}}{(36)^{7}}$

$\simeq 0.057$

3) In eight throws of a die 5 or 6 is conditional success, find mean & SD.

Sol $n(S) = 6$

$S = \{1, 2, 3, 4, 5, 6\}$

$P(\text{getting } 5) = \frac{1}{6}$

$P(\text{getting } 6) = \frac{1}{6}$

$P(\text{success } 5 \text{ or } 6) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

$q = 1 - p$

$q = 1 - \frac{1}{3}$

$\boxed{q = \frac{2}{3}}$

$\boxed{n = 8}$

mean of a BD $= np$

$\mu = np$

$\mu = 8\left(\frac{1}{3}\right)$

$\boxed{\mu = \frac{8}{3}}$

$V(X) = npq$

$\sigma^{2} = 8\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)$

$= \frac{16}{9}$

$\sigma = \sqrt{\frac{16}{9}}$

$\sigma = \frac{4}{3}$,,

) In 256 sets of 12 tosses of a coin, ~~In~~ how many cases one can expect 8 heads and 4 tails.

So $\because P(r) = n_{C_r}\, p^r q^{n-r}$

$P(r) = 12_{C_8} \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^4$

$P(8) = \frac{12!}{4!\,8!} \left(\frac{1}{2^{12}}\right)$

$= \frac{11 \times 15 \times 9}{2^{12}}$

$= 0.120 \times 256$

$= 30.93$

$\simeq 31$,,

5) The mean and variance of a binomial distribution are 4 and 4/3 respectively. Find $P(X \geq 1)$

So Given that mean of B.D $= 4$

$np = 4$ — ①

$V(X) = \frac{4}{3}$

$npq = 4/3$ — ①

$\frac{②}{①} = \frac{npq}{np} = \frac{\frac{4}{3}}{\frac{4}{1}}$

$\Rightarrow \boxed{q = \frac{1}{3}}$

$p = 1 - q$

$= 1 - \frac{1}{3}$

$\boxed{p = \frac{2}{3}}$

$np = 4$

$n(\frac{2}{3}) = 4$

$n = \cancel{\frac{12}{2}}\ 6$

$\boxed{n=6}$

$\therefore\ P(X \geq 1) = 1 - P(X \leq 1)$

$= 1 - P(X=0)$

$= 1 - P(0) \qquad \boxed{r=0}$

$= 1 - 6_{C_0} p^0 q^6$

$= 1 - 1 \times 1 \times (\frac{1}{3})^6$

$= 0.998$

5) Fit a BD to the following frequency

x	0	1	2	3	4	5	6
f	13	25	52	58	32	16	4

Sol Given data

$\boxed{n=6}$

$N = \Sigma f_i$

$N = 13 + 25 + 52 + 58 + 32 + 16 + 4$

$N = 200$

$\text{Mean} = \frac{\Sigma x_i f_i}{N}$

$= \frac{0 + 25 + 104 + 174 + 128 + 80 + 24}{200}$

$\boxed{\text{Mean} = 2.675}$

$\text{Mean} = np$

$np = 2.675$

$p = \frac{2.675}{6}$

$p = 0.445$

$q = 1-p$

$q = 0.555$

Hence BD is $= N(q+p)^n$

$= 200\ (0.55 + 0.445)^6$

$= 200\{ 6c_0\ (0.555)^6 + 6c_1\ (0.555)^5\ (0.445)^1$

$+ 6c_2\ (0.555)^4\ (0.445)^2 + 6c_3$

$\Rightarrow (0.555)^3\ (0.445)^3 + 6c_4\ (0.55)^2\ (0.445)^4 + 6c_5\ (0.555)^1$

$(0.445)^5 + 6c_6\ (0.445)^6$

$= 200\ (0.029 + 0.141 + 0.281 + 0.207 + 0.177 + 0.058 + 0.076)$

$= 199.1$

:)

x	0	1	2	3	4	5	6
f	6	28	56	60	36	12	2

Fit a binomial distribution to the following data

x	0	1	2	3	4
f(x)	20	62	46	10	2

So Given data

Here n = 4

x_i	$f(x_i)$	$x_i f(x_i)$
0	30	0
1	62	62
2	46	92
3	10	30
4	2	8
	$\Sigma f(x_i) = 150$	$\Sigma x_i f(x_i) = 192$

$$\text{Mean} = \frac{\Sigma x_i f(x_i)}{\Sigma f(x_i)} = \frac{192}{150}$$

Here $\Sigma f(x_i) = N = 150$

Mean $= 1.28$

$\therefore$ Binomial distribution

$\mu = \text{Mean} = 1.28$

Mean $= np$

$\mu = np$

$$p = \frac{\mu}{n}$$

$$p = \frac{1.28}{4}$$

$$\boxed{p = 0.32}$$

$q = 1 - p$

$q = 1 - 0.32$

$$\boxed{q = 0.68}$$

Hence Binomial Distribution is

$BD = N(p+q)^n$

$= 150(0.32 + 0.68)$

$\therefore (x+y)^n = n_{c_0} x^n + n_{c_1} x^{n-1} y + n_{c_2} x^{n-2} y^2 + \dots n_{c_n} y^n$

$= 100\,(4_{C_0}(0.32)^4 + 4C_1\,(0.32)^3(0.68) + 4C_2\,(0.32)^2$

$(0.68)^2 + 4C_3\,(0.32)(0.68)^3 + 4C_4\,\,(0.68)^4$

$= 100[(0.32)^4 + 4(0.32)^3(0.68) + 6(0.32)^2(0.68)^2] + 4(0.32)$

$(0.68)^3 + (0.68)^4]$

Prob Fit a binomial distribution to the following data

x	0	1	2	3	4
$f(x)$	30	62	46	10	2

Sol

x	Expected freq	Theoritical form
0	30	32
1	162	60
2	46	43
3	10	13
4	2	2

Q) Two players A and B play tennis game. Their chances of winning a game are in the ratio 3:2. Find A's chance of winning at least two games out of four games played

Sol Ratio of A & B winning the game

$A:B = 3:2$ Here $n=4$

$p =$ Probability of A, winning the game

$p = 3/5\,,\quad q = 2/5$

$P(X=r) = n_{C_r}\,p^r q^{n-r} \qquad r = 0,1,2,3,4$

$= 4_{C_r}(3/5)^r(2/5)^{4-r}$

$P(X \geq 2) = 1 - P(X<2)$

$= 1 - [P(X=0) + P(X=1)]$

$= 1 - [4_{C_0}(3/5)^0(2/5)^{4-0} + 4C_1\,(3/5)^1(2/5)^{4-1}]$

$= 0.82$

Possion distribution:

A random variable 'x' is said to follow the possion distribution if its probability mass function is given by

$$\boxed{P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}} \quad x = 0,1,2,3 \ldots$$

where $\lambda = np$ is the parameter of the distribution.

Note 1: Possion distribution is a discrete distribution as random variable 'x' is the discrete value.

2. Possion distribution is a limiting case of binomial distribution as 'n' goes ∞ & p goes to 0.

$n \to \infty$ $\quad$ $p \to 0$.

3. If 'n' trials constitutes an experiment which is repeated 'N' times the frequency function of the possion distribution is given by

$$\boxed{f(x) = NP(x)}$$

Mean of a Poisson distribution:

Proof:

We've $PD = P(X=x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ —①

where $x = 0,1,2,3$

Mean $= E(X)$

$\mu = \sum x\, p(x)$

$$\mu = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!}$$

$$\mu = e^{-\lambda} \frac{\left[\sum_{x=0}^{\infty} x \lambda^x\right]}{x!}$$

Scanned by CamScanner

$$\mu = e^{-\lambda}\left[\sum_{x=0}^{\infty} \frac{x\lambda^x}{x(x-1)!}\right]$$

$$\mu = e^{-\lambda}\left[\sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!}\right]$$

$$\mu = e^{-\lambda}\left[\frac{\lambda^1}{0!} + \frac{\lambda^2}{1!} + \frac{\lambda^3}{2!} + \cdots\right]$$

$$\mu = e^{-\lambda}\lambda\left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\right]$$

$$\mu = e^{-\lambda} \cdot \lambda e^{\lambda}$$

$$\mu = \lambda e^{\lambda - \lambda}$$

$$\mu = \lambda \cdot e^0$$

$$\boxed{\mu = \lambda} \quad \boxed{\text{Mean of PD} = \lambda}$$

$$e^{-x} = 1 - \frac{x}{1!} + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots$$

$$e^{x} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots$$

Variance of a Poisson Distribution:

Proof:

P.D $\quad P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}, \; x = 0, 1, 2, 3 \ldots$

where 'λ' mean of PD

$$Var(x) = E(X^2) - (E(X))^2 \quad \text{--- (1)}$$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 p(x)$$

$$= \sum_{x=0}^{\infty} ((x^2 - x) + x)\, p(x)$$

$$E(X^2) = \sum_{x=0}^{\infty} (x^2 - x)\, p(x) + \sum_{x=0}^{\infty} x\, p(x)$$

$$= \sum_{x=0}^{\infty} (x^2 - x) \frac{e^{-\lambda}\lambda^x}{x!} + \lambda$$

$$= \sum_{x=0}^{\infty} \frac{\cancel{x(x-1)}\, e^{-\lambda}\lambda^x}{\cancel{x(x-1)}(x-2)!} + \lambda$$

$$E(x^2) = \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-2)!} + \lambda$$

$$= \frac{e^{-\lambda}\lambda^2}{0!} + \frac{e^{-\lambda}\lambda^3}{1!} + \frac{e^{-\lambda}\lambda^4}{2!} + \dots + \lambda$$

$$= e^{-\lambda}\lambda^2 \left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots\right] + \lambda$$

$$= e^{-\lambda}\lambda^2 e^{\lambda} + \lambda$$

$$\therefore \boxed{E(x^2) = \lambda^2 + \lambda}$$

from ①

$$\text{Var}(x) = E(x^2) - (E(x))^2 \qquad \because \boxed{\mu = E(x) = \lambda}$$

$$= \lambda^2 + \lambda - \lambda^2$$

$$\boxed{\text{Var}(x) = \sigma^2 = \lambda}$$

Note: For poisson Distribution

$$\boxed{\text{Mean} = \text{Variance} = \lambda}$$

<u>Mode</u> <u>of</u> <u>a</u> <u>Poisson</u> <u>Distribution</u>:

Mode is the value of x for which $P(x)$ is maximum

case (1): If 'λ' is an integer

Then, λ, $(\lambda-1)$ are two modes of a P.D

case (2): If 'λ' is not an int

Then the integral part of λ is the mode

Note: Recurrence relⁿ of P.Dⁿ

we've PD $P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2 \dots$

Replace x by $(x+1)$

$$P(x+1) = \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!}$$

$$P(x+1) = \frac{e^{-\lambda}\lambda^x \cdot \lambda}{(x+1)(x!)}$$

Scanned by CamScanner

$$P(x+1) = \frac{\lambda}{x+1}\left[\frac{e^{-\lambda}\lambda^x}{x!}\right]$$

$$P(x+1) = \left(\frac{\lambda}{x+1}\right) P(x) \longrightarrow ③ \text{ [using ①]}$$

Now replace x by $(x-1)$ in ②

$$P(x-1+1) = \frac{\lambda}{(x-1+1)} P(x-1)$$

$$P(x) = \frac{\lambda}{x} P(x-1)$$

1) In a Poisson Distribution if $3P(X=2) = P(X=4)$ then find Probability at $x=3$.

Sol Given

we've PD

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}, x=0,1,2$$

$$3P(X=2) = P(X=4)$$

$$\frac{3\cancel{e^{-\lambda}}\cancel{\lambda^2}}{\cancel{2!}} = 1 \frac{\cancel{e^{-\lambda}}\lambda^{4}{}^{2}}{4!\,2}$$

$$\frac{3}{2} = \frac{\lambda^2}{24}$$

$$\lambda^2 = \frac{3 \times \cancel{24}^{12}}{\cancel{2}}$$

$$\lambda^2 = 36$$

$$\lambda = \pm 6$$

$$\because \lambda > 0$$

so $\lambda = 6$

$$2.\ P(X=3) = \frac{e^{-\lambda}\lambda^3}{3!}$$

$$= \frac{e^{-6}(6)^{\cancel{3}2}}{\cancel{6}}$$

$$= e^{-6} \cdot 36$$

$$= 0.089$$

Scanned by CamScanner

2) The avg no of accidents on one day on a national highway is 1.8. Determine the probability that no. of accidents are

① At most one ② At least one

So Given avg no of accidents on 1 day.

$$\boxed{\mu = \lambda = 1.8}$$

we've PD $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2 -$

(i) P (at most one) = $P(x \le 1)$

$= P(x=0) + P(x=1)$

$= \frac{e^{-\lambda}\lambda^x}{x!} + \frac{e^{-\lambda}\lambda^x}{x!}$

Substitute all values

$$P(X \le 1) = \frac{e^{-1.8}(1.8)^0}{0!} + \frac{e^{-1.8}(1.8)^1}{1!}$$

$$= \frac{e^{1.8} \times 1}{1} + \frac{e^{-1.8}(1.8)}{1}$$

$$\boxed{P(X \le 1) = 0.462}$$

2) P(at least one) = $P(X \ge 1) = 1 - P(X=0)$

$$P(X \ge 1) = 1 - \frac{e^{-1.8}(1.8)^0}{0!}$$

$$= 1 - 0.165 = 0.83$$

3) 2% of items of a factory are defective the items are packed in boxes. What is the probability that there will be

(i) 2 defective items

(ii) atleast 3 defectives item in a box of 100 items

So let P(success) = P(defective items)

$p = 2\%$

$p = \frac{2}{100} = 0.02$ $\boxed{n=100}$

$\because \lambda = np$

$\lambda = 100 \times \frac{2}{100}$

$\boxed{\lambda = 2}$

$\therefore$ we've PD $P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, x = 0, 1, --$

(i) P(2 defective items) = $P(X=2)$

$= \frac{e^{-2}(2)^2}{2!} = 0.27$

(ii) P(at least 3 defective items) = $P(X \geq 3)$

$= 1 - P(X=0) + P(X=1) + P(X=2)$

$= 1 - \left[\frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-2}2^2}{2!}\right]$

$= 1 - [0.135 + 0.27 + 0.27]$

$\Rightarrow \boxed{P(X \geq 3) = 0.32}$

4) Fit a Poisson Distribution to the following data

x	0	1	2	3	4
$f(x)$	122	60	15	2	1

Sol: Given data ?

Fitting PD is nothing but calculating expected / theoritical frequency

$$\boxed{f(x) = Np(x)}$$

where $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2, 3, 4 \ldots$

$N = \Sigma f(x)$

$N = 122 + 60 + 15 + 2 + 1$

$$\boxed{N = 200}$$

To find Mean of PD

$\lambda = \frac{1}{N} \Sigma x f(x)$

$\lambda = \frac{1}{200} [(0 \times 122) + (1 \times 60) + 2(15) + 3(2) + 4(1)]$

$\lambda = \frac{1}{200} (0 + 60 + 30 + 6 + 4)$

$\lambda = \frac{100}{200}$

$$\boxed{\lambda = 0.5}$$

Expected frequency are given by $= Np(x)$

$= 200 \frac{e^{-\lambda}\lambda^x}{x!}$

$f(x=0) = 200 e^{-0.5} \frac{(0.5)^0}{0!}$

$= 121.3 \simeq 121$

$f(x=1) = \frac{200 e^{-0.5}}{1!} = 60.6 \cong 61$

$f(x=2) = \frac{200\, e^{-0.5} (0.5)^2}{2!} = 15.1 \cong 15$

$f(x=3) = \frac{200\, e^{-0.5} (0.5)^3}{3!} = 2.5 \cong 2$

$f(x=4) = \frac{200\, e^{-0.5} (0.5)^4}{4!} = 0.315 = 0.$

x	0	1	2	3	4
Obv	122	60	15	2	1
Exp	121	61	15	2	0

H/w - Fit a PD to the following data

x	0	1	2	3	4
f(x)	109	65	22	3	1

H/w If x is a Poisson variable ST $3P(X=4) = \frac{1}{2}P(X=2) + P(X=0)$

Find ① λ ② $P(X \le 2)$

⊛ Moment Generating Function (mgf)

The mgf of a random variable v, about the origin having the probability function f(x) is and is denoted by $M_x(t)$ and defined as

$$M_x(t) = \begin{cases} \sum e^{tx} p(x) & \text{in DRV} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\,dx & \text{in CRV} \end{cases}$$

Note : $\boxed{M_x(t) = E(e^{tx})}$

① Find the mgf of PD.

Sol: We've PD is $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2 \ldots$ where $\lambda = np$

Moment Generating Function is given by

$$M_x(t) = E(e^{tx})$$

$$= \sum_{x=0}^{\infty} e^{tx} p(x)$$

$$= \sum_{x=0}^{\infty} (e^{tx}) \frac{e^{-\lambda}\lambda^x}{x!}$$

$$M_x(t) = e^{-\lambda}\left(\sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}\right) \quad \left[\because e^x = 1 \times \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right]$$

$$= e^{-\lambda}\left[1 + \frac{(\lambda e^t)}{1!} + \frac{(\lambda e^t)^2}{2!} + \cdots \frac{(\lambda e^t)^r}{r!}\right]$$

$$= e^{-\lambda}\left[e^{\lambda e^t}\right]$$

$$\boxed{M_x(t) = e^{\lambda[e^t - 1]}}$$

② Find mgf of a Binomial Distribution

Sol: We've BD $P(X=x) = n_{c_x} p^x q^{n-x}$

mgf is given by

$$M_x(t) = E(e^{tx})$$

$$= \sum_{x=0}^{n} e^{tx} p(x)$$

$$= \left[\sum_{x=0}^{n} e^{tx} n_{c_x} p^x q^{n-x}\right]$$

$$M_x(t) = \sum_{x=0}^{n} n_{c_x} (pe^t)^x (q^{n-x})$$

$$= n_{c_0} (pe^t)^0 q^n + n_{c_1} (pe^t) q^{n-1} + \cdots + n_{c_n} (pe^t)^n q^0$$

$$= 1 \times 1 \cdot q^n + n_{c_1} (pe^t) q^{n-1} + \cdots + (pe^t)^n$$

$$\boxed{M_x(t) = (q + pe^t)^n}$$

Note: MGF of BD

$M_x(t) = (q + pe^t)^n$

$M_1 = \frac{d}{dt} M_x(t)$

$M_1 = \frac{d}{dt} (q+pe^t)^n$

$= n(q+pe^t)^{n-1} (0+pe^t)$

$= npe^t (q+pe^t)^{n-1}$

put $t=0$

$\mu_1 = npe^0 (q+pe^0)^{n-1}$

$= np(1)(q+p)^{n-1}$

$= np(1)^{n-1}$

$\boxed{\mu_1 = np}$

$\boxed{\mu_1 = \mu}$

First moment of a B.D $\quad \because p+q=1$

$\boxed{M_x'(t) = \mu_1' = np}$

2nd moment of a B.D

$\because M_x'(t) = npe^t (q+pe^t)^{n-1}$

diff wrt 't' above

$M_x''(t) = \frac{d}{dt} [npe^t (q+pe^t)^{n-1}]$

Apply u.v rule $uv = uv' + vu'$

$= np [e^t (n-1)(q+pe^t)^{n-2} (pe^t) + (q+pe^t)^{n-1} e^t]$

$M_x''(t) = np[e^0 (n-1)(q+p.e^0)^{n-2} (p.e^0] + (q+pe^0)^{n-1} e^0]$

$= np[(n-1)(q+p)^{n-2} p + (q+p)^{n-1}]$

$= np[(n-1)(1)^{n-2} p + (1)^{n-1}]$

$= np[(n-1)p+1]$

$[M_x''(t)]_{\text{at } t=0} = np[(n-1)p+1]$

$= np[(n-1)p + p+q]$

$= np[np - \cancel{p} + \cancel{p} + q]$

$= n^2p^2 + npq$

$$\boxed{M_x''(t)_{\text{at } t=0} = (np)^2 + npq}$$

$E(x) = \mu_1' = np$

$E(x^2) = (np)^2 + npq$

$Var(x) = E(x^2) - (E(x))^2$

$= \cancel{(np)^2} + npq - \cancel{(np)^2} = npq$

$$\boxed{Var(x) = npq}$$

1st moment of PD:

We have m.g.f of a PD:

$M_x(t) = e^{\lambda(e^t-1)}$ —— ①

1st moment

$M_x'(t) = \frac{d}{dt}(e^{\lambda(e^t-1)})$

$= e^{\lambda(e^t-1)}(\lambda e^t)$

$M_x'(t) = \lambda[e^t e^{\lambda(e^t-1)}]$ — ②

at $t=0$

$M_x'(t) = \lambda[e^0 e^{\lambda(e^0-1)}]$

1st moment $= M_x'(t)_{t=0} = \lambda = E(x)$.

In P.D

1st moment $\mu_1' = [M_x'(t)]_{t=0} = \lambda = E(X)$

$\because$ we've

$M_x'(t) = \lambda\ [e^t e^{\lambda(e^t-1)}]$ — ②

Now diff w.r.t 't'

$M_x''(t) = \frac{d}{dt}[\lambda e^t e^{\lambda(e^t-1)}]$

$= \lambda[e^t e^{\lambda(e^t-1)} \lambda e^t + e^{\lambda(e^t-1)} e^t)$

at $t=0$

$= \lambda[\lambda+1]$

$= \lambda^2+\lambda$

$E(X^2) = M_x''(0) = \lambda^2+\lambda$

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

$= \cancel{\lambda^2}+\lambda-\cancel{\lambda^2}$

$\boxed{Var(X) = \lambda}$

Moments about a point $x = A$ (OR) Arbitrary Moments

rth Moment

i) $\mu_r' = \frac{\Sigma(x_i-A)^r}{n}$

n = no. of observation

ii) $\mu_r' = \frac{\Sigma f_i (x_i-A)^r}{N}$

where $N = \Sigma f_i$

Moment about the mean $(\bar{x}=\mu)$ (OR) central moments

i) $\mu_r = \frac{\Sigma(x_i-\bar{x})^r}{n}$ $\quad [(x_i-A-(\bar{x}-A)]^r$

$[(x_i-A-\mu_1]^r$

ii) $\mu_r = \dfrac{\sum f_i (x_i - \bar{x})^r}{N}$ $N = \sum f_i$

Relation between central moments interms of Arbitrary moments vice versa

Central Moments

① $\mu_1 = 0$

② $\mu_2 = \mu_2' - (\mu_1')^2$

③ $\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$

④ $\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$

Arbitrary moment

① $\mu_1' = \bar{x} - A$

② $\mu_2' = \mu_2 + (\mu_1')^2$

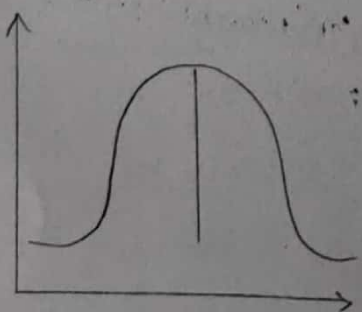
③ $\mu_3' = \mu_3 + 3\mu_2\mu_1' + (\mu_1')^3$

④ $\mu_4' = \mu_4 + 4\mu_3\mu_1' + 6\mu_2(\mu_1')^2 + (\mu_1')^4$

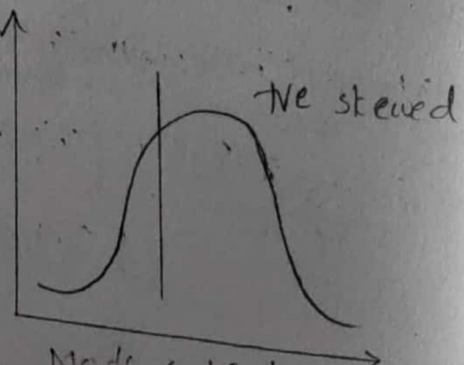
Note: Moment of order zero in both ie $\boxed{\mu_0 = \mu_r' = 1}$

Skewness: It means lack of symmetry.

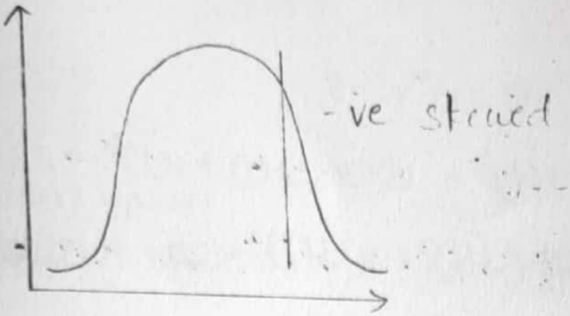
A freq distribution is said to be symmetrical if the va of equidistance from mean have same frequencies, ie in symmetrical distribution mean, median & mode are identical.



Mean = Median = Mode



Mode < Median < Mean



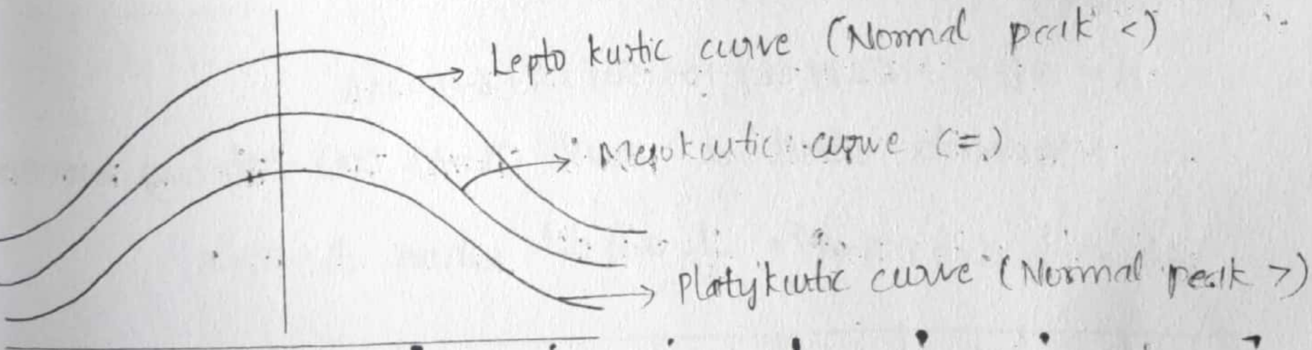
Mean < Median < Mode

Coefficient of skewness:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

kurtosis: It measures the degree to which a curve of frequency distribution is packed (or) flat topped. The coefficient of kurtosis is:

$$\beta_2 = \frac{\mu_4}{\mu_2}$$



2. The first four moments of a R.V 'x' about a point x=5 are 1, −5, 15, 30. Find the corresponding four moments about mean. Also find β_1 & β_2

So Let μ_r' → arbitrary (or) moments about a point

μ_r → moments about mean (central)

Given that moment about a point $x=5$

$\mu_1' = 1$, $\mu_2' = -5$, $\mu_3' = 15$, $\mu_4' = 30$.

$\mu_1 = 0$

$\mu_2 = \mu_2' - (\mu_1')^2 = -5-(1)^2 = -6$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3 = 15-3(-5)(1)+2(1)^3 = 32$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 = 30-4(15)(1)+6(-5)(1)$

$=$

$\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{(32)^2}{(-5)^3} = -8.192$

$\beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{30}{25} = 1.2$

Q. calculate the first four moments of the following distribution about the mean & hence find β_1 & β_2

x	0	1	2	3	4	5	6	7	8
f	1	8	28	56	70	56	28	8	1

Sol: Given ↑

$N = \Sigma f_i = 1+8+28+56+70+56+28+8+1 = 256$

∴ Moments about a point $(x=A)$ (or) Arbitary momen

$\mu_r' = \frac{1}{N}\Sigma f_i (x_i - A)^r = \frac{1}{N}\Sigma f_i d_i^r$ where $d_i = x_i - A$

x_i	f_i	$d_i = x_i - 4$	$f_i d_i$	$f_i d_i^2$	$f_i d_i^3$	$f_i d_i^4$	d_i^2	d_i^3	d_i^4
0	1	-4	-4	16	-64	256	16	-64	256
1	8	-3	-24	72	-216	648	9	-27	81
2	28	-2	-56	112	-224	448	4	-8	16
3	56	-1	-56	56	-56	56	1	-1	1
4	70	0	0	0	0	0	0	0	0
5	56	1	56	56	56	56	1	1	1
6	28	2	56	112	224	448	4	8	16
7	8	3	24	72	216	648	9	27	81
8	1	4	4	16	64	256	16	64	256
	256		0	512	0	2816			

$\mu_1 = 0$

$\mu_2 = 2$

$\mu_3 = 0 - 3(2)(0) + 0 = 0$

$\mu_4 = 11$

$\beta_1 = 0$

$\beta_2 = \frac{11}{4} = 2.75$

$\mu_1' = \frac{1}{256}(0) = 0$

$\mu_2' = \frac{1}{256}(512) = 2$

$\mu_3' = 0$

$\mu_4' = 11$

13/9/19
Friday

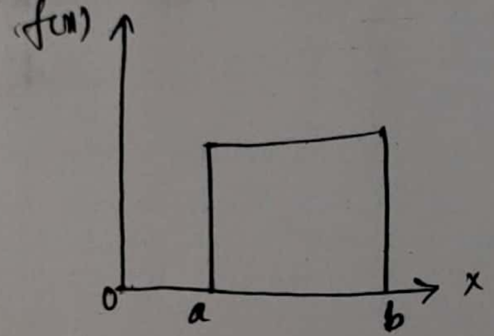
UNIT - III

CONTINUOUS PROBABILITY DISTRIBUTION

I. Uniform (or) Rectangular distribution :

A continuous random variable 'x' is said to follow uniform distribution if its pdf is given by f(x)

$$f(x) = \begin{cases} \frac{1}{b-a} ; & a<x<b \\ 0 ; & \text{elsewhere} \end{cases}$$



imp

$$f(x) = \begin{cases} \frac{1}{b-a} ; & a<x<b \\ 0 ; & \text{elsewhere} \end{cases}$$

∵ 'x' is a CRV

mean = E(x)

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

$$= \left[\int_{-\infty}^{a} + \int_{a}^{b} + \int_{a}^{\infty} \right] x f(x) dx$$

$$y = \int_{-\infty}^{a} x f(x) dx + \int_{a}^{b} x f(x) dx + \int_{a}^{\infty} x f(x) dx$$

$$y = 0 + \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b + 0$$

$$y = \frac{1}{b-a} \left(\frac{b^2-a^2}{2} \right) = \frac{b+a}{2}$$

$$\Rightarrow \boxed{E(x) = y = \frac{a+b}{2}}$$

Scanned by CamScanner

* Variance:-

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

$$E(x) = y = \frac{b+a}{2}$$

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

$$= \int_a^b x^2 \frac{1}{b-a} dx$$

$$= \left(\frac{1}{b-a}\right)\left[\frac{x^3}{3}\right]_a^b$$

$$= \frac{1}{b-a} \frac{b^3-a^3}{3}$$

$$= \frac{1}{b-a} \frac{(b-a)(b^2+ab+a^2)}{3}$$

$$E(x^2) = \frac{b^2+ab+a^2}{3}$$

$$Var(x) = \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2$$

$$= \left(\frac{b^2+ab+a^2}{3}\right) - \frac{1}{4}(b^2+2ab+a^2)$$

$$= \frac{4(b^2+ab+a^2) - 3(b^2+2ab+a^2)}{12}$$

$$= \frac{4b^2+4ab+4a^2-3b^2-6ab-3a^2}{12}$$

$$= \frac{b^2-2ab+a^2}{12}$$

$$= \frac{(b-a)^2}{12}$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$

$$\sigma = SD = \sqrt{\frac{(b-a)^2}{12}}$$

$$\sigma = \frac{b-a}{2\sqrt{3}}$$

Scanned by CamScanner

Moment Generating function:

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

$$= \frac{1}{b-a} \int_a^b e^{tx}\, dx$$

$$M_x(t) = \frac{1}{t(b-a)} (e^{tb} - e^{ta})$$

Q) If x is uniformly distributed with mean 1 & variance 4/3 find $P(x<0)$

Sol: In UD mean = 1

$$\frac{b+a}{2} = 1$$

$$b+a = 2 \quad \text{—(1)}$$

Variance $= \frac{4}{3}$

$$\frac{(b-a)^2}{12} = \frac{4}{3}$$

$$b-a = 4 \quad \text{—(2)}$$

$$b-a = -4 \quad \text{—(3)}$$

from (1) & (2)

$a = -1,\ b = 3$

from (1) and (3)

$a = 3;\ b = -1$

$\therefore\ a<x<b$ & $a=b \Rightarrow a=-1;\ b=3$

$$f(x) = \begin{cases} \frac{1}{b-a}; & a<x<b \\ 0; & \text{elsewhere} \end{cases}$$

& $a<b$

so: $a=-1,\ b=3$

$$P(x<0) = P(-1<x<0)$$

$$= \int_{-1}^{0} f(x)\, dx = \int_{-1}^{0} \frac{1}{b-a}\, dx$$

$$= \frac{1}{b-a} \int_{-1}^{0} dx = \frac{1}{3-(-1)} [x]_{-1}^{0}$$

$$= \frac{1}{4} [0-(-1)]$$

$$= \frac{1}{4}$$

Scanned by CamScanner

2) A random variable x has a uniform distribution in $(-3,3)$. If $P(X>k)=\frac{1}{3}$. Find 'k' also evaluate $P(X<2)$ and $P(|x-2|<2)$

Sol: we've uniform distribution

$$f(x)=\begin{cases}\frac{1}{(b-a)}, & a<x<b\\ 0, & \text{elsewhere}\end{cases}$$

$$f(x)=\begin{cases}\frac{1}{3-(-3)}, & -3<x<3\\ 0, & \text{elsewhere}\end{cases}$$

$$f(x)=\begin{cases}\frac{1}{6}, & -3<x<3\\ 0, & \text{elsewhere}\end{cases} \longrightarrow (1)$$

Given $P(x>k)=\frac{1}{3}$

$$1-P(X\leq k)=\frac{1}{3}$$

$$1-[P(-3\leq x\leq k)]=\frac{1}{3}$$

$$1-\int_{-3}^{k} f(x)\,dx=\frac{1}{3}$$

$$1-\int_{-3}^{k} \frac{1}{6}\,dx=\frac{1}{3}$$

$$1-\frac{1}{6}[x]_{-3}^{k}=\frac{1}{3}$$

$$1-\frac{1}{6}[k+3]=\frac{1}{3}$$

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

$$\frac{2}{\cancel{3}}=\frac{k+3}{\cancel{6}\,2}$$

$$4=k+3$$

$$\boxed{k=1}$$

Now, $P(X<2)=P(-3<x<2)$

$$=\int_{-3}^{2}\frac{1}{6}\,dx$$

$= \frac{1}{6}(x)^2_{-3}$

$= \frac{1}{6}(2-(-3))$

$= \frac{1}{6}(2+3)$

$\therefore \boxed{P(X<2) = \frac{5}{6}}$

$|x|<k$

$P(|X-2|<2) = P(-2<(x-2)<2)$

$= P((2-2)<x<2+2)$

$= P(0<x<4)$

$= \int_0^3 f(x)dx = \int_0^3 \frac{1}{6}dx$

$= \frac{1}{6}(x)_0^3$

$= \frac{1}{6}(3-0) = \frac{1}{2}$

3) A UD is given by $f(x)=1, 0\le x\le 1$.. Find mean & variance

So: We've UD

$$f(x) = \begin{cases} \frac{1}{b-a}, & a<x<b \\ 0, & \text{otherwise} \end{cases}$$

$$f(x) = \begin{cases} 1, & 0\le x\le 1 \\ 0, & o/w \end{cases}$$

From UD

Mean $= E(x) = \frac{b+a}{2} = \frac{1+0}{2}$

$\boxed{\mu = E(x) = \frac{1}{2}}$

Variance $= \sigma^2 = \frac{(b-a)^2}{12} = \frac{(1-0)^2}{12} = \frac{1}{12}$ //

* Moments of a Uniform Distribution:

we've MGF of UD

$$M_x(t) = E(e^{tx}) = \frac{(e^{bt} - e^{at})}{(b-a)t}$$

$\because e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots$

$$M_x(t) = \frac{1}{(b-a)t}\left\{1 + \frac{(bt)^1}{1!} + \frac{(bt)^2}{2!} + \frac{(bt)^3}{3!} + \cdots\right] \cdots$$

$$\left[1 + \frac{(at)^1}{1!} + \frac{(at)^2}{2!} + \frac{(at)^3}{3!} + \cdots\right]$$

$$M_x(t) = \frac{1}{(b-a)t}\left[\frac{(b-a)t}{2} + \frac{(b^2-a^2)t^2}{2} + \frac{(b^3-a^3)t^3}{3} + \cdots\right]$$

$$= \left[1.t^0 + \frac{(b^2-a^2)t^2}{2(b-a)t} + \frac{(b^3-a^3)t^3}{6(b-a)t} + \cdots\right.$$

$$= \left[1.t^0 + \left[\frac{(b-a)(b+a)}{2(b-a)}\right]t^1 + \frac{(b-a)(b^2+ab+a^2)t^2}{6(b-a)} + \cdots\right.$$

$$= \left[1.t^0 + \frac{(b+a)}{2}\frac{(t^1)}{1!}\right] + \frac{(b^2+ab+a^2)}{3}\left[\frac{t^2}{2!}\right] + \cdots$$

Moments about origin $= \mu_r' =$ the coefficient of $\frac{t^r}{r!}$

$\mu_0' = 1$

$\mu_1' = \left(\frac{b+a}{2}\right) = E(x)$

$\mu_2' = \frac{(b^2+ab+a^2)}{3} = E(x^2)$

Variance $(t) = \sigma^2 = E(x^2) - (E(x))^2$

$$= \left(\frac{b^2+ab+a^2}{3}\right) - \left(\left(\frac{b+a}{2}\right)^2\right)$$

$$= \left(\frac{b-a}{2}\right)^2$$

Similarly obtain the coeff of t^3 & t^4

we get

$$\mu_3' = \frac{(a+b)(a^2+b^2)}{4}$$

$$\mu_4' = \frac{(a^4 + a^3b + a^2b^2 + ab^3 + b^4)}{5}$$

and compute central moments

$\mu_1 = 0$

$\mu_2 = \mu_2' - (\mu_1')^2$

μ_3 & μ_4 and also find β_1, β_2

Exponential Distribution :

→ A continuous variable 'x' is said to follow exponential distribution if its PDF is given $f(x) = \begin{cases} \theta \cdot e^{-\theta x} & ; x \geq 0 \\ 0 & ; \text{otherwise} \end{cases}$

where $\theta > 0$

① Mean of Exponential Distribution :

Mean = $E(X)$

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

we have Exp. distribution

$$f(x) = \begin{cases} \theta \cdot e^{-\theta x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\mu = E(x) = \left[\int_{-\infty}^{0} + \int_{0}^{\infty}\right] x f(x) dx$$

$$= \int_{-\infty}^{0} 0 \cdot dx + \int_{0}^{\infty} x f(x) dx$$

$$= \int_{0}^{\infty} x \, \theta \cdot e^{-\theta x} dx$$

$$\mu = E(x) = \theta \int_{x=0}^{\infty} x \cdot e^{-\theta x} dx \Big].$$

$\int uv dx = uv_1 - u'v_2 + u''v_3$

$$= \theta \left\{ x \left[\frac{e^{-\theta x}}{(-\theta)}\right] - 1 \left[\frac{e^{-\theta x}}{(-\theta)^2}\right]_0^{\infty} \right.$$

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

$$= \theta \left[-\frac{1}{\theta^2}\left[0 - 1\right]\right].$$

$= \theta \left[\frac{1}{\theta^2} \right]$

$$\boxed{\mu = E(X) = \frac{1}{\theta}}$$

Variance of Exponential distribution

we have Exp dis

$$f(x) = \begin{cases} \theta \cdot e^{-\theta x}, & x \geq 0, \ \theta > 0 \\ 0, & o/w \end{cases}$$

$\therefore var(x) = \sigma^2 = E(x^2) - [E(x)]^2 - (1)$

First compute

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

$$= \int_{0}^{\infty} x^2 \theta \cdot e^{-\theta x} dx$$

$$E(x^2) = \theta \int_{x=0}^{\infty} (x^2 e^{-\theta x}) dx \qquad \begin{aligned} u &= x^2 \\ u &= e^{-\theta x} \end{aligned}$$

$$= \theta \left[x^2 \left[\frac{e^{-\theta x}}{-\theta} \right] - (2x) \left[\frac{e^{-\theta x}}{(-\theta)^2} \right] + 2 \left[\frac{e^{-\theta x}}{(-\theta)^3} \right]_0^{\infty} \right.$$

$$= \theta \left[0 - 0 - \frac{2}{\theta^3} \left[e^{-\theta x} \right]_{x=0}^{\infty} \right]$$

$$= \theta \left[\frac{-2}{\theta^3} \left[e^{-\infty} - e^{0} \right] \right]$$

$$= \cancel{\theta} \left[\frac{-2}{\theta^{3\,2}} [0 - 1] \right]$$

$$= \frac{2}{\theta^2}$$

$$\boxed{E(x^2) = \frac{2}{\theta^2}}$$

from (1)

$$var(x) = \sigma^2 = \frac{2}{\theta^2} - \left(\frac{1}{\theta} \right)^2$$

$$= \frac{2-1}{\theta^2}$$

$$\boxed{Var(x) = \frac{1}{\theta^2}}$$

$$S.D = \sqrt{\sigma^2} = \sqrt{\frac{1}{\theta^2}}$$

$$\boxed{SD = \frac{1}{\theta}}$$

Moment Generating function (MGF) of Exp distribution

We've Exp distribution $f(x) = \begin{cases} \theta \cdot e^{-\theta x}, & x \geq 0 \\ 0, & o/w \end{cases}$

$$MGF = M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

$$= \int_0^{\infty} e^{tx} \theta e^{-\theta x}\,dx$$

$$= \theta \left[\int_0^{\infty} e^{-(\theta - t)x}\,dx \right]$$

$$= \theta \left[\frac{e^{-(\theta - t)x}}{-(\theta - t)} \right]_{x=0}^{\infty}$$

$$MGF = M_x(t) = \theta \frac{(e^{-\infty} - e^0)}{-(\theta - t)}$$

$$= -\left(\frac{\theta}{\theta - t}\right)[0 - 1]$$

$$= \left(\frac{\theta}{\theta - t}\right)$$

$$= \left(\frac{\theta - t}{\theta}\right)^{-1}$$

$$= \left(1 - \frac{t}{\theta}\right)^{-1}$$

$$\because [1 - x]^{-1} = 1 + x + x^2 + x^3 + \ldots$$

$$MGF = M_x(t) = \left[1 - \frac{t}{\theta}\right]^{-1}$$

$$= \left[1 + \left(\frac{t}{\theta}\right) + \left(\frac{t}{\theta}\right)^2 + \ldots\right]$$

$$M_x(t) = \sum_{n=0}^{\infty} \left(\frac{t}{\theta}\right)^n$$

Moments of an exponential distribution

We've mgf of Exp. dis

$$MGF = M_X(t) = \sum_{n=0}^{\infty} \left(\frac{t}{\theta}\right)^n$$

$\therefore$ Moments ~~that~~ about the origin is given by

$\mu_r' =$ The coefficient of $\frac{t^r}{r!}$

ie $\mu_r' =$ The coeff. of in $M_X(t)$

$$\mu_r' = \frac{r!}{\theta^r}$$

1st moment put $r=1$

$$\mu_1' = \frac{1!}{\theta^1} = \frac{1}{\theta} = E(X) = \mu$$

put $r=2$

$$\mu_2' = \frac{2!}{\theta^2} = \frac{2}{\theta^2} = E(X^2)$$

$$\mu_3' = \frac{3!}{\theta^3} = \frac{6}{\theta^3}, \quad \mu_4' = \frac{24}{\theta^4}$$

Then find central moment

$\mu_1 = 0$

$$\mu_2 = \mu_2' - (\mu_1')^2$$

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

$$= \frac{2}{\theta^2} - \frac{1}{\theta^2}$$

$$= \frac{1}{\theta^2}$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$

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

$$= \cancel{\frac{6}{\theta^3}} - \cancel{\frac{6}{\theta^3}} + \frac{2}{\theta^3}$$

$$= \frac{2}{\theta^3}$$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$

$= \frac{24}{\theta^4} - 4\left(\frac{6}{\theta^3}\right)\left(\frac{1}{\theta}\right) + 6\left(\frac{2}{\theta^2}\right)\left(\frac{1}{\theta}\right)^2 - 3\left(\frac{1}{\theta}\right)^4$

$= \frac{24}{\theta^4} - \frac{24}{\theta^4} + \frac{12}{\theta^4} - \frac{3}{\theta^4}$

$= \frac{9}{\theta^4}$ //

n|9|19

Q) Subway trains on a certain line run every half-hour b/w midnight and six in the morning. What is the Probability that a man entering the station at a random time during this period will have to wait at least twenty minute?

Sol: Let 'x' be a RV waiting timer

Given $(a,b) = (0,30)$

we've UD

$$f(x) = \begin{cases} \frac{1}{b-a}, & a<x<b \\ 0, & o/w \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{30-0}, & 0<x<30 \\ 0, & o/w \end{cases}$$

$P(X \geq 20) = \int_{20}^{30} f(x)\,dx$

$= \int_{20}^{30} \frac{1}{30}\,dx$

$= \frac{1}{30}(x)_{20}^{30}$

$= \frac{1}{30}(30-20)$

$= \frac{10}{30} = \frac{1}{3}$

Scanned by CamScanner

Normal - Distribution (ND) [Gaussian Distribution]

Define N.D ?

Normal Distribution: A continuous random variable 'x' is said to follow a normal distribution if its p.d.f is given by

$f(x) = f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$

where, $-\infty < x < \infty$

$-\infty < \mu < \infty$

$\sigma > 0$

μ = mean

σ = SD are two parameters of a ND

Note:

$x \sim N(\mu, \sigma^2)$

Standard Normal variate:

Let $z = \frac{x-\mu}{\sigma}$, then $E(z) = 0$, $Var(z) = 1$

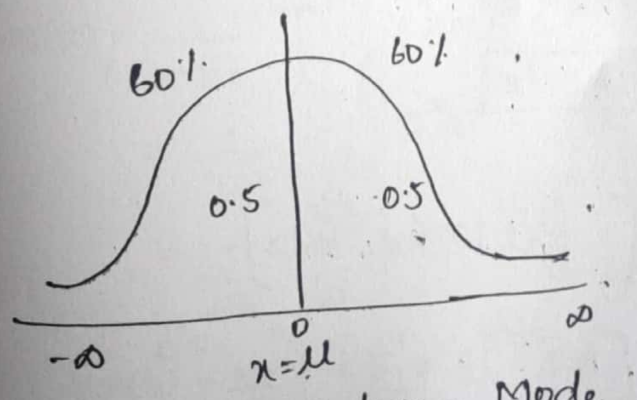
Thus 'z' is called SNV

$z \sim N(0,1)$

The pdf of a SNV 'z' is given by $\phi(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2}$, $-\infty < z < \infty$

Characteristics of ND:

1. The graph of the ND $y = f(x)$ in the xy plane is known as normal curve.



Mean = Median = Mode

2. Normal curve is bell shape and symmetrical about the line $x = \mu$

3. Area under the normal curve represents the total population.

4. For the normal curve mean = median = mode.
So N.C is unimodal.

Q) Find the mean of ND

Sol We've ND

$$f(x) = N(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$$

let $\mu = b$

$$f(x) = N(x, b, \sigma) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-b}{\sigma}\right]^2}$$

where, $-\infty < b < \infty$

$\sigma > 0$

$$\text{Mean} = E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

$$= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-b}{\sigma}\right]^2} dx$$

$$E(x) = \frac{1}{\sigma\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} x . e^{-\frac{1}{2}\left[\frac{x-b}{\sigma}\right]^2} dx\right] - (2)$$

let $\boxed{\frac{x-b}{\sigma} = z}$

$x - b = \sigma z \Rightarrow x = b + \sigma z$

$\boxed{dx = \sigma dz}$

Sub in (2)

$$E(x) = \frac{1}{\sigma\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} (b + \sigma z) e^{-z^2/2} \sigma dz\right]$$

$$E(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} b e^{-z^2/2} dz + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz$$

$\int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx$, If $f(x)$ E.F

$= 0$ If $f(x)$ is odd.F

$$F(X) = \frac{2}{\sqrt{2\pi}} \int_0^\infty b e^{-z^2/2}\, dz$$

$$= \frac{2b}{\sqrt{2\pi}} \int_0^\infty e^{-z^2/2}\, dz$$

$$= \frac{2b}{\sqrt{2}\cancel{\sqrt{\pi}}} \times \frac{\cancel{\sqrt{\pi}}}{\sqrt{2}} \qquad \because \int_0^\infty e^{-x^2/2}\,dx = \sqrt{\pi/2}$$

$$= \frac{2b}{2} = b = \mu$$

Find the variance of a Normal Distribution

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

(or)

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

we have normal distribution

$$f(x, b, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left[\frac{x-b}{\sigma}\right]^2}$$

there $b = \mu =$ mean where $-\infty < x,\ b < \infty,\ \sigma > 0$

$$Var(X) = E[(x-b)^2] = \int_{-\infty}^{\infty} (x-b)^2 \cdot f(x)\,dx$$

$$= \int_{-\infty}^{\infty} (x-b)\, e^{-1/2\left[\frac{x-b}{\sigma}\right]^2} dx$$

$$\frac{x-b}{\sigma} = z \Rightarrow x-b = \sigma z \Rightarrow x = \sigma z + b \Rightarrow dx = \sigma\, dz$$

$$Var(X) = \frac{1}{\sigma\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} (z\sigma)^2 e^{-1/2[z]^2}\, \sigma\, dz\right]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 z^2 e^{-1/2 z^2} dz$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \cdot e^{-z^2/2} dz$$

$$Var(X) = \frac{\sigma^2}{\sqrt{2\pi}}\, 2\int_0^\infty z^2 e^{-z^2/2} dz$$

put $\frac{z^2}{2} = t \Rightarrow z^2 = 2t$

$$\cancel{2}z\,dz = \cancel{2}dt$$

$$dz = \frac{dt}{z}$$

$$dz = \frac{dt}{\sqrt{2t}}$$

$$Var(x) = \frac{2\sigma}{\sqrt{2\pi}} \int_0^\infty (2t)\, e^{-t}\, dt/\sqrt{2t}$$

$$= \frac{2\sigma^2}{\sqrt{2}\sqrt{\pi}} \cdot \frac{2}{\sqrt{2}} \int_0^\infty e^{-t}\, t^1 \cdot t^{-1/2}\, dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty e^{-t}\, t^{1/2}\, dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty e^{-t}\, t^{3/2-1}\, dt$$

$$Var(x) = \frac{2\sigma^2}{\sqrt{\pi}} \sqrt{\tfrac{3}{2}}$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \sqrt{\tfrac{1}{2}+1} \qquad \sqrt{n} = \int_0^\infty e^{-t}\, t^{n-1}\, dt$$

$$= \frac{\cancel{2}\sigma^2}{\sqrt{\pi}} \frac{1}{\cancel{2}} \sqrt{\tfrac{1}{2}}$$

$$= \frac{\sigma^2}{\cancel{\sqrt{\pi}}} \cancel{\sqrt{\pi}}$$

$$= \sigma^2$$

** Find the mgf of a ND

So We've ND $f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 \left[\frac{x-\mu}{\sigma}\right]^2} \rightarrow (1)$

$-\infty < \mu, x < \infty, \quad \sigma > 0$

$$m.g.f = M_x(t) = E(e^{tx})$$

$$= \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \cdot e^{-1/2 \left[\frac{x-\mu}{\sigma}\right]^2} dx \quad -(2)$$

Scanned by CamScanner

put $\frac{x-\mu}{\sigma} = z$

$x-\mu = \sigma z \Rightarrow x = \sigma z + \mu$

$dx = \sigma dz$

same limits

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\sigma z+\mu)} e^{-z^2/2} \sigma dz$$

$$= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} e^{\sigma z t} \cdot e^{\mu t} \cdot e^{-z^2/2} dz\right]$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} e^{\sigma z t} \cdot e^{-z^2/2} dz\right]$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} e^{(-z^2/2 + \sigma z t)} dz\right]$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} e^{-1/2 (z^2 - 2\sigma t z)} dz\right]$$

$$M_x(t) = \frac{e^{\mu t}}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} e^{-1/2 [z^2 - 2\sigma t z + \sigma^2 t^2 - \sigma^2 t^2]} dz\right]$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} e^{-1/2 (z^2 - 2\sigma t z + \sigma^2 t^2)} e^{\frac{\sigma^2 t^2}{2}} dz\right]$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} e^{-1/2 (z-\sigma t)^2} e^{(\sigma t)^2/2} dz\right]$$

$$= \frac{e^{\mu t} \cdot e^{(\sigma t)^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2 (z-\sigma t)^2} dz$$

$$= \frac{e^{\left[\mu t + \frac{(\sigma t)^2}{2}\right]}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2 (z-\sigma t)^2} dz \quad \text{--- (3)}$$

Put $z - \sigma t = p$

$dz = dp$

From (3)

$$M_X(t) = \frac{e^{\left(\mu t + \frac{(\sigma t)^2}{2}\right)}}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} e^{-p^2/2} dp \right]$$

$$= \frac{e^{\left(\mu t + \frac{(\sigma t)^2}{2}\right)}}{\sqrt{2\pi}} \cdot 2 \int_0^{\infty} e^{-p^2/2} dp$$

$$= \frac{2e^{\left(\mu t + \frac{(\sigma t)^2}{2}\right)}}{\sqrt{2\pi}} \cdot \sqrt{\frac{\pi}{2}}$$

$$= \frac{\cancel{2}e^{\left(\mu t + \frac{(\sigma t)^2}{2}\right)}}{\cancel{\sqrt{2\pi}}} \cancel{\sqrt{\frac{\pi}{2}}}$$

$$= e^{\left(\mu t + \frac{(\sigma t)^2}{2}\right)}$$

$$= e^{\left(\mu t + \frac{(\sigma t)^2}{2}\right)}$$

$\therefore$ M.G.F of ND $= e^{\left(\mu t + \frac{(\sigma t)^2}{2}\right)}$

Prob

For a ND variate with mean 1 & SD 3 find the

i) $P(3.43 \le x \le 6.19)$

ii) $P(-1.43 \le x \le 6.19)$

Sol X is Normally distributed variate

Mean $=1$

$\Rightarrow \mu = 1$ and SD $=3$

$$\boxed{\sigma = 3}$$

we've standard Normal variate

$$z = \frac{x-\mu}{\sigma}$$

$P(3.43 \le x \le 6.19) = P(x_1 \le x \le x_2)$

let $x_1 = 3.43$ let $x_2 = 6.19$

$z_1 = \frac{x_1 - \mu}{\sigma}$ $z_2 = \frac{x_2 - \mu}{\sigma}$

$z_1 = \frac{3.43 - 1}{3}$ $z_2 = \frac{6.19 - 1}{3}$

$\boxed{z_1 = 0.81}$ $\boxed{z_2 = 1.73}$

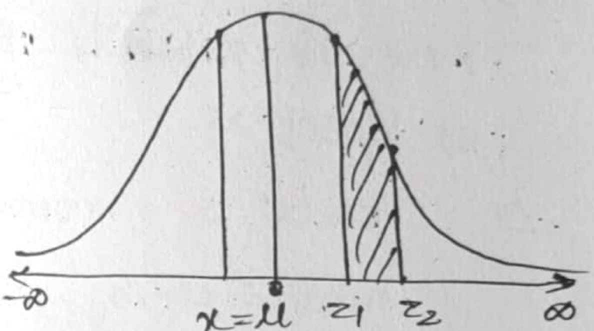
$P(x_1 < x \le x_2) = P(z_1 \le z \le z_2)$

$= P(0.81 \le z \le 1.73)$

$= P(z_2) - P(z_1)$

$= |A(z_2) - A(z_1)|$

$= |A(1.73) - A(0.81)|$



From the normate table.

$= 0.4582 - 0.2910$

$= 0.1672$

ii) $P(-1.43 \le x \le 6.19)$

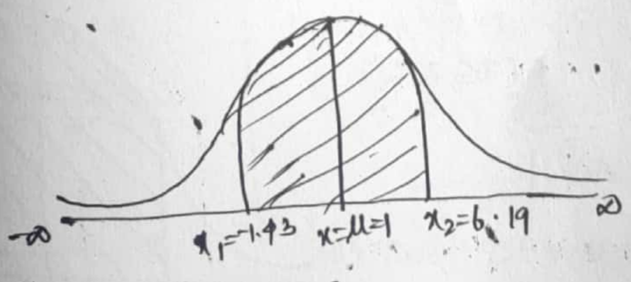
$z = \frac{x - \mu}{\sigma}$ $\mu = 1$

$\sigma = 3$

Take $x = x_1 = -1.43$ $x = x_2 = 6.19$

$z = z_1 = \frac{-1.43 - 1}{3}$ $z = z_2 = \frac{6.19 - 1}{3}$

$z_1 = -0.81$ $z_2 = 1.73$



$z_1 = -0.8$, $z = 0$, $z_2 = 1.73$

$P(-1.4) \le x \le 6.19) = P(-0.8 \le z \le 1.73)$

$= P(-0.8 \le z \le 0) + P(0 \le z \le 1.73)$

$= |A(0.8) + A(1.73)|$

$= 0.291 + 0.4582$ from the standard normal table

$\therefore \boxed{P(-1.43 \le x \le 6.19) = 0.7492}$

2) x is a normal variate with mean 30 and SD is 5. Find the probability that i) $26 \le x \le 40$ ii) $x \ge 45$ and iii) $|x-30| > 5$

So Given 'x' is a normal variate

mean $= \mu = 30$, $SD = \sigma = 5$

we have standard normal variate

$z = \frac{x-\mu}{\sigma}$ - ①

i) $P(26 \le x \le 40) = P(x_1 \le x \le 40)$

where $x = x_1 = 26$

from ①

$z = z_1 = \frac{26-30}{5}$

$z_1 = \frac{-4}{5}$

$\boxed{z_1 = -0.8}$

where $x = x_2 = 40$

$z = z_2 = \frac{40-30}{5}$

$\boxed{z_2 = 2}$

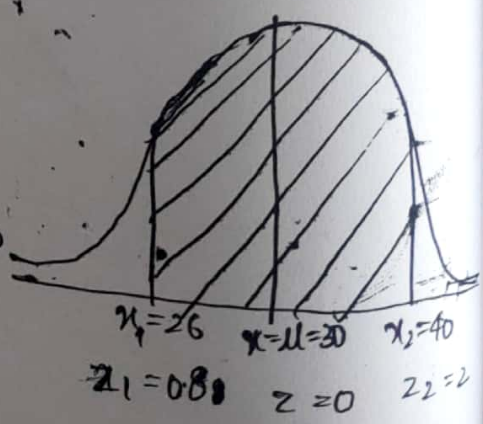
$P(26 \le x \le 40) = P(-0.8 \le z \le 2)$

$\Rightarrow P(-0 \le z \le 0.8) + P(0 \le z \le 2)$

$= |A(0.8) + A(2)|$

$= 0.2881 + 0.4772$ (from ND tables)

$= 0.7653$



ii) $P(x \geq 45)$

$\because z = \frac{x-\mu}{\sigma}$

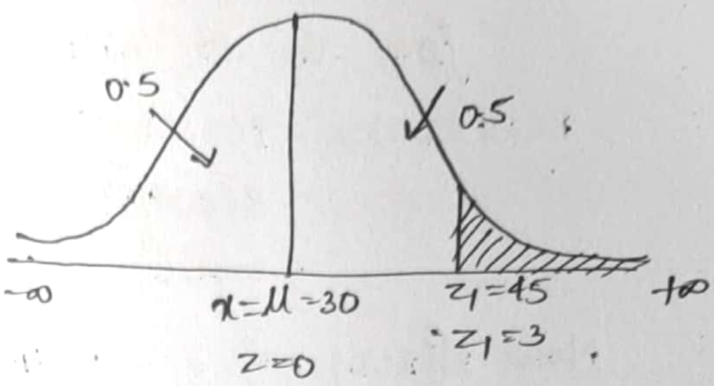
$z = \frac{x-30}{5}$

where $x = x_1 = 45$

$z = z_1 = \frac{45-30}{5}$

$z_1 = \frac{15}{5}$

$\boxed{z_1 = 3}$



$\therefore P(x \geq 45) = P(z \geq 3)$

$= 0.5 - P(0 \leq z \leq 3)$

$= 0.5 - |A(3)|$

$= 0.5 - 0.4987$ from ND table

$\boxed{P(x \geq 45) = 0.0013}$

iii) $P(|x-30| \geq 5)$

first compute

$P(|x-30| \leq 5)$

$P(-5 \leq |x-30| \leq 5)$

$P(-5+30 \leq x \leq 5+30)$

$P(25 \leq x \leq 35)$

$= P(x_1 \leq x \leq x_2)$

where $x = x_1 = 25$

$z_1 = \frac{25-30}{5}$

$\boxed{z_1 = -1}$

where $x = x_2 = 35$

$z = z_2 = \frac{35-30}{5}$

$\boxed{z_2 = 5}$

$P(25 \leq x \leq 35) = P(-1 \leq z \leq +1)$

$= P(0 \leq z \leq 1) + P(0 \leq z \leq 1)$

$= 2P(0 \leq z \leq 1)$

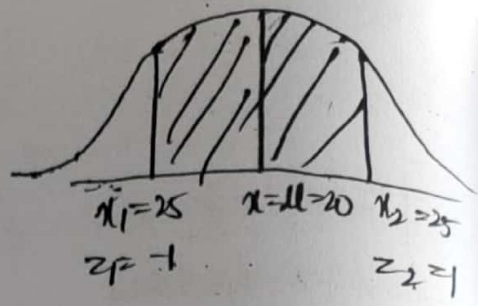
$= 2 \cdot |A(1)|$

from the ND table

$P(25 \le x \le 35) = P(-1 \le z \le 1)$

$= 2(0.3413)$

$= 0.6826$



Now $P(|x-30| > 5) = 1 - P(|x-30| \le 5)$

$= 1 - 0.6826$

$= 0.3174$

(XXX) In a normal distribution 31% of the items are under 45 and 8% are over 64. Find the mean and SD of the distribution.

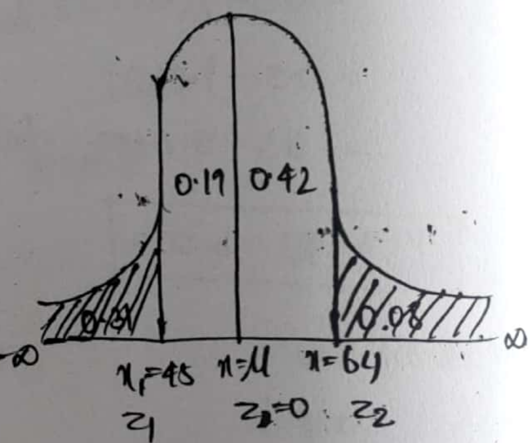
Sol 31% items are under 45

$P(x \le 45) = 31\%$

$= \frac{31}{100} = 0.31$

$P(x > 64) = 8\%$

$= 0.08$



$\because z = \frac{x-\mu}{\sigma}$

Put $x=45$, $\frac{45-\mu}{\sigma} = -z_1$ — ①

Put $x=64$, $\frac{64-\mu}{\sigma} = z_2$ — ②

Also given

$\therefore P(0 \le z \le z_1) = 0.19$ (from the table corresponding z value

$= 0.5 = z_1$

$\because P(0 \le z \le z_2) = 0.42$ (from the table corresponding z va

$= 1.4 = z_2$

sub z_1 & z_2 in eq ① & ②

$\frac{45-\mu}{\sigma} = 0.5$

$\Rightarrow 45-\mu = (-0.5)\sigma$ — ③

$\Rightarrow \frac{64-\mu}{\sigma} = 1.9$

$\Rightarrow 64-\mu = (1.4)\sigma$ — ④

solving ③ & ④ by substitute method

$$45-\mu = (-0.5)\sigma$$
$$64-\mu = (1.4)\sigma$$
$$- \quad + \qquad -$$

$$-19 = -1.9\sigma$$

$$\sigma = \frac{19}{1.9}$$

$$\boxed{\sigma = 10}$$

$SD = \sigma = 10$

sub 'σ' values in ③

$$45-\mu = (-0.5)10$$

$$45-\mu = -5$$

$$\boxed{\mu = 50}$$

mean $= 50$//

H.w The marks obtained in mathematics by 1000 students is normally distributed with mean 78% and SD 11%. Determine

(i) How many students got marks above 90%.

(ii) What was the highest marks obtained by the lowest 10% of the students

→ If 'x' is a N.V find the area A

i) to the left of $z = -1.78$

ii) to the right of $z = -1.45$

iii) Corresponding to $-0.8 \le z \le 1.53$

iv) to the left of $z = -2.52$ and the result of $z = 1.83$

→ Moments about mean

We have mgf of a ND:

$$M_x(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} = E[e^{tx}] \quad (1)$$

Now mgf about a mean

$$M_x(t) = E(e^{t(x-\mu)}) = E[e^{tx} \cdot e^{-\mu t}]$$

$$= e^{-\mu t}[E(e^{tx})]$$

from (1)

$$M_x(t) = e^{-\mu t}\left[e^{\mu t + \frac{\sigma^2 t^2}{2}}\right]$$

$$= e^{-\mu t} e^{\mu t} e^{\frac{[\sigma t]^2}{2}}$$

$$M_x(t) = E(e^{t(x-\mu)}) = e^{\left(\frac{\sigma^2 t^2}{2}\right)} \quad (2)$$

in the m.g.f about mean

$$M_x(t) = \left[1 + \frac{\left(\frac{\sigma^2 t^2}{2}\right)}{1!} + \frac{\left(\frac{\sigma^2 t^2}{3}\right)^2}{2!} + \frac{\left(\frac{\sigma^2 t^2}{3}\right)^3}{3!} + \cdots\right]$$

$$M_x(t) = \left[1 + \sigma^2 \frac{t^2}{2!} + 3\sigma^4 \frac{t^4}{4!} + \cdots\right]$$

Moments about a mean are given by

μ_r = The coefficient of $\left(\frac{t^r}{r!}\right)$

$\mu_0 = 1$, $\mu = 0$

$\mu_2 = \sigma^2$

$\mu_3 = 0$

$\mu_4 = 3\sigma^4$

$\mu_{2n+1} = 0$

That is all odd moments are zero

$\mu_1 = \mu_3 = \mu_5 = \cdots 0$

Skewness

$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

$\beta_1 = \frac{0}{\sigma^2}$

$\boxed{\beta_1 = 0}$

$\boxed{\beta_2 = 3}$ and $\boxed{\beta_1 = 0}$

Kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{(\sigma^2)^2} = 3$

Scanned by CamScanner

25/9/19
Wednesday

V. TEST OF SIGNIFICANCE:

Population: It is the collection of objects (living or non-living)

Sample: A finite sub-set of a population is called a sample.
Sample size is denoted by 'n'.
If $n \geq 30$, then the sample is large sample
If $n < 30$, then the sample is small sample

→ The measurements of the population such as the mean (μ), variance (σ^2), standard Deviation are known as parameters.

→ Measurements of the samples such as the mean, variance, standard deviation etc are known as statistics.

Hypothesis:

→ It is an assumption about the population parameters.

Null Hypothesis (H_0): A definite element about the population parameters such a hypothesis is usually a hypothesis of no difference.

Alternative Hypothesis (H_1): Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis

$$\boxed{\overline{H_0} = H_1}$$

Ex: NH (H_0) = $\mu = \mu_0$

NH(H_1) = $\mu \neq \mu_0$ (two failed test)
$\mu > \mu_0$ (right failed test)
$\mu < \mu_0$ (left failed test)

Q. Define Type-I & Type-II Errors:

Type I Rejection of null hypothesis if it is true

Error:

Type II Error: Acceptance of null hypothesis when it is false

The sizes of Type I and Type II errors ie (α, β) are also known as producers and consumers list.

* **Working rule of Hypothesis:**

step-I: set up null hypothesis (H_0)

step II: Set up the alternative hypothesis (H_1) this gives whether we have to use 2-tailed test (or) 1-tailed test

step III: Level of significance: In general, $\alpha = 5\%$ (or) 2%.

step IV: Test-statistics $|t_{cal}|$; $|F_{cal}|$; $|\chi^2_{cal}|$; $|Z_{cal}|$

step V: Conclusion, if calculated value is less than tabulated value accept null hypothesis.

Student's t-test

Formulae: (single mean)

The test-statistics

$t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ where $s^2 = \dfrac{\Sigma(x-\bar{x})^2}{n}$

$s = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n}}$

s = SD of the population

Note:

If SD of the sample 's' gives $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}}$

Q) A random sample of 10 boys had the IQs 70, 120, 110, 101, 88, 83, 95, 98, 107 & 100. Do these data support the assumption of a population mean. IQ of 160 (tabulated value (α) $t_{0.09} = 2.26$)

Sol: Given

The size of the sample

i.e $n=10$.

$\because$ $n<30$ (so given sample space is small sample)

$\because$ Problem related. mean so we have to use t - distribution test.

x	$(x-\bar{x})$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
$\Sigma x = 972$		$\Sigma = 1866.6$

$$\bar{x} = \frac{\Sigma x}{10} = \frac{972}{10} = 97.2$$

$\mu = 160$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$$s^2 = \left(\frac{1866.6}{10-1}\right)$$

$$s^2 = \frac{1866.6}{9}$$

$s = \sqrt{207.4}$

$s = 14.40$

step 1: Null Hypothesis $H_0 :- \mu = 160$

(The population mean of IQ is 160)

step 2: Alternative Hypothesis: $H_1 :- \mu \neq 160$

step 3: Level of significance $\alpha = 1\%$.

ν = degree of freedom

$\nu = n-1$

$\nu = 10-1 = 9$

$$\boxed{\begin{array}{l} t_{tab}\ 0.09 = 2.26 \\ \text{with } \nu = 9 \end{array}}$$

step 4: The test statistic

$$t_{cal} = \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right|$$

$$t_{cal} = \left| \frac{97.2 - 160}{\frac{14.40}{\sqrt{10}}} \right|$$

$t_{cal} = 13.79$

step 5: Conclusion: $\because t_{cal} > t_{tab}$

(so reject the Null hypothesis 'H_0'.)

Prob 8: A number of candidates appeared to a selection trail for recruitment in the array. A random sample of heights of 10 candidates are.

in cms	162	170	168	169	173	171	165	166	161	160

can be conclude that the average height of the candidates is greater than 165 cm) Test at 5% level of significance ($t_\alpha = 1.833$)

Sol: Given $n = 10 < 30$

so small sample

$\mu = 165$ (cm) (population mean)

Sample mean: $\bar{x} = \frac{\Sigma x}{n} = \frac{1665}{10} = 166.5$

x	$(x-\bar{x})$	$(x-\bar{x})^2$
162	-4.5	20.25
170	3.5	12.25
168	1.5	2.25
169	2.5	6.25
173	6.5	42.25
171	4.5	20.25
165	-1.5	2.25
166	-0.5	0.25
161	-5.5	30.25
160	-6.5	42.25
$\Sigma x = 1665$		$\Sigma(x-\bar{x})^2 = 178.5$

Variance:

$\because s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$

$s^2 = \frac{178.5}{10-1}$

$s^2 = \frac{178.5}{9}$

$s^2 = 19.83$

$s = \sqrt{19.83}$

$\boxed{s = 4.45}$

Step 1: Null Hypothesis $H_0: \mu = 165$ cm

Step 2: Alternative hypothesis $H_1: \mu > 165$ cm (Right-tailed test)

Step 3: level of significance $\alpha = 5\%$

ν = degree of freedom

$\nu = n-1$

$\nu = 10-1 = 9$

t_{tab} at $5\% = 1.833$

with $\nu = 9$

Step 4:- The test statistic:

$$t_{cal} = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right|$$

$$t_{cal} = \left|\frac{166.5-165}{\frac{4.45}{\sqrt{10}}}\right|$$

$t_{cal} = 1.06$

Step 5:- Conclusion $t_{cal} < t_{tab}$

(so accept the null hypothesis H_0 at $\alpha = 5\%$)

Hwo A sample of 20 items has mean 42 units and SD 5 units. Test the hypothesis that it is a random sample from a normal population with mean 45 units.

$n = 20 < 30$ (small sample)

$\bar{x} = 42$ (mean of sample)

$\mu = 45$ (" population)

$s = 5$

Scanned by CamScanner

Step ①: Null Hypothesis H_0: $\mu = 45$ unit

Step ②: Alternative Hypothesis H_1: $\mu \neq 45$ (2-tailed test)

Step ③: level of significance $\alpha = 5\%$

$\therefore \alpha = 0.05$

ν = degree of freedom

$\nu = n-1$ $\alpha = 0.05$

$\nu = 20-1 = 19$ $\alpha/2 = 0.025$

$t_{\alpha/2 = 0.025} = 2.093$

with $\nu = 19$

$$\boxed{t_{tab} = 2.093}$$

Step ④: The test statistic

$$t_{cal} = \left| \frac{\bar{x} - \mu}{s/\sqrt{n-1}} \right|$$

$$t_{cal} = \left| \frac{42-45}{5/\sqrt{20-1}} \right| = \left| \frac{42-45}{\frac{5}{\sqrt{19}}} \right|$$

$t_{cal} = 2.61$

Step ⑤: Conclusion: $t_{cal} > t_{tab}$

(Reject Null Hypothesis 'H_0')

Test of significance for different mean (Two means)

Formula:

The test statistic $t = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

(or)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s^2 = \dfrac{\Sigma(x_i-\bar{x})^2 + \Sigma(x_i-\bar{y})^2}{n_1+n_2-2}$

Note ①: degree of freedom

$\nu = n_1+n_2-2$

Note ②: If the SD of the samples s_1 & s_2 are given direct

then $s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$

Note: formula;

confidence (or) Fiducial limits

1) at $\alpha = 5\%$

$\bar{x} \pm t_{0.05} \dfrac{s}{\sqrt{n}}$

2) at $\alpha = 1\%$

$\bar{x} \pm t_{0.01} \dfrac{s}{\sqrt{n}}$

Prob: Samples of two types of electric light bulbs were tested for length of life and following data were obtained

Type-I	Type-II
$n_1 = 8$ $\bar{x} = 1234$ hours $s_1 = 36$ hrs	$n_2 = 7$ $\bar{y} = 1036$ hrs $s_2 = 40$ hrs

Is the difference in the means sufficient to warrant that type-I is superior to type-II regarding length of life.

Sol: Given

size of 1st sample $n_1 = 8$

" " 2nd sample $n_2 = 7$

Mean of 1st sample $\bar{x} = 1234$ hrs

$\bar{y} = 1036$ hrs

SD of 1st sample $s_1 = 36$ hrs
$s_2 = 40$ hrs

1. NH: $H_0: \mu_1 = \mu_2$
2. AH: $\mu_1 > \mu_2$ (Right-tailed test)
3. Level of significance $\alpha = 5\%$

$\alpha = 0.05$

$\nu = d.f = n_1 + n_2 - 2$

$= 8 + 7 - 2$

$= 15 - 2$

$\nu = 13$

$t_{\alpha = 5\%} = 1.77$

4. Test-Statistic $t_{cal} = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$t = 9.39$

$\therefore \; t_{cal} > t_{tab}$

Conclusion : = Reject H_0.

Q) Two independent samples of 7 items had the following values.

Sample-I	11	11	13	11	15	9	12	14
Sample-II	9	11	10	13	9	8	10	-

Is the diff b/w the means of sample significant.
(Given $t_\alpha = t_{0.05} = 2.16$)

Sol: Given data

Let n_1 = size of sample-I

$n_1 = 8$

n_2 = size of sample-II

$n_2 = 7$

(Both are small samples)

$\because$ SD are not given so,

$$\hat{S}^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2\right]$$

$$\because \bar{x} = \frac{\Sigma x_i}{n}$$

$$\text{i.e. } \bar{x} = \frac{11+11+13+11+15+9+12+14}{8}$$

$$\bar{x} = 12$$

$$\bar{y} = \frac{\Sigma y_i}{n}$$

$$\bar{y} = \frac{9+11+10+13+9+8+10}{7}$$

$$\bar{y} = 10$$

x_i	$x_i - \bar{x}$	$(x_i-\bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i-\bar{y})^2$
11	-1	1	9	-1	1
11	-1	1	11	1	1
13	1	1	10	0	0
11	-1	1	13	3	9
15	3	9	9	-1	1
9	-3	9	8	-2	4
12	0	0	10	0	0
14	2	4			

$\Sigma(x_i-\bar{x})^2 = 26$ $\qquad$ $\Sigma(y_i-\bar{y})^2 = 16$

Now, $\hat{S}^2 = \frac{1}{8+7-2}[26+16]$

$= 3.23$

$SD = \sqrt{3.23} = 1.79$

Step 1:- Null hypothesis:- $H_0 : \mu_1 = \mu_2$

Step 2:- Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ [Two-tailed test]

Scanned by CamScanner

Step3: level of significance: $\alpha = 5\%$

$\alpha = 0.05$

with degree of freedom $\nu = n_1 + n_2 - 2$

$\nu = 8 + 7 - 2$

$\nu = 13$

$t_{0.05} = 2.16$ with $\nu = 13$

Step4: test - statistic $t_{cal} = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$t_{cal} = \dfrac{12 - 10}{1.79\sqrt{\frac{1}{8} + \frac{1}{7}}}$

$t_{cal} = 2.158$

$\boxed{t_{0.05} = 2.16}$

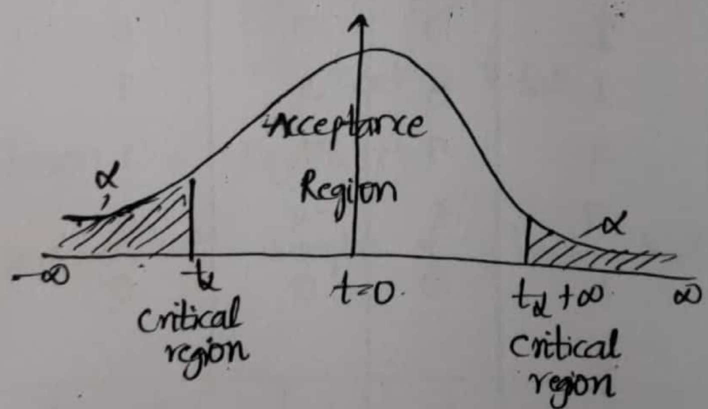
⇓

t_{tab}

Step5: $t_{cal} < t_{tab}$

Accept H_0

8) Properties of t-Distribution.



1. The shape of t-Distribution is a bell shape, which is similar to normal distribution

2. x-axis is an assymptote to the curve of the t-distribution

Application of t-Distribution:

1. To test if the sample mean differs significantly from the hypothetical value 'μ' of the population mean

2. To test the significance b/w two sample means

3. To test the significance of observed partial & multiple correlation coefficient.

F-test (or) Variance Ratio test.

The objective of the test is to determine whether there is a significant difference b/w the two population variant or whether the two samples may be recorded as drawn from the populations having same variance ($\sigma_1^2 = \sigma_2^2$).

formula for F-test.

The test statistic is given by

$$F_{cal} = \frac{s_1^2}{s_2^2} = \frac{\text{Greater variance}}{\text{Smaller variance}}$$

where $s_1^2 = \frac{1}{n_1-1}\left[\Sigma (x_i - \bar{x})^2\right]$

$s_2^2 = \frac{1}{n_2-1}\left[\Sigma (y_i - \bar{y})^2\right]$

Note:

F- Distribution follows two degree of freedom

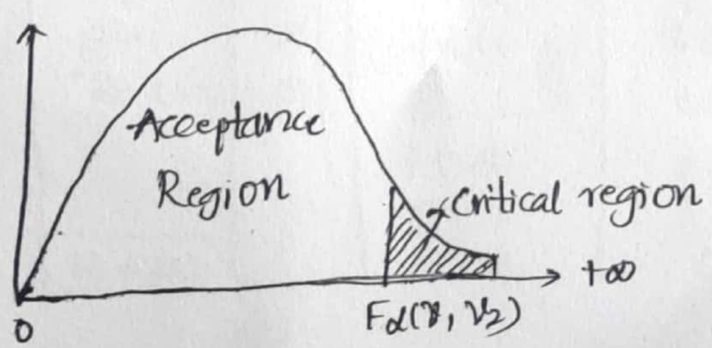
$\nu_1 = n_1 - 1, \quad \nu_2 = n_2 - 1$

$d.f = (\nu_1, \nu_2)$

Properties of F- Distribution:

F- Distribution curve entirely lives in 1st quadrant.

$F_\alpha(\nu_1, \nu_2)$ is the value of F with ν_1 & ν_2 degrees of freedom s/t area under the F- Distribution to right of F_α is α



Note:

① When the SD s_1 & s_2 are given, then the test statistic

$$F_{cal} = \frac{\frac{n_1 s_1^2}{(n_1-1)}}{\frac{n_2 s_2^2}{(n_2-1)}}$$

②) For two tailed test, in F-distribution 'α' is 'α' only.

Q. Two random samples of size 9 & 6, have the following value

Sample-I	15	22	28	26	18	17	29	21	24
Sample-II	8	12	9	16	15	10	-	-	-

Test the difference of estimate of the population variance 5% loss

Sol Given data (box)

$n_1 = 9$ & $n_2 = 6$

$\bar{x} = \frac{\sum x_i}{n} = \frac{200}{9}$

$\boxed{\bar{x} = 22.2}$

$\bar{y} = \frac{\sum y_i}{n_2} = \frac{70}{6}$

$\boxed{\bar{y} = 11.6}$

$\because s_1^2 = \frac{1}{n_1-1}\left[\sum (x_i-\bar{x})^2\right]$ & $s_2^2 = \frac{1}{n_2-1}\left[\sum (y_i-\bar{y})^2\right]$

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	y_i	$(y_i-\bar{y})^2$	$(y_i-\bar{y})$
15	-7.2	51.84	8	12.96	-3.6
22	-0.5	0.25	12	0.16	0.4
28	5.8	33.64	9	6.76	-2.6
26	3.8	14.44	16	19.36	4.4
18	-4.2	17.64	15	11.56	3.4
17	-5.2	27.04	10	2.56	-1.6
29	6.8	46.24			
21	-1.2	1.44			
24	1.8	3.44		$\boxed{\Sigma = 53.36}$	

$s_1^2 = \frac{1}{9-1}[195.5]$ $\quad$ $s_2^2 = \frac{1}{6-1}(53.36)$

$s_1^2 = 24.43$ $\quad$ $s_2^2 = 10.67$

$s_1 = 4.93$ $\quad$ $s_2 = 3.26$

① $H_0: \sigma_1^2 = \sigma_2^2$

② $H_1: \sigma_1^2 \neq \sigma_2^2$ (Two-tailed test)

③ LOS: $\alpha = 5\%$.

$\alpha = 0.05$

$v_1 = n_1 - 1$ $\quad$ $v_2 = n-1$

$v_1 = 9-1$ $\quad$ $v_2 = 6-1$

$v_1 = 8$ $\quad$ $v_2 = 5$

DOF $= (v_1, v_2) = (8, 5)$

t-tab at $\alpha = 5\%$ $= 4.82$ (From F-table)

④ TTS: $F_{cal} = \frac{\text{Greater variance}}{\text{Smaller variance}} = \frac{24.43}{10.67}$

$F_{cal} = 2.28$

⑤ Conclusion:

$\therefore F_{cal} < F_{tab}$

Accept Null hypothesis

*→ Sample of size of 9 and 8 give the sum of squares of deviations from their respective means equal to 160 and 91 respectively. Test the diff of variance of normal distribution (Given $F_{tab} = 3.73$)

So: Given $n_1 = 9$, $n_2 = 8$

$\Sigma(x_i - \bar{x})^2 = 160$

$s_1^2 = \frac{1}{n-1}(\Sigma(x_i - \bar{x})^2)$

$= \frac{160}{9-1} = \frac{160}{8}$

$\boxed{s_1^2 = 20}$

$\Sigma(y_i - \bar{y})^2 = 91$

$s_2^2 = \frac{1}{n_2 - 1}[\Sigma(y_i - \bar{y})^2]$

$= \frac{1}{8-1}[91] = \frac{91}{7}$

$\boxed{s_2^2 = 13}$

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$ (Two-tailed test)

LOS: $\nu_1 = n_1-1$; $\nu = n_2-1$

$(\nu_1, \nu_2) = (9-1, 8-1) = (8, 7)$

$F_{tab} = 3.73$

TTS : $F_{cal} = \frac{20}{13} = 1.53$

Conclusion

$F_{cal} < F_{tab}$

$\therefore$ Accept Null hypothesis

→ The two random samples reveal the following data

sample	size.	Mean	variance
I	16	440 $\bar{x}$	40 s_1^2
II	25	410 $\bar{y}$	42 s_2^2

Test whether the samples drawn from the same normal population

figure this Statement then we have to do 1st Ftest than T test.

Chi- square Test :-

(χ^2-Test) or (χ^2-distribution)

If O_i is a set of oberved frequencies and E_i [i= 1 to n] E_i are expected frequencies then

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{O_i - E_i}{E_i} \right]^2 \text{ with } (n-1) \text{ degree of freedom}$$

→ chi- square is used to test whether the diff b/w observed and expected frequency is significant or not

Note: conditions to apply χ^2-test

1) The sample observation should be independent
2) The tot freq (N) should be large- (atleast 50)
3) No. of theoritical frequency be small ($E_i \leq 10$)

Applications of (chi square test) χ^2-test

- χ^2-test is used to test of goodness of fit
- To test independence of attributes.
- To test homogenity of independent estimate of the population variance and population correlation coefficient

Note:- If χ^2 test in the binomial distribution degree of freedom

$\nu = n-1$

$\Rightarrow$ PD $\nu = n-2$

ND $\nu = n-3$

Q: The no. of automobile accidents per week in a certain communities are as follows 12, 8, 20, 2, 14, 10, 15, 6, 9, 4 are the frequencies in argument with the belief that accident conditions were the same during the 10 week period ($\chi^2_{5\%} = 16.9$)

Sol: $n = 10$

expected freq each week $= \frac{100}{10} = 10$

Step ①: Null hypothesis H_0: The accident conditions is same during these 10 week period

② Alternative hypothesis H_1: Not same

Tabular form

Observed freq O_i	Expec freq E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
12	10	2	4	0.4
8	10	-2	4	0.4
20	10	10	100	1.0
2	10	-8	64	6.4
14	10	4	16	1.6
10	10	0	0	0
15	10	5	25	2.5
6	10	-4	16	1.6
9	10	-1	1	0.1
4	10	-6	36	3.6
$\Sigma O_i = 100$	$\Sigma E_i = 100$			$\Sigma \frac{(O_i - E_i)^2}{E_i} = 26.6$

③ Level of significance:

$\alpha = 5\%$

$\alpha = 0.025$

$$\chi' = \chi'_{\alpha} = \chi'_{5\%} = \chi^2_{0.05} = 16.9$$

from the table with $\nu = n-1$

$\nu = 10-1$

$\boxed{\nu = 9}$.

④ The test-statistic:

$$\chi^2_{cal} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 26.6$$

⑤ Conclusion: $\chi^2 = 26.6$, $\chi^2 tab = 16.9$

$\chi^2_{cal} > \chi^2_{tab}$

So reject NH H_0 at $\alpha = 5\%$.

∴ Accident conditions were not same during this 10 week period.

→ A sample analysis of examination results of 500 students was made. It may found that 200 students had failed, 170 had secured a 3rd class 90 and have place in 2nd class and 20 got a 1st class. Do these figures commensurate with the general examination result which is in ratio of 4:3:2:1 for the various categories ($\chi^2_{5\%} = 7.8\%$)

Step 1:- Null hypothesis H_0: The observed results commensurate with the general examination results

Step 2: H_1: The observed results not commensurate

Given ratio:- 4:3:2:1

N = 500 = Total frequency   sum of the ratio = 10

Categories	Obs freq O_i	Expec freq E_i	$O_i - E_i$	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
Failed	120	$\frac{500 \times 4}{10} = 100$	20	400	$\frac{400}{200} = 2$
3rd class	170	$\frac{500 \times 3}{10} = 150$	20	400	$\frac{400}{150} = 2.66$
2nd class	90	$\frac{500 \times 2}{10} = 100$	-10	100	$\frac{100}{100} = 1$
1st class	20	$\frac{500 \times 1}{10} = 50$	-35	900	$\frac{900}{50} = 18$

Step 3: Level of significance: $\alpha = 5\%$.

$$\chi^2_\alpha = \chi^2_{5\%} = 7.81 \text{ (with } \nu = n-1 = 4-1 = 3)$$

Step 4: The test-statistic: $\chi^2_{cal} = \sum_{i=1}^{n} \frac{[O_i - E_i]^2}{E_i}$

$$\chi^2_{cal} = 23.66$$

Step 5: Conclusion: $\chi^2_{cal} > \chi^2_{tab}$

So reject N.H: H_0

**

A die is thrown 276 times and the results of these thrown are given below.

No. appeared on the die	1	2	3	4	5	6
Frequencies	40	32	29	59	57	59

Test whether the die is biased or not

$(\chi^2_{5\%} = 11.09)$

Sol: $N = 276$

under H_0: The expected frequencies of each digit $= \frac{276}{6} = 46$

Digit	Obs Freq O_i	Exp Freq E_i	$O_i - E_i$	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
1	40	46	-6	36	0.78
2	32	46	-12	144	3.13
3	29	46	-17	289	6.28
4	59	46	13	169	3.67
5	57	46	11	121	2.6
6	59	46	131	169	3.67
	$\Sigma O_i = 276$				

$$\sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 20.13$$

H/w.

→ The theory predicts the proportion of beans in the four graphs G_1, G_2, G_3 and G_4 should be in the ratio 9:3:3:1. In an experienced with 1600 beans the numbers in the four groups were 882, 313, 287 and 118. Does the experimental results support the theories $(\chi^2_{5\%} = 7.8)$

Sol $N = 1600$

Group	Obs Freq O_i	Exp Freq E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
G_1	882	$\frac{1600 \times 9}{16} = 900$			
G_2	313	$\frac{1600 \times 3}{16} = 300$			
G_3	287	$\frac{1600 \times 3}{16} = 300$			
G_4	118	$\frac{1600 \times 1}{16} = 100$			

$\Sigma O_i = 1600 \quad \Sigma E_i = 1600$

→ Fit the poisson distribution to the following data and test the goodness of fit at 5% level of significance

x_i	0	1	2	3	4
$f(x)$ obs	419	352	154	56	19
Exp E_i	406	366	165	49	11

Mean $= \lambda = \frac{1}{N}(\Sigma x_i f_i x_i)$

$\boxed{\lambda = 0.9}$

$P(0) = \frac{e^{-0.9}(0.9)^0}{0!}$

$P(0) = 0.406 \quad \Rightarrow f(0) = N \times P(0) = 406$

~~f(x) = 1000 × P(x)~~

$= \cancel{1000 \times 2.45}$

$f(1) = N(P(1))$

$= 1000 \times \frac{e^{-0.9} (0.9)^1}{1!} = 366$

$f(2) = 165$ $\quad f(3) = 49$ $\quad f(4) = 11$

x	0	1	2	3	4
$f(x)\ O_i$	419	352	154	56	19
E_i	406	366	165	49	11