24th NOV 2022.

Mathematics for CS Engineers.

Page:
Date:

Syllabus.

Module 3 : Statistics

Module 4 : Probability Distribution.

Module 5 : Simplicity Distribution and optimisation techniques.

Module 1 : Vector spaces.

Module 2 : Eigen value and Eigen vectors.

variance = (standard deviation)^2.

Module 3 : Statistics.

mean : $\bar{x} = \dfrac{\Sigma x}{n} = \dfrac{x_1 + x_2 + x_3 \dots x_n}{n}$.

variance = $\dfrac{\sum_{i=1}^{n} (x-\bar{x})^2}{n}$ $\quad \dfrac{1}{n} \Sigma\, x^2 + \bar{x}^2 - 2x\bar{x}$. $\quad \bar{x}$ = constant

$\dfrac{1}{n}\left[\Sigma x^2 + n\bar{x}^2 - 2\bar{x}\Sigma x\right]$ $\quad \Sigma\bar{x} = n = \bar{x}\Sigma 1$ as $\sum^{n}$ constant = n.

here $\bar{x} = \dfrac{\Sigma x}{n}$

$\dfrac{1}{n}\left[\Sigma x^2 + \cancel{n}\dfrac{(\Sigma x)^2}{n^{\cancel{2}}} - 2\dfrac{\Sigma x\,\Sigma x}{n}\right.$

$\dfrac{1}{n}\left[\Sigma x^2 - \dfrac{(\Sigma x)^2}{n}\right] = \dfrac{\Sigma x^2}{n} - (\bar{x})^2$ $\hookrightarrow$ 2nd form.

Combined mean of 2 data sets.

$n_1 \rightarrow$ series 1 — $\bar{x}_1$ (mean for series 1).

$n_2 \rightarrow$ series 2 — $\bar{x}_2$ (mean for series 2).

$\therefore$ combined mean = $\dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$.

Standard deviation for combination of 2 series/groups.

$(n_1+n_2)\sigma^2 = n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 D_1^2 + n_2 D_2^2$

Here n_1 — stan. dev = σ_1 $\quad D_1 = \bar{x}_1$ − combined mean.

n_2 − stan. dev = σ_2. $\quad D_2 = \bar{x}_2$ − combined mean

$[D_i = m_i - M]$

m_i = mean of i series $\quad M$ = combined mean.

median : data in ascending/desending order.

1) If no. of terms are odd, median = $\frac{n+1}{2}$ th term.

2) If no of terms are even, median = $\frac{\left(\frac{n}{2}\right)\text{term} + \left(\frac{n}{2}+1\right)\text{term}}{2}$.

median for grouped data:

median = $L + \frac{h}{f}\left(\frac{N}{2} - c\right)$

L - lower limit of median class.

h - width of median class.

f - frequency of median class.

N - total frequency

C - Cumulative frequency up to class preceding median class.

standard deviation.
grouped

$$\frac{\sum f(x_i-\bar{x})^2}{\sum f}$$

Mode: The value of variable which occurs most frequently in data or the value of maximum frequency.

* while mean, median are singular - mode for data could be more than one value.

Mode for grouped data: $L + \frac{H \cdot f_1}{f_1+f_2}$.

L - lower limit of the mode class.

H - width of modal class.

f_1 - excess of modal frequency over freq. of preceding class.

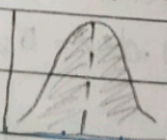
f_2 - excess of modal frequency over followin[g] class.

Median can be applied even for data sets with extreme values. while mean for the same is accurate.

for mean - a certain range does exist.

median - applicable to extreme values (without a specified ra[nge])

distributions (normal/gaussian) are symmetric about mean



Mean

Camlin

Page :
Date :

Q1> Find mean, median and mode for the data set

Value.	5	10	15	20	25	30	35	40	45
freq.	29	224	465	582	634	644	650	653	655.

mean = $\bar{x} = \frac{\Sigma f x}{\Sigma f} = \frac{134515}{4536.} = 29.65$

median =

X	f.	Cf.
5	29	29.
10	224	253
15	465	718
20	582	1300
25	634	1934
30	644	2578
35	650	3228
40	653	3881
45	655.	4536.

$\frac{N}{2}$ term = $\frac{4536}{2.} = 2268.$

✓ median class = 25−30.

from Cf ∴ take value below

=> median 30.

mode = 45 (has highest freq.).

Q2>

	X (mid value)	f.	Cf
0−10	5	7	7
10−20	15	8	$\boxed{15}$ — median class.
$\boxed{20-30}$	25	20 (mode)	$\boxed{35}$
30−40	35	10	45
40−50.	45.	5	50.

mean = $\frac{\Sigma f X}{\Sigma f} = \frac{1230}{50.} = 24.6.$

median = $\frac{N}{2} = \frac{50}{2} = 25.$

∴ median interval = 20−30.

median = $20 + \frac{10}{20}\left(25-15\right) = 20 + \frac{1}{2}(10) = 25.$

Camlin

$\text{mode} = L + \frac{hf_1}{f_1+f_2}$ = mode class = 20-30.

$f_1 = 20-8$.

$f_2 = 20-10$.

$$\text{mode} = 20 + \frac{10(20-8)}{(20-10)+(20-8)}.$$

$$= 20 + \frac{12(10)}{12+10} = 20 + \frac{120}{22} = 25.45.$$

Q) The median and mode are given to be Rs 25 and 24 respectively. Calculate the missing frequency

Class	Freq.	x_i	cf
0-10	14	5	14
10-20	x	15	14+x
(20-30)	$\boxed{27}$	$\boxed{25}$	41+x.
30-40	y	35	41+x+y.
40-50	15.	45	56+x+y.

median class & mode class.

$$\text{median} = L + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

$$25 = 20 + \frac{10}{27}\left(\frac{56+x+y}{2} - 14+x\right)$$

$$5 = \frac{10}{27}\left(\frac{56-28-x+y}{2}\right).$$

$$27 = 28 - x + y.$$

$$\boxed{x - y = 1} \text{ — (1)}.$$

$$\text{mode} = L + \frac{hf_1}{f_1+f_2}.$$

$$24 = 20 + \frac{10(27-x)}{(27-x)+(27-y)}$$

$$4 = \frac{10(27-x)}{(27-x)+(27-y)}$$

$$4 = \frac{10(27-x)}{54-x-y}.$$

$$216 - 4x - 4y = 270 - 10x$$

$$6x - 4y = 54.$$

$$3x - 2y = 27$$

$$(x - y = 1)2.$$

$$3x - 2y = 27$$

$$\underline{{}_{(-)}2x \mathbin{\underset{(+)}{-}} 2y = \underset{(-)}{2}}$$

$$x = 25.$$

$$y = 24.$$

∴ Missing frequencies = x = 25.

y = 24.

MODE CLASS will be subsequent class of the class having highest frequency.

Camlin

8) Median = 33.5. } Total frequency = 230.
mode = 34.
find x, y, z

Median & mode lies b/w 30-40
⇒ median mode class - 30-40

Class	Freq.	Cf
0 – 10	4	4
10 – 20	16	20
20 – 30	x	20+x
30 – 40	y	20+x+y
40 – 50	z	20+x+y+z
50 – 60	6	26+x+y+z
60 – 70	4	30+x+y+z

Median & mode class

1) $30+x+y+z = 230$
$\Rightarrow x+y+z = 200.$

2) Median $= L + \frac{h}{f}\left(\frac{N}{2} - c\right)$.

$33.5 = 30 + \frac{10}{y}\left(\frac{30+x+y+z - 40 - 2x}{2}\right)$

$3.5 = \frac{5}{y}(-10-x+y+z)$.

$\frac{3.5y}{5} = -10-x+y+z$.

$0.7y = -10-x+y+z$.

$0.3y+z-x = 10$.

$3.5 = \frac{5}{y}(230-40-2x)$.

$\frac{3.5y}{5} = 190-2x$.

$0.7y = 190-2x$.

$0.7y+2x = 190$.

x = 60
z = 40
y = 100

3) mode $= L + \frac{hf_1}{f_1+f_2} = 30 + \frac{10(y-x)}{y-x+y-z} = 34.$

$\frac{10(y-x)}{2y-x-z} = 4.$

$\Rightarrow 10y-10x = 8y-4x-4z.$

$2y-6x+4z = 0.$

$y-3x+2z = 0.$

$y = 3x-2z.$

$x+z+3x-2z = 200.$

$4x-z = 200.$

$z = 4x-200.$

$x+y+z = 200.$
$y = 3x+2z$
$z = 3x - y/2.$

95-y
$\frac{3.5y}{10} =$
$3.5y = 950-10x.$
$3.5y+10x = 950$
$\frac{230}{2}$
95-y
115-20-x

Camlin

Q) For a group of 200 students, the mean, st. deviation of scores were found to be 40 & 15. Later on it was discovered the score 43 and 35 mentioned - was missed as 34 and 53.
Find the corrected st. deviation.

for 200 students

mean = 40.

$\frac{\Sigma x}{200} = 40.$

$\Sigma x = 40(200).$

$\Sigma x = 8000.$

corrected value

$= 8000 - (43 - 35) + (34 + 53).$ remove incorrect, add correct.

$= 8009.$

σ^2 = st. dev.

$\sigma^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2.$

$[(15)^2 + (40)^2](200) = \Sigma x^2.$

$\Sigma x^2 = 365000$ (Before correction)

corrected value of Σx^2.

$= 365000 - 43^2 - 35^2 + 34^2 + 53^2.$

$= 365891.$

corrected st. dev $= \frac{\Sigma x^2}{n} - (\bar{x})^2_{\text{correct}} =$

$= \frac{365891}{200} - \left(\frac{8009}{200}\right)^2 = 225.8529\ (\sigma^2).$

$\sigma_{\text{correct}} = 15.0284.\ (\sqrt{\sigma^2}.)$

Q). The scores obtained by 2 batsmen A and B in 10 matches are given below. Calculate mean, st. dev and coeff. of variance. Determine who is more efficient and who is more consistant

A -	30	44	66	62	60	34	80	46	20	38
B -	34	46	70	38	55	48	60	34	45	30

n = 10.

$\bar{X}_A = 48.$ $\bar{X}_B = 46.$ $\sigma_A^2 = \frac{\Sigma x_A^2}{n} - (\bar{X}_A)^2.$

$\Sigma x = 480$ $\Sigma x = 460$

$\Sigma x^2 = 26152.$ $\Sigma x^2 = 22626.$ $\sigma_A^2 = 311.2$ $\sigma_B^2 = 146.6.$

$\sigma_A = 17.64$ $\sigma_B = 12.10.$

coefficient of variance = $\frac{\text{std.dev}}{\text{mean}} = \sigma/\bar{x}$

$COV_A = 0.3675 = 36.75\%$

$COV_B = 0.2630 = 26.3\%$ (cov ∝ 1/constancy)

If coefficient of variance is less then data set is more consistent

as $COV_B < COV_A$.

* Player B – more consistent

Efficiency determined by mean (mean ∝ efficiency)

$\bar{X}_A > \bar{X}_B$

* Player A – more efficient

Q) An analysis of monthly wages paid to worker of two company's A & B belonging to the same industry gives the results.

Company	Worker	Mean monthly wages	Variance
A	500	186	81
B	600	175.	100.

a) which company has larger wages bill.

b) Greater variability in individual wages.

c) Calculate mean and st. dev of wages of all worker in the company taken together (A+B).

a) $\bar{X}_A = \frac{\Sigma x_A}{n_A}$

$\Sigma X_A = 186(500) = 93{,}000$

$\Sigma X_B = 175(600) = 1{,}05{,}000$

company B.

b) $COV_A = \bar{X}_A / \text{std.dev}_A$

$= \frac{9}{186}$

$= 0.04838$

$= 4.838\%$

$COV_B = \frac{10}{175} = 5.714\%$

$\uparrow COV \Rightarrow \uparrow$ vari

$\therefore COV_B > COV_A$

$\Rightarrow$ company B.

c) mean combined $= \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1+n_2} = 180$.

$D_1^2 = 36 \quad D_2^2 = 25$.

$(n_1+n_2)\sigma^2 = n_1\sigma_1^2 + n_2\sigma_2^2 + n_1D_1^2 + n_2D_2^2 = 1{,}33{,}500$.

$\sigma^2 = 133500/500+600 = 121.3636$, $\sigma = 11.0165$.

2nd December 2022.

Correlation: Defines the degree of relationship between 2 data sets.
denoted as r.
r lies b/w $-1 \le r \le 1$.

* for linear relationships i.e. if data sets are linearly dependent on one another - the correlation is a positive value.

* for non linear relationships i.e. if data sets are inversely dependent on one another - negative correlation.

* If the data sets are not related - the value of r becomes 0.

$r = 0$ no correlation
$r = +1$ perfect positive correlation.
$r = -1$ perfect negative correlation.

Ref: B.V. Ramanna

formula. $r = \dfrac{\text{covariance}(x,y)}{\sigma_x \sigma_y}$ (variance calculated for 2 different series)

$$\text{covariance}(X,Y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}.$$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\sigma_x \sigma_y}. \quad [r = \text{dimensionless}].$$

Karl Pearson coeff. of correlation.

$$r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{(n\Sigma X^2 - (\Sigma X)^2)(n\Sigma Y^2 - (\Sigma Y)^2)}}.$$

5th dec.

Q) find the covariance of the following pairs of observations of two variables.
(1,4), (2,2), (3,4), (4,8), (5,9), (6,12).

X : 1 2 3 4 5 6. $\bar{X} = 3.5$.
Y : 4 2 4 8 9 12. $\bar{Y} = 6.5$.

$\cancel{\dfrac{\Sigma(X-\bar{X})(Y-\bar{Y})}{n} = \dfrac{(-2.5)\ldots}{6} = \dfrac{2039.337158}{6} = 339.889}$

$X-\bar{X}$: -2.5, -1.5, -0.5, 0.5, 1.5, 2.5.
$Y-\bar{Y}$: -2.5 -4.5, -2.5, 1.5, 2.5, 5.5.
$(X-\bar{X})(Y-\bar{Y}) =$ 6.25, 6.75, 1.25, 0.75, 3.75, 13.75.
$\Sigma(X-\bar{X})(Y-\bar{Y}) = 32.5$.
covariance $= \dfrac{32.5}{6} = 5.4166$.

Q) Find the covariance.

x	y	$(x-\bar{x})$	$(y-\bar{y})$
98	15	9.67	3.34
87	12	−1.33	0.34
90	10	1.67	−1.66
85	10	−3.33	−1.66
95	16	6.67	4.34
75	7	−13.33	−4.66

$\Sigma(x-\bar{x})(y-\bar{y}) = 125.6662$

covariance = 20.944.

$\bar{x} = 88.33$

$\bar{y} = 11.66$

$$\frac{\Sigma(x-\bar{x})(y-\bar{y})}{n} = \frac{1}{n}\left[\Sigma xy - \Sigma x\bar{y} - \Sigma\bar{x}y + \Sigma\bar{x}\bar{y}\right]$$

$$= \frac{1}{n}\left[\Sigma xy - \bar{y}\Sigma x - \bar{x}\Sigma y + n\bar{x}\bar{y}\right]$$

Use 2 variable calculation

$$\frac{1}{n}\left[6309 - 11.66(530) - 88.33(70) + 6(88.33)(11.66)\right] = 20.9$$

(or) $$\frac{1}{n}\left[\Sigma xy - \frac{\Sigma y\,\Sigma x}{n} - \frac{\Sigma x\,\Sigma y}{n} + \cancel{n}\frac{\Sigma x}{\cancel{n}}\frac{\Sigma y}{n}\right]$$

$$= \frac{1}{n}\left[\Sigma xy - \frac{\Sigma y\,\Sigma x}{n}\right]$$

$$= \frac{1}{n}\Sigma xy - \frac{\Sigma x\,\Sigma y}{n^2} \qquad \frac{1(6309)}{6} - \frac{70(530)}{36}$$

6th dec.

- Normalised value → bringing the value of a quantity between the va[lue] of $-1 \le \text{value} \le 1$.
- r is normalised by dividing covariance$(x,y)/\sigma_x\sigma_y$.
- Correlation coeff. is independent of change of origin and scale.

$$u = \frac{x-x_0}{h} \quad v = \frac{y-y_0}{k} \quad \Big]\ \text{shift of origin}$$

$\Rightarrow r(x,y) = r(u,v)$.

Establish the formula (derivation)

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

ekumbh.aicte.india

let $x - y = z$

$\sigma_{x-y}^2 = \sigma_z^2 = \dfrac{\Sigma(z-\bar{z})^2}{n}$

here $\bar{z} = \overline{x-y} = \bar{x} - \bar{y}$

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

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

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

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

$\dfrac{\Sigma(z-\bar{z})^2}{n} = \sigma_x^2 + \sigma_y^2 - 2\,\text{covariance}(x,y)$

$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2\,\text{covariance}(x,y)$.

$\sigma_x^2 + \sigma_y^2 - \sigma_z^2 = 2\,\text{covariance}(x,y)$.

as covariance $= r\sigma_x\sigma_y$.

$$\therefore r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x\sigma_y}.$$

(Q) Find the correlation coefficient b/w x and y from the data.

$n = 8$.

Use calculation

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

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

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

$r = 0.948786$.

- feed data.
- go to AC.
- OPTN
- scroll down
- opt 1 summation.

num = 12728

$\sqrt{1799763136} = 13415.033$

Linear Regression :-

8 Dec 2022. finding the mathematical relationship b/w 2 variables.

linear regression Y on X $\longrightarrow$ X independent - Y depends on X.

Regression of Y on X $-(Y-\bar{Y}) = b_{yx}(X-\bar{X})$.

$\bar{Y}$ = mean of y series.

$\bar{X}$ - mean of x series. $\qquad Y = AX + B$

b_{yx} - regression coeff. of Y on X = $\dfrac{r\sigma_y}{\sigma_x}$.

regression of X on Y = $(X-\bar{X}) = b_{xy}(Y-\bar{Y})$. $\qquad X = CY + D$.

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

13 Dec.

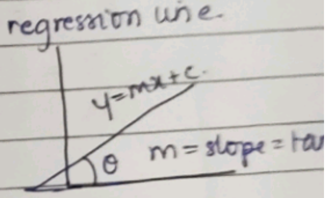
Q) If θ is the angle between the two regression lines relating variable X and Y, show that $\tan\theta = \dfrac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\left(\dfrac{1-r^2}{r}\right)$.

$\theta = \theta_2 - \theta_1$

θ_2 - angle by regression line X on Y then $X-\bar{X} = b_{xy}(Y-\bar{Y})$

here θ_2 for line with eqn $y = mx+c$ or $x = cy+d$.

$\theta_2 = \tan^{-1}(m)$ or $\tan^{-1}(c)$



X on Y $\longrightarrow$ $X-\bar{X} = b_{xy}(Y-\bar{Y})$.

$$\Rightarrow Y-\bar{Y} = \frac{1}{b_{xy}}(X-\bar{X})$$

$$Y = mx.$$

$$\Rightarrow \theta_2 = \tan^{-1}m = \tan^{-1}\left(\frac{1}{b_{xy}}\right) = \tan^{-1}\left(\frac{\sigma_y}{r\sigma_x}\right).$$

$\theta_1 \rightarrow$ for line Y on X $\quad Y-\bar{Y} = b_{yx}(X-\bar{X})$

$$Y = mx.$$

$$\Rightarrow \theta_1 = \tan^{-1}(b_{yx}) = \tan^{-1}\left(\frac{r\sigma_y}{\sigma_x}\right).$$

$$\Rightarrow \theta = \theta_2 - \theta_1$$

$$= \tan^{-1}\left(\frac{\sigma_y}{r\sigma_x}\right) - \tan^{-1}\left(\frac{r\sigma_y}{\sigma_x}\right)$$

$$\tan^{-1}A - \tan^{-1}B = \tan^{-1}\left(\frac{A-B}{1+AB}\right).$$

$$\theta = \tan^{-1}\left[\frac{\sigma_y}{r\sigma_x} - \frac{r\sigma_y}{\sigma_x} \Big/ 1 + \frac{\sigma_y}{\cancel{r}\sigma_x} \times \frac{\cancel{r}\sigma_y}{\sigma_x} \right]$$

$$\theta = \tan^{-1}\left[\frac{\dfrac{\sigma_y\sigma_y - r^2\sigma_y\sigma_x}{r\sigma_x^{\cancel{2}}}}{\dfrac{\sigma^2_x + \sigma^2_y}{\sigma^{\cancel{2}}_x}} \right].$$

$$\theta = \tan^{-1}\left[\frac{\sigma_y\sigma_x(1-r^2)}{r(\sigma^2_x+\sigma^2_y)} \right]$$

$$\Rightarrow \tan\theta = \frac{\sigma_y\sigma_x}{(\sigma^2_x+\sigma^2_y)} \times \frac{(1-r^2)}{r}.$$

Q) Find the correlation coeff. given $\sigma_y = 2\sigma_x$ and $\theta = \tan^{-1}\frac{3}{5}$.

$$\tan\theta = \frac{\sigma_x\sigma_y}{\sigma^2_x+\sigma^2_y}\left(\frac{1-r^2}{r}\right).$$

$$\frac{3}{5} = \frac{\sigma_x(2\sigma_x)}{\sigma^2_x+4\sigma^2_x}\left(\frac{1-r^2}{r}\right)$$

$$\frac{3}{\cancel{5}} = \frac{2\cancel{\sigma^2_x}}{\cancel{5}\cancel{\sigma^2_x}}\left(\frac{1-r^2}{r}\right)$$

$$\frac{3}{2} = \frac{1-r^2}{r}. \qquad 3r = 2-2r^2.$$

$$2r^2+3r-2=0.$$

$$r = 0.5,\ -2.$$

as $r \rightarrow -1 \leq r \leq 1$

$$\Rightarrow r = 0.5.$$

13 Dec 2022.

module 4 : probability distribution.

Sample space = The set of all possible outcomes of a random event / experiment is called sample space.

Random variable = Variable which can not be predicted accurately.

Event = subset of sample space.

Total sum of probability is always less than or equal to 1.

S = 2 ^ (no of obj) / outcomes (coin) example : tossing a single coin = $S = 2^1 = \{H, T\}$.

tossing two coins = $2^2 = 4 = \{HT, TH, HH, TT\}$.

S = 6 ^ (no. of obj) (dice) throwing 2 dies = 6^2 — no of dice.

↓ outcomes from a dice

$$\text{probability} = \frac{\text{number of occurrances of given event}}{\text{total no. of occurrances.}}$$

discrete random variable = If random variable takes finite or countably infinite number of values, then it is called discrete random variable.

Ex. tossing a coin and observing the outcomes.

$S = \{H, T\}$ where H value = 1 ; t value = 0.

∴ range of probability = $\{0, 1\} = X$.

tossing two coins.

$S = \{HH, HT, TH, TT\}$.

probability of 0 Head = $\{TT\} = 0.25$.

probability of Head occurrance = either 0 times $\{TT\}$

either 1 time $\{HT, TH\}$.

either 2 times $\{HH\}$.

=> range of head / Tail = $\{0, 1, 2\}$.

tossing 3 coins.

=) range = $\{0, 1, 2, 3\}$.

TTT	TTH, THT, HTT	HHT, THH, HTH	HHH
// head / tail occurs 0 times /	1 time /	2 times /	3 time

(head case).

* Discrete Random variable – can only be integer values within specified range.
* Continuous Random variable – any value (real values) within specified range.

Discrete Probability Distribution

For each value x_i of a discrete random variable X, we assign a real number $p(x_i)$ such that 1) $p(x_i) \geq 0$ 2) $\Sigma p(x_i) = 1$ then function $p(x_i)$ is called propability function. If the probability that X takes the values x_i is p_i or $p(x_i)$ the $P(X = x_i) = p(x_i)$ or p_i. Then $P(X)$ is probability density function (pdf) or probability mass function (pmf)

mean $(\mu) = \Sigma x_i p(x_i)$

variance $(V) = \Sigma (x_i - \mu)^2 p(x_i) = \Sigma x_i^2 p(x_i) - \mu^2$

Standard Deviation $\sigma = \sqrt{V}$.

8) Find k such that the following distribution represents a finite probability distribution. Hence find the mean, variance, $p(x \leq 1)$, $p(x > 1)$, $p(-1 < x \leq 2)$

x	-3	-2	-1	0	1	2	3.
$p(x)$	k	$2k$	$3k$	$4k$	$3k$	$2k$	k.

$k + 2k + 3k + 4k + 3k + 2k + k = \underline{1}$.

$16k = 1$.

$\Rightarrow k = 1/16$.

mean $= \Sigma x_i p(x_i)$

$= \frac{-3}{16} + \frac{(-4)}{16} - \frac{3}{16} + 0 + \frac{3}{16} + \frac{4}{16} + \frac{3}{16} = 0$.

variance $= \Sigma x_i^2 p(x_i) - \cancel{\mu^2}\, 0$.

$= \Sigma x_i^2 p(x_i) = 5/2$.

Std. dev $= \sqrt{\frac{5}{2}} = 1.5811$.

$p(x \leq 1) = P\{-3, -2, -1, 0, 1\}$

$= \frac{1}{16} + \frac{2}{16} + \frac{3}{16} + \frac{4}{16} + \frac{3}{16} = \frac{13}{16}$ //.

$P(x > 1) = P\{2, 3\}$

$= 3k = 3/16$.

$P(-1 \leq x \leq 2) = P\{0, 1, 2\}$

$= 4k + 3k + 2k$.

$= 9/16$.

$9^9 - 9^8$

$9^8(9-1)$

$= 9^8(8)$

Camlin

Q) 1) mean 4) $p(x>6)$
2) variance 5) k.
3) $P(x<6)$

Page :
Date :

X	0	1	2	3	4	5	6	7.
P(X)	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$.

sum $P(X) = 1$

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

$10k^2 + 9k - 1 = 0.$

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

$\therefore k = 1/10.$

mean $= \frac{1}{10} + \frac{4}{10} + \frac{6}{10} + \frac{12}{10} + \frac{5}{100} + \frac{12}{100} + \frac{49}{100} + \frac{7}{10}.$

$= \frac{30}{10} + \frac{66}{100} = \frac{183}{50} = 3.66.$

variance $= \Sigma x_i^2 p(x_i) - \mu^2.$

$= \left(\frac{1}{10} + \frac{8}{10} + \frac{18}{10} + \frac{48}{10} + \frac{25}{100} + \frac{72}{100} + \frac{343}{100} + \frac{49}{10}\right) - (3.66)^2.$

$= 16.8 - 13.36.$

$= 3.41.$ $\sigma = 1.846.$

$P(x<6) = P(0,1,2,3,4,5)$

$= k + 2k + 2k + 3k + k^2.$

$= 8k + k^2.$

$= 8/10 + 1/100.$

$= 0.81.$

$P(x>6) = P(7)$

$= 7k^2 + k.$

$= \frac{7}{100} + \frac{1}{10} = 0.17.$

CIE-I → Probability Distribution (module 3) and Statistics (module 4)

Page :
Date :

Geometric Distribution

Probability – can yield 2 values – like success / failure.

The total sum of success and failures must be 1.

$\therefore$ P(success rate / probability) + Q(failure probability) = 1

or $P = 1-Q$; $Q = 1-P$.

Geometric Distribution is used to find when will the I^{st} success value occurance.

$p(x) = q^x p \quad x = 0, 1, 2, 3 \ldots n$

$x = 0 \quad p(x) = q^0 p = p$

$x = 1 \quad p(x) = qp$.

as $q = 1-p$.

$\Rightarrow p(x) = (1-p)^x p$.

similarly $p(x) = q^{x-1} p$.

$\Rightarrow x = 1, 2, 3, 4 \ldots n$

If p be the probability of success and x be the number of failures preceding the I^{st} success, then this distribution is given by

$p(X = x) = q^x p$

Properties i) $p(x) > 0$ for $0 \le p \le 1$.

ii) total sum of probability $= \sum_0^\infty p(x)$

$= p + qp + q^2 p + q^3 p \ldots q^n p \ldots \infty$.

$= p\,[\underbrace{1 + q + q^2 + q^3 \ldots \infty}]$.

$= p\left(\frac{1}{1-q}\right)$

sum of infinite series $= a/1-r$

$r = q$.

$a = 1$.

but $1-q = p$.

$\Rightarrow \cancel{p}\left(\frac{1}{\cancel{p}}\right) = 1$.

1) mean $(\mu) = \sum x\,P(x) = \sum x(q^x p)$

as p is constant $\Rightarrow p\sum_0^\infty x q^x$.

$\Rightarrow p\sum_1^\infty x q^{x-1} \cdot q$.

$\Rightarrow pq \sum_1^\infty x q^{x-1}$.

$\Rightarrow pq \sum_1^\infty d/dq\, q^x$ //

$\cong pq \frac{d}{dq}\left(\underbrace{\sum_1^\infty q^x}_{\text{geometric progression upto } \infty}\right) = pq \frac{d}{dq}\left[\frac{q}{1-q}\right]$.

Camlin

$= pq\left[\frac{1(1-q) - q(-1)}{(1-q)^2}\right] = pq\left[\frac{1-q+q}{(1-q)^2}\right] = \frac{pq}{(1-q)^2}$

Page :
Date :

$1-q=p \Rightarrow \frac{pq}{p^2} = \frac{q}{p}$ //.

for $p(x) = q^x p$ mean$(\mu) = \frac{q}{p}$.

- $p(x) = q^{x-1}p$ mean$(\mu) = 1/p$.

Variance $= \sum x^2 p(x) - \mu^2$.

$\sum x^2 (q^x p) - \left(\frac{q}{p}\right)^2$ or $\sum_1^\infty x^2(q^{x-1}p) - \frac{1}{p^2}$

$\frac{d}{dq}(q^x) = xq^{x-1}$

$\frac{d^2}{dq^2}(xq^{x-1}) = x(x-1)q^{x-2}$.

to substitute for x^2.

$\frac{d^2}{dq^2} = (x^2-x)q^{x-2}$.

For $x^2 \Rightarrow x(x-1)+x$.

$V = p\sum_1^\infty [x(x-1)+x]q^{x-1} - \frac{1}{p^2}$.

mean exp $= 1/p$.

$= p\sum_1^\infty x(x-1)q^{x-1} + p\sum_1^\infty xq^{x-1} - \frac{1}{p^2}$.

$= qp\sum_1^\infty x(x-1)q^{x-2} + \frac{1}{p} - \frac{1}{p^2}$.

$= qp\sum_2^\infty \frac{d^2}{dq^2}(q^x) + \frac{p^2-1}{p^3}$.

$= \frac{d^2}{dq^2} qp \sum_2^\infty q^x + \frac{1}{p} - \frac{1}{p^2}$. → sum of infinite series

$= \frac{d^2}{dq^2} qp\left(\frac{q^2}{1-q}\right) + \frac{1}{p} - \frac{1}{p^2}$.

$= \cancel{qp\left(\frac{pq}{p^7}\right) + \frac{1}{p} - \frac{1}{p^2} = q^2}$

$$\text{i)}\ \frac{d}{dq}\left(\frac{q^2}{1-q}\right) = \frac{2q(1-q) - q^2(-1)}{(1-q)^2} = \frac{2q - 2q^2 + q^2}{(1-q)^2}$$

$$= \frac{2q - q^2}{(1-q)^2}$$

$$\frac{d^2}{dq^2}\left(\frac{q^2}{1-q}\right) = \frac{d}{dq}\left(\frac{2q - q^2}{(1-q)^2}\right) = \frac{(2-2q)(1-q)^2 - 2(1-q)(-1)(2q-q^2)}{(1-q)^4}$$

$$= \frac{(2-2q)(1+q^2-2q) + 2(1-q)(2q-q^2)}{(1-q)^4}$$

$$= \frac{2 + \cancel{2q^2} - 4q - 2q - \cancel{2q^3} + 4q^2 + \cancel{4q} - \cancel{2q^2} - \cancel{4q} + \cancel{2q^3}}{(1-q)^4}$$

$$= \frac{4q^2 - 6q + 2}{(1-q)^4} = \frac{2(2q^2 - 3q + 1)}{(1-q)^4}$$

$$= \frac{2(2q^2 - 2q - q + 1)}{(1-q)^4} = \frac{2(2q(q-1) - q(q-1))}{(1-q)^4}$$

$$= \frac{2((q-1)(2q-q))}{(1-q)^4} = \frac{-2((1-q)(2q-1))}{(1-q)^3}$$

$$= \cancel{\frac{-4q+2}{(1-q)^3}}$$

$$= \frac{2(1-q)(1-q)^2 + 2(1-q)q(2-q)}{(1-q)^4}$$

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

$$= \frac{2(1-q)^3 + 2q(2-q)}{(1-q)^3}$$

$$= 2 + \frac{2q(2-q)}{(1-q)^3}$$

or on simplification.

$$D_i = \frac{2}{p^3}$$

$$V = \cancel{p}q\left(\frac{2}{p^2}\right) + \frac{1}{p} - \frac{1}{p^2} = \frac{2q}{p^2} - \frac{1}{p^2} + \frac{1}{p} \quad \frac{2q-1+p}{p^2}$$

$\frac{2(1-p)-1+p}{p^2}$ as $q = 1-p$.

$\Rightarrow 2-2p-1+p/p^2 = \frac{1-p}{p^2} = \frac{q}{p^2}$.

Q). 3% of product produced by machine is found to be defective. Find the I st defective occurs in the 1) 5th item inspected
2) I st five inspected 3) mean 4) variance.

As I st defective is to be found - geometric probability distribution

$P(X) = q^{x-1}p \quad x = 1, 2, 3 \ldots \infty$.

$P(X) = $ / probability of defective $= 3\% = \frac{3}{100} = p$.

$q = 1-p = 0.97$.

$P(X) = (0.97)^{x-1} p$.

1) 5th item $\Rightarrow x = 5$. $\quad P(5) = (0.97)^{5-1}(0.03)$

$= (0.97)^4 (0.03)$.

$= 0.0265$.

2) I st five inspected.

$= \cancel{\sum_{1}^{5} (0.97)^{x-1}(0.03)} = \sum_{1}^{5} q^{x-1} p$.

$= 0.03 + 0.0291 + 0.028227 + 0.02738 + 0.02655$.

$= 0.141257$.

3) mean $= \frac{q}{p} = \frac{1}{0.03} = \frac{10}{3} = 3.33\ldots$

4) Variance $= \frac{q}{p^2} = \frac{0.97}{(0.03)^2} = 1077.77$.

Poisson Distribution.

Page :
Date :

It is discrete probability distribution with no upper bonds and its p.d.f is given by.

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

Here m = parameter of distribution.

Posion Distribution - special limiting case of binomial distribution when number of event is infinite but probability is very less.

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

$np = m$ - finite quantity.

Ex. no of vehicles produced $\to \infty$.

probability of accidents - less $\to 0$.

Total sum of probability $= 1$.

PROOF

$$\sum_{x=0}^{\infty} P(x) = \sum_{x=0}^{\infty} \frac{e^{-m} m^x}{x!} = e^{m} \left[\frac{m^0}{0!} + \frac{m}{1} + \frac{m^2}{2!} + \frac{m^3}{3!} \ldots\ldots \right]$$

$$= e^{m} \cdot e^{-m} = e^0 = 1.$$

$$\text{mean}(\mu) = \sum_0^{\infty} x P(x) = \sum_0^{\infty} x \cdot \frac{e^{-m} \cdot m^x}{x!} = \sum_0^{\infty} \frac{\cancel{x} \cdot e^{-m} \cdot m^x}{\cancel{x}(x-1)!}$$

$$\sum_0^{\infty} \frac{e^{-m} m^x}{(x-1)!} = e^{-m} \left[\sum_0^{\infty} \frac{m \, m^{(x-1)}}{(x-1)!} \right] \Rightarrow e^{-m} \cdot m \left[\sum \frac{m^{(x-1)}}{(x-1)!} \right]$$

$$= \cancel{e^{-m}} \cdot m \cdot \cancel{e^{m}} = m.$$

$\to e^m$ series.

$$\text{Variance}(V) = \sum_0^{\infty} x^2 P(x) - \mu^2 = \sum_0^{\infty} x^2 \frac{e^{-m} m^x}{x!} - m^2.$$

other way to derive x^2.

$$\sum_0^{\infty} \frac{[x(x-1) + x] \, e^{-m} m^x}{x!} - m^2$$

$$= e^{-m} \left[\sum_0^{\infty} \frac{x(x-1) m^x}{x!} + \sum_0^{\infty} \frac{x m^x}{x!} \right] - m^2.$$

$$e^{-m}\left[\sum_{2}^{\infty}\frac{m^x}{(x-2)!} + \sum_{1}^{\infty}\frac{m^x}{(x-1)!}\right] - m^2$$

$$e^{-m}\left[m^2\sum_{2}^{\infty}\frac{m^{x-2}}{(x-2)!} + m\sum_{1}^{\infty}\frac{m^{(x-1)}}{(x-1)!}\right] - m^2.$$

$$e^{-m}\left[m^2 e^m + m e^m\right] - m^2.$$

$$= m^2 + m - m^2.$$

$\Rightarrow$ Variance $(V) = m$.

Q) A certain screw making machine produces on an average 2 defective out of 100 and pack them in boxes of 500. Find the probability that box contains 1) 3 defective 2) at least one defective.

Ans. probability of defective screw $= \frac{2}{100} = 0.02$ $\quad n = 500$.

$m = np = 500 \times 0.02 = 10.$

Poisson [as $n \to \infty$ (higher) $p - 0.02$ (lower value)].

1) $P(3) = \dfrac{e^{-10}(10)^3}{3!}$

$= 7.5666 \times 10^{-3}$

$= 0.0076.$

2) At least one defective.

$P(X \geq 1) = 1$ (Total sum) $- P(X = 0)$.

$= 1 - \left[\dfrac{e^{-10}\,10^0}{0!}\right]$

$= 0.9991.$