Data Science and Artificial Intelligence

Probability and Statistics

Random Variables

Lecture No.- 03

By- Rahul Sir

Topics to be Covered

Topic Expectation of Random Variables (One Dimensional)

(Univariate)

✓ One Dimensional random — Discrete / continuous

✓ (bivariate Random — Discrete / continuous.

✓ expectation OR Average.

Expected value:

Average OR
center of mass
OR
weighted mean.

Average:

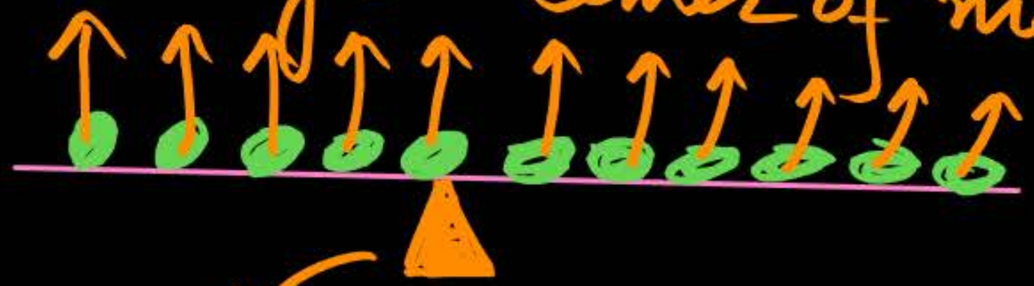
n Different Points (Data)

$x_1 \; x_2 \; x_3 \; x_4 \; x_5 \; x_6$

$1, 2, 3, 4, 5, 6 = \dfrac{\sum x}{n}$

$$\text{Average} = \frac{1+2+3+4+5+6}{6} = 3.5$$

Average – Center of mass.



balance point

balance point
= (Average)

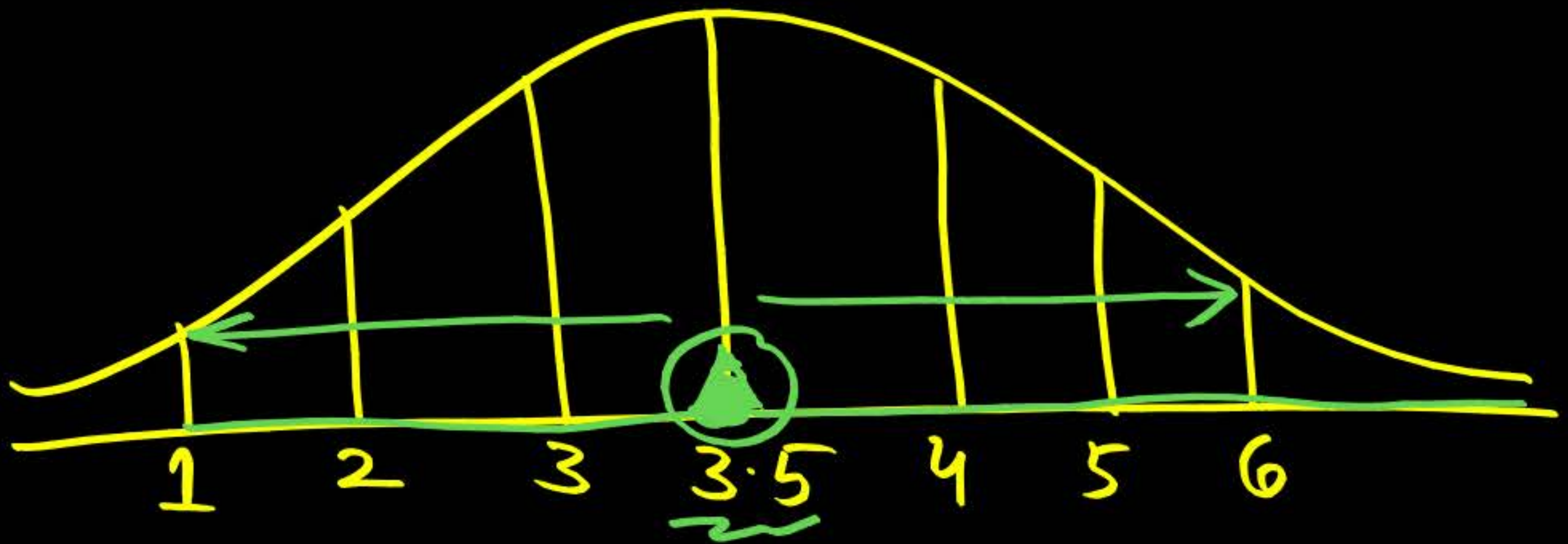
✓ 50% 50%

→ balace point

Die 1, 2, 3, 4, 5, 6 Mean = 3.5

Large No. of trials

10



dataset

balance point

Class - 1 - 10 students

Number (marks)

1 I_1

2 I_2

3 I_3

4 I_4

9 I_5

8 I_6

7 I_7

2 I_8

5 I_9

6.5 I_{10}

Average

$$= \frac{1+2+3+4+9+8+7+2+5+6.5}{10}$$

$\boxed{\text{Single value}}$

Class Performance

Center of mass = Mean OR Average :

Median = mid value

n even → median $\frac{n}{2}$, $\frac{n}{2}+1$

n odd → $\frac{n+1}{2}$ mid value

mid value ⇒ 1, 2, 3, 4, 5, 6 — dataSET EVEN

Median = $\frac{6}{2}$, $\frac{6}{2}+1$

Mode : 1, 1, 1, 1, $\boxed{2, 2, 2, 2, 2, 2}$, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5

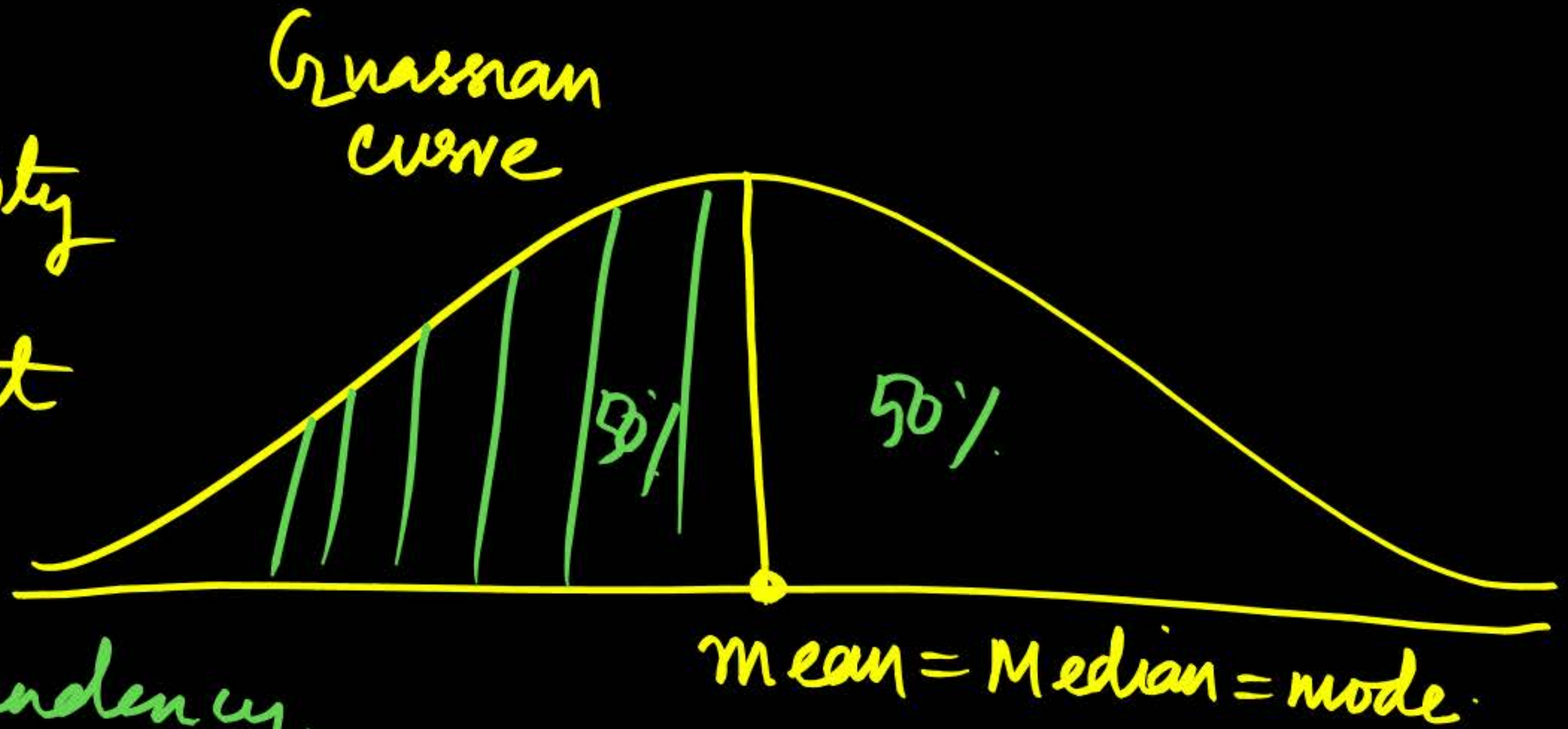
Mode = Highest occur value
= Highest frequent value

2

SAME

Centeral value

{ MEAN → center of gravity
median → mid value
mode → Highest Frequent value



Centeral measure of Tendency.

Highest Frequent form:

dataset-repeat

1, 1, 1, 1, 3, 3, 2, 2, 2, 2, 2, 5, 5, 5, 5, 5, 5, 5, 5

most likely no = 5 = mode.

 Die 1, 2, 3, 4, 5, 6

$P(\text{wining}) = \frac{4}{6} = \frac{2}{3}$] win ↑

$P(\text{Loosing}) = \frac{2}{6} = \frac{1}{3}$] LOOSE ↓

This is a Discrete Random variable.

X	1	2	-3
$P[X=x_i]$	$\frac{3}{6}$	$\frac{1}{6}$	$\frac{2}{6}$

Die Throw one time

for n trials

Play a GAME (CASINO) $X = 1, 2, -3$

1	Re 3	(LOOSE)	-3
2	Re 1	win	$+1$
3	Re 1	win	$+1$
4	Re 1	win	$+1$
5	Re 2	win	$+2$
6	3 ₹	LOOSE	-3 ✓

Should we Play This GAME OR Not?

For n trials

$$\text{Total Payoff} = \underbrace{n \times \frac{3}{6} \times (1)}_{1 \text{ Payoff}} + \underbrace{n \times \left(\frac{1}{6}\right) \times 2}_{2 \text{ Payoff}} + n \times \frac{2}{6} \times (-3)$$

X	1	2	-3
$P(X=x_i)$	$\frac{3}{6}$	$\frac{1}{6}$	$\frac{2}{6}$

$$n \times \frac{1}{2} \times 1 + n\left(\frac{1}{6}\right) \times 2 + n \times \frac{1}{3}(-3) = \boxed{-\frac{n}{6}}$$

quit This GAME

$$\Rightarrow n \times P[X=x_1] \cdot x_1 + n\, P[X=x_2] \cdot x_2 + n\, P[X=x_3] \cdot x_3 = -\frac{n}{6}$$

Total Payoff for n trials

$$= \boxed{P[X=x_1]\, x_1 + P[X=x_2] \cdot x_2 + P[X=x_3] \cdot x_3 = -\frac{1}{6}}$$

$$\boxed{\sum_{i=1}^{3} P[X=x_i]\, x_i} = -\frac{1}{6}$$

Average OR Expected value.

$$\underbrace{\sum_{i=1}^{3} P[X=x_i]\cdot x_i}_{\text{Expected value}} = -\frac{1}{6}$$

quit this GAME because Payoff = Neg.

Average = EXPECTED value $= \boxed{\sum_{i=1}^{n} P[X=x_i]\cdot x_i}$

$= \mu = E[X]$

✓ How to evaluate It for Discrete

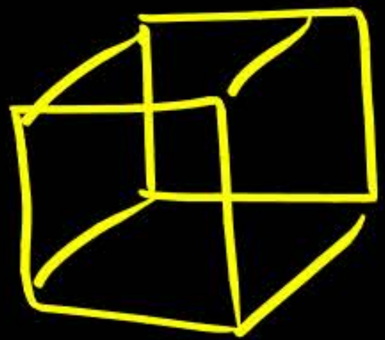
X	x_0	x_1	x_2	x_3	x_n
$P[X=x_i]$	p_0	p_1	p_2	p_3	p_n

$$E[X] = \mu = \frac{x_0p_0 + x_1p_1 + x_2p_2 + \cdots + x_np_n}{p_0 + p_1 + p_2 + \cdots + p_n}$$

$$E[X] = \frac{x_0p_0 + x_1p_1 + x_2p_2 + \cdots + x_np_n}{1}$$

$$E[X] = x_0 p_0 + x_1 p_1 + x_2 p_2 + \dots + x_n p_n$$
$$= \sum_{i=0}^{n} x_i p_i$$

$$\boxed{E[X] = \sum_{i=0}^{n} x_i P[X = x_i)}$$



X = No. of dots

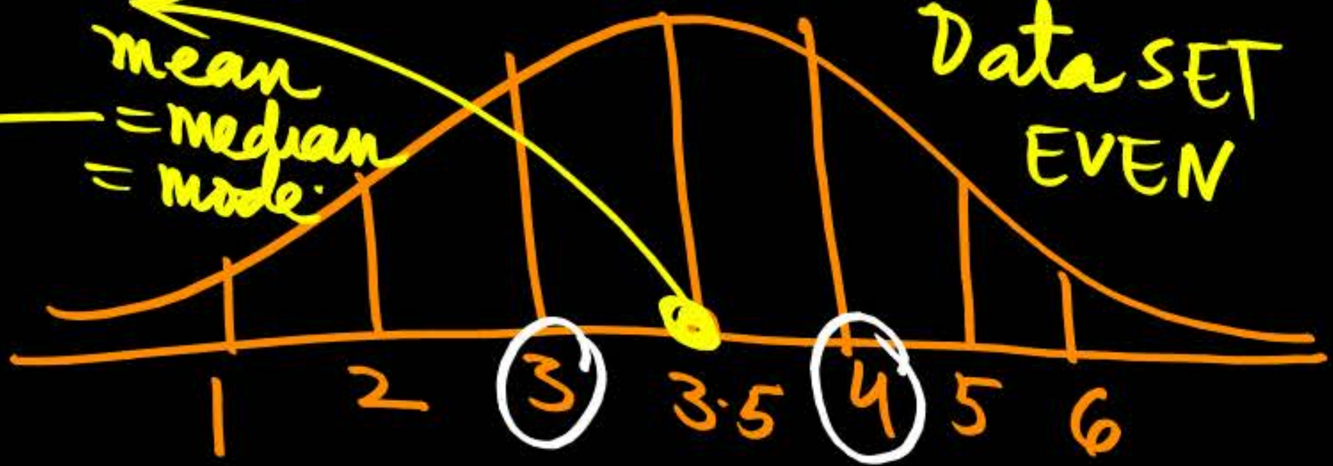
$X = 1, 2, 3, 4, 5, 6$

X	~~1~~	2	3	4	5	6
$P[X=x_i]$	1/6	1/6	1/6	1/6	1/6	1/6

$$E[X] = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + 6\cdot\frac{1}{6}$$
$$= 3.5$$

$$\boxed{E[X] = 3.5}$$

mean = median = mode



Data SET EVEN

Properties of EXPECTATIONS

X → Rohit ○ coin
Y → deepanshu □ Die

Joint Expec

If X and Y be Two Independent random Variable.

$$\boxed{E[X+Y] = E[X] + E[Y]}$$

X : Tossing A coin

(H, T)

X	HEAD(1)	Tail(0)
$P[X=x_i]$	$\frac{1}{2}$	$\frac{1}{2}$

$E[X] = 1 \times \frac{1}{2} + 0 \times \frac{1}{2}$

$= \frac{1}{2} = \underline{0.5}$

Y : Throwing A Die

$Y = 1, 2, 3, 4, 5, 6$

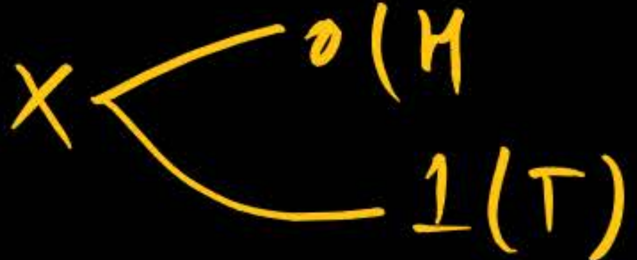
Y	1	2	3	4	5	6
$P[Y=y_j]$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

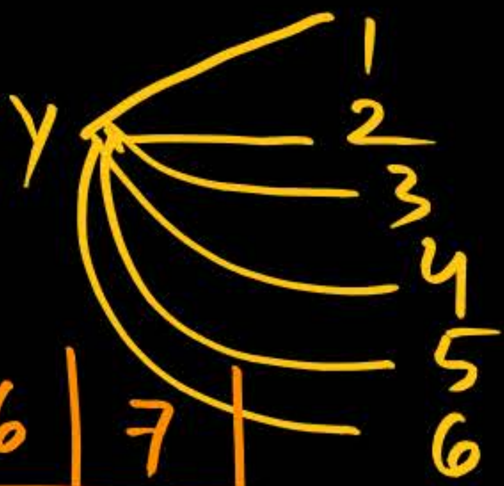
$$\boxed{E[Y] = 3.5}$$

$E[X] + E[Y] = 0.5 + 3.5$

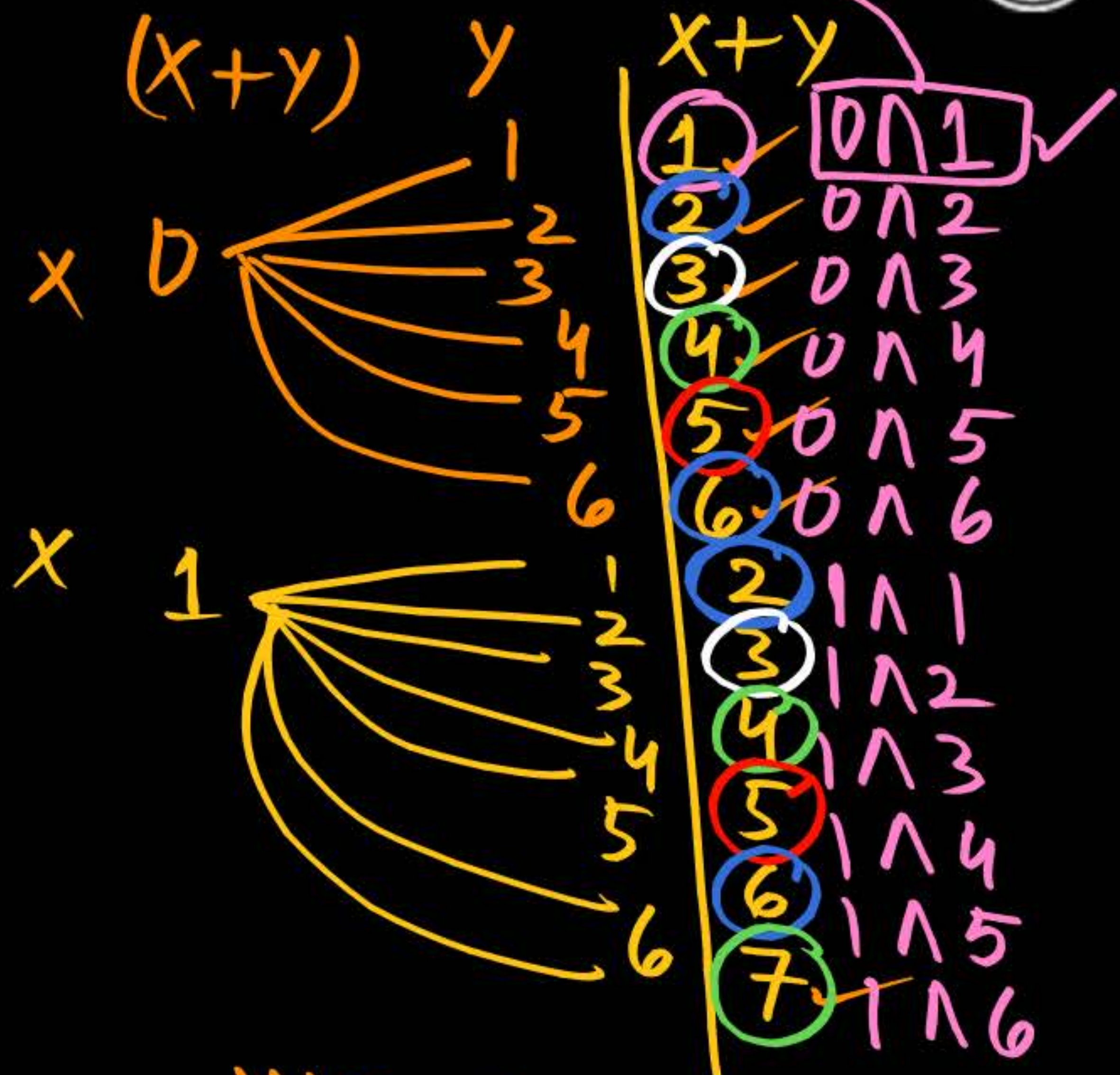
$= \underline{4.0}$

$$P(0)P(1) = \frac{1}{2}\times\frac{1}{6}$$





$X+Y$	1	2	3	4	5	6	7
$P(X+Y = x_i+y_j)$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{1}{12}$



$$P(0\cap 2) + P(1\cap 1)$$

$$= \underset{T}{\underline{P(0)}}\ \underset{2}{\underline{P(2)}} + \underset{H}{\underline{P(1)}}\ \underset{1}{\underline{P(1)}}$$

$$= \frac{1}{2}\times\frac{1}{6} + \frac{1}{2}\times\frac{1}{6} = \frac{2}{12}$$

$X+Y = 1, 2, 3, 4, 5, 6, 7$

$$P(0\cap 1) = P(0)\,P(1) = \frac{1}{2}\times\frac{1}{6} = \frac{1}{12}$$

$\cap$ both Indep.

$$E[X+Y] = 1\times\frac{1}{12}+2\times\frac{2}{12}+3\times\frac{2}{12}+4\times\frac{2}{12}+5\times\frac{2}{12}+6\times\frac{2}{12}+7\times\frac{1}{12}$$

$$\Rightarrow \frac{48}{12} = 4$$

$$\boxed{E[X+Y] = E[X]+E[Y]}$$

superposition of wave

(Linearity)

If $X, Y, Z, T, U, V \ldots$ Are Independent Random variables.

$$E[X+Y+Z+T+U+V+\ldots] = E[X]+E[Y]+E[Z]+\ldots$$

THREE Indep. events

X = Tossing A coin

Y = Throwing A Die

Z = Pick a Deck of draw

$E[cX] = c\,E[X]$

$E[10X] = 10\,E[X]$

Throwing A Die

X	1	2	3	4	5	6
$P(X=x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

$E[X] = 3.5$

$E[11X] = 11\,E[X]$

$= \underline{11 \times 3.5}$ Ans

$= \underline{38.5}$

$E[aX+b] = a\,E[X]+b$

$E[2X+3] = 2\,E[X]+3$

$= 2 \times 3.5 + 3$

$= 10$ Ans

$E\left[\frac{1}{X}\right] \neq \frac{1}{E[X]}$

$E\left[\frac{1}{X^2}\right] \neq \frac{1}{E[X^2]}$

If X and Y Are Dependent Then:

$E[XY] = E[X]\,E\left[\frac{Y}{X}\right]$

If X, Y, Z Are Dependent variable.

$$\boxed{E[XYZ] = E[X]\,E\left[\frac{Y}{X}\right]\,E\left[\frac{Z}{Y \cap X}\right]}$$

Variance : (deviation)

$$\text{variance} = E[X^2] - [E[X]]^2$$

X	x_0	x_1	x_2		x_n
$P(X=x_i)$	p_0	p_1	p_2		p_n

$$E[X^2] = x_0^2 p_0 + x_1^2 p_1 + x_2^2 p_2 + \cdots + x_n^2 p_n$$

$$E[X^2] = \sum_{i=0}^{n} x_i^2 p_i$$

$$\boxed{\text{variance} = \text{var}(X) = \sigma_X^2 = E[X^2] - [E[X]]^2}$$

1, 2, 3, 4, 5, 6

$$\frac{\sum x}{n} = 3.5$$

$(1-3.5)^2$

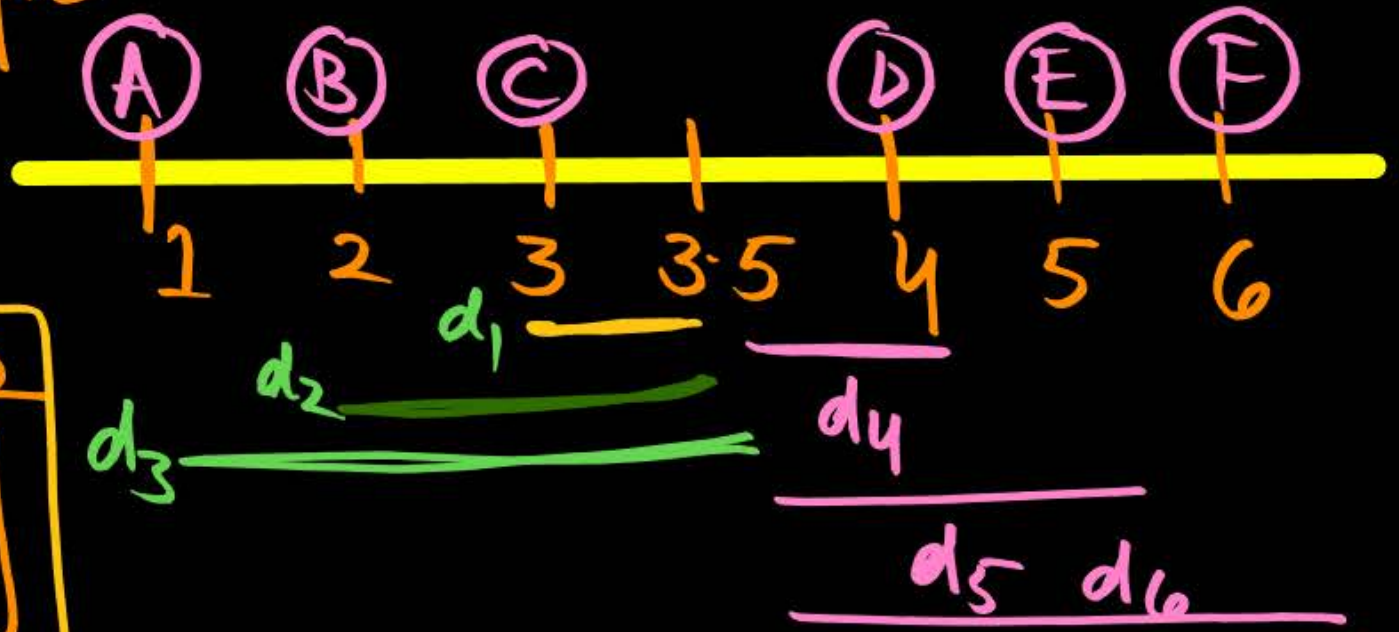
$(2-3.5)^2$

$(3-3.5)^2$

$(4-3.5)^2$

$(5-3.5)^2$

$(6-3.5)^2$



σ_X^2 = variance = $E[X^2] - [E[X]]^2$

variance can't be Negative

standard deviation = $\sqrt{\text{variance}}$

$\text{var}(X) > 0$

$\sigma_X = \sqrt{E[X^2] - [E[X]]^2}$

Continuous Random Variable :

In continuous Random var.

$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx \qquad -\infty \le X \le \infty$$

X is a Cont. random var.

$$E[X] = \int_a^b x f(x)\,dx$$

$$E[X^n] = \int_a^b x^n f(x)\,dx$$

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

$$\mathrm{var}(X) = \int_a^b x^2 f(x)\,dx - \left[\int_a^b x f(x)\,dx\right]^2$$

$$E[X^2] = \int_a^b x^2 f(x)\,dx$$

$$E[X^3] = \int_a^b x^3 f(x)\,dx$$

standard deviation $= \sqrt{\text{variance}}$

THANK - YOU