

# PROBABILITY

## what we will study

- Basic definitions of prob. (axioms)
- sample space, events
- Baye's theorem
- Random variables
- prob. distributions

## what we'll not study

combinatorics type of prob. is in  
discrete maths  
permutation & comb.

## PROBABILITY ?

what does it mean to say -

Probability of appearing head on toss of fair coin is  $\frac{1}{2}$ .

does this mean, if you toss 2 times, you'll get 1 time head. No  
4 times 2 times head. NO  
100 times 50 times head. NO

However, if you toss the same coin 1000 times, the outcomes will be close to half heads and half tails.

English statistician Karl Pearson who tossed a coin 24,000 times with a result of 12,012 heads, one of the authors tossed a coin 2,000 times. The results were 996 heads.

→ Probability tells how likely an event is to occur. enough no.  $\approx$  equally of times likely

$$\lim_{n \rightarrow \infty} \frac{n(H)}{n} = \frac{1}{2}$$

↑ no. of times head  
↓ total tosses

50% chance  $\Rightarrow$  head

50% chance  $\Rightarrow$  tails

ques. Toss a coin 100 times, is it possible to get 100 heads :  $H, H, H, \dots, H$   
100 times

↳ yes it is possible but the prob. is very low,  $(\frac{1}{2})^{100}$

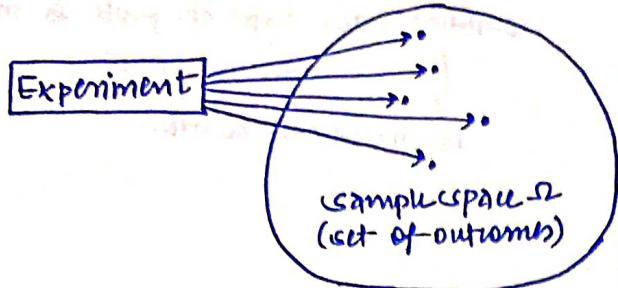
ques. Is it possible :

100 times  $\rightarrow$  0 times head  
tossed  $\rightarrow$  60 times tail

↳ yes it is also possible

## Sample Space

- Every probabilistic model involves an underlying process, called the experiment, that will produce exactly one out of several possible outcomes.
- The set of all possible outcomes is called the sample space of the experiment.

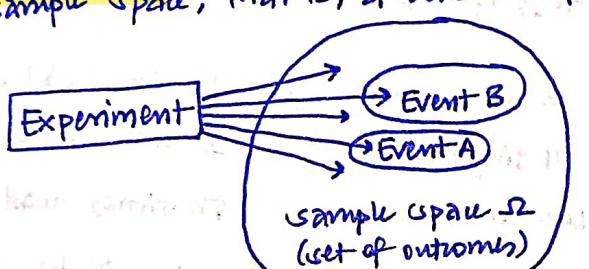


(ex)

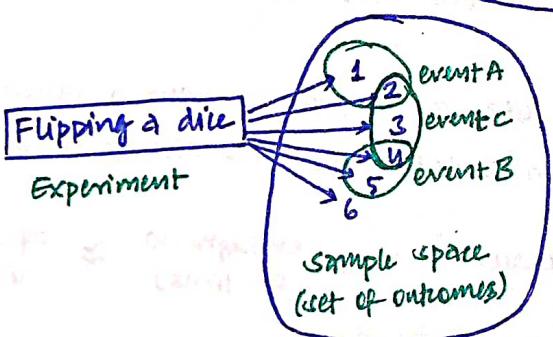
Tossing a coin →  $\{H, T\}$   
Experiment                                  all possible outcomes.

## Event

A subset of the sample space, that is, a collection of possible outcomes, is called an event.



(ex)

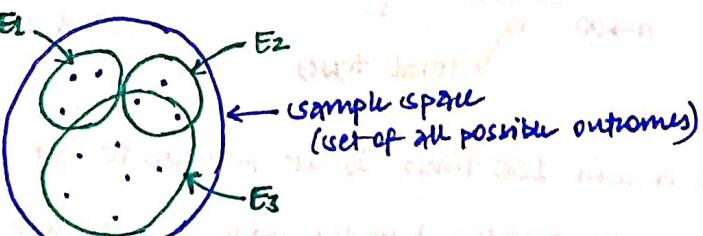


$$\rightarrow E_1 = \{1, 2\}; P(E_1) = 2/6$$

$$E_2 = \{3, 4, 5\}; P(E_2) = 3/6$$

$$E_3 = \{5\}; P(E_3) = 1/6$$

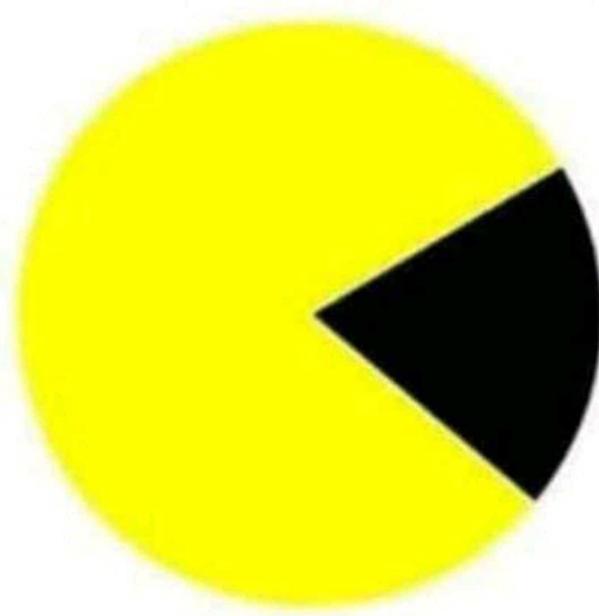
$$\Omega = \{1, 2, 3, 4, 5, 6\}$$



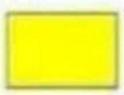
∴ total no. of Events possible

$$= \text{no. of subsets of } \Omega = 2^{\Omega}$$

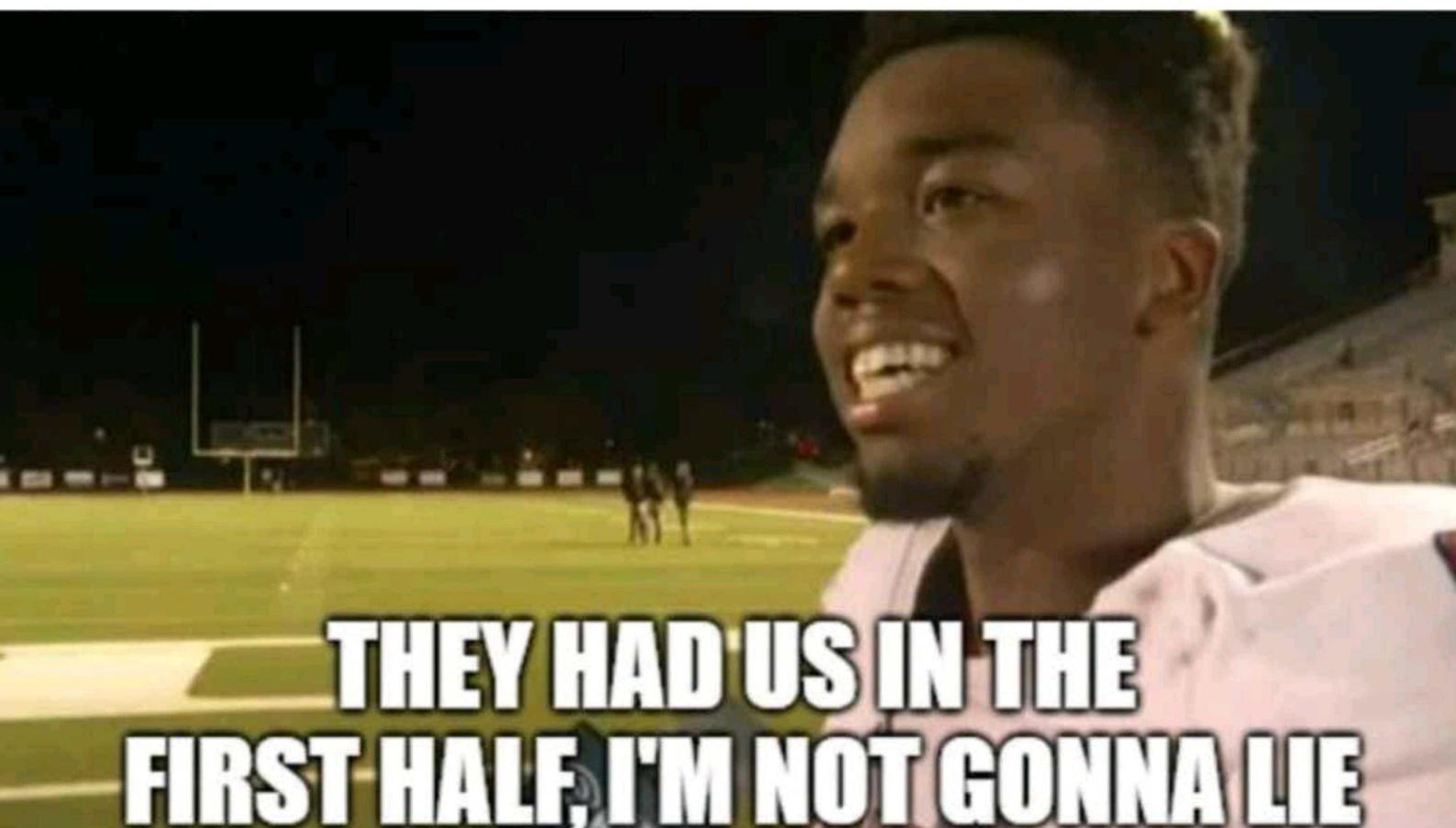
# Probability of me finding a gf



*None*



*Also none, but in yellow to make Pac-Man*



**THEY HAD US IN THE  
FIRST HALF, I'M NOT GONNA LIE**

## (ex) Experiment: Flipping a coin

sample space,  $\Omega = \{H, T\}$

All possible events =  $2^n$

$$\left\{ \{H\}, \{T\}, \{H, T\}, \{\emptyset\} \right\}$$

$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$   
 $P(H) = \frac{1}{2} \quad P(T) = \frac{1}{2} \quad P(\{H, T\}) = \frac{1}{2} \quad P(\emptyset) = 0$

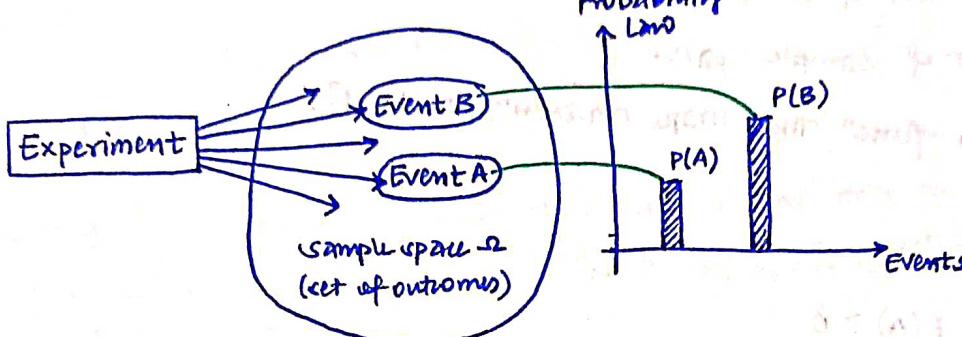
## Probability

Func<sup>n</sup> that maps an event to  $[0, 1]$ .  $\rightarrow$  i.e.  $P : 2^n \rightarrow [0, 1]$

$\rightarrow$  prob. is defined over event.

i.e.  $P(\{\cdot\})$

event



- The main ingredients of a Probabilistic model.

## Properties of Sample Space

o must have "list" (set) of possible outcomes.

o list must be :

- Mutually exclusive

- Collectively Exhaustive

o sample space for tossing a coin :  $\{H, T\}$

sample space for tossing 2 coins :  $\{HH, HT, TH, TT\}$   
 or  
 tossing coin 2 times

sample space for rolling a dice :  $\Omega = \{1, 2, 3, 4, 5, 6\}$   
 (or s)

o if the outcome of an experiment consists in the determination of the sex of a newborn child, then  $S = \{g, b\}$  where the outcome g means the child is a girl. b " boy.

o if the outcome of an experiment is the order of finishing in a race among the 7 horses having post positions 1, 2, 3, 4, 5, 6 and 7, then

$S = \{\text{all } 7 \text{! permutations of } (1, 2, 3, 4, 5, 6, 7)\}$

The outcome  $(2, 3, 1, 6, 5, 4, 7)$  means, for instance, that the number 2 horse comes in first, then the no 3 horse, then the no 1 horse, and so on.

→ Sample space depends not on what we are interested in.

② sample space of tossing coin 2 times : {HH, HT, TH, TT}

but if we are only interested in what comes last ; then  $S = \{H, T\}$

③ if we are interested in order of all horses :  $7!$  permutations

knowing who has come first  $\Rightarrow S = \{1, 2, 3, 4, 5, 6, 7\}$

knowing 1<sup>st</sup> two positions  $\Rightarrow S = \{12, 13, 14, 15, 16, 17, 23, 24, 25, 26, 27, \dots\}$   
↓  
7×6 possible outcomes.

### NOTE →

→ Sample space : set of all possible outcomes.

→ Event : subset of sample space.

→ Probability : A func' that maps an event to [0, 1].

## Axioms of Probability

1. Nonnegativity :  $P(A) \geq 0$

2. Normalization :  $P(\Omega) = 1$  {Probability of all the outcomes combined is 1}

3. Additivity : If  $A \cap B = \emptyset$ , then  $P(A \cup B) = P(A) + P(B)$

→  $S = \{E_1, E_2, E_3\}$  ;  $P(\Omega) = P(E_1) + P(E_2) + P(E_3) = 1$   
(or  $\Omega$ )  
mutually exclusive events

→  $P(\{H\} \text{ or } \{T\}) = P(\{H\}) + P(\{T\})$        $P(H \cup T) = P(H) + P(T)$   
↓  
U ← union      H & T are disjoint

→  $S = \{S_1, S_2, S_3, \dots, S_n\}$         
 $P(\{S_1, S_2, \dots, S_n\}) = P(S_1) + P(S_2) + \dots + P(S_n)$

4. Countable Additivity Axiom : (extended 3<sup>rd</sup> Axiom) if no. of events are infinite

If  $A_1, A_2, \dots$  are disjoint events, then

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

→ Let, all outcomes be equally likely, then

$$P(A) = \frac{\text{no. of elements of } A}{\text{total no. of sample space points}}$$

$$\hookrightarrow S = \{s_1, s_2, s_3, \dots, s_n\}$$

$$P(s_1) = P(s_2) = \dots = P(s_n) \quad (\text{equally likely outcomes})$$

$$\hookrightarrow P(\{s_1\} \cup \{s_2\} \cup \{s_3\} \cup \dots \cup \{s_n\}) = 1 \quad (\text{using 2nd Axiom})$$

$$\Rightarrow P(s_1) + P(s_2) + \dots + P(s_n) = 1$$

$$\Rightarrow P(s_i) = \frac{1}{n}$$

Now, if  $A = \{s_1, s_2, \dots, s_k\}$   
an event

$$\Rightarrow P(A) = \underbrace{P(s_1) + P(s_2) + \dots + P(s_k)}_{\frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n}}$$

$$P(A) = \frac{k}{n} \rightarrow \frac{\text{no. of elements in } A}{\text{total elements}}$$

→ If everything is not equally likely then this doesn't hold true.

Ex:  $\{s_1, s_2, s_3, s_4, s_5, s_6\}$  Now,  $P(s_1, s_2, s_3) \neq \frac{3}{6} = \frac{1}{2}$

$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow$ $\frac{1}{6} \quad \frac{1}{6} \quad \frac{2}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6}$	$P(s_1, s_2, s_3) = \frac{1}{6} + \frac{1}{6} + \frac{2}{6} = \frac{4}{6} = \frac{2}{3}$
--	--

Ques. If 2 dice are rolled, what is the probability that the sum of the upturned faces will equal 7?

$$\hookrightarrow \text{sum} = 7$$

Possible Outcomes: (1,1) (1,2) (1,3) (1,4) (1,5) (1,6)  
(2,1) (2,2) ...  
36 outcomes

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

$$\Rightarrow P(A) = \frac{6}{36} = \frac{1}{6}$$

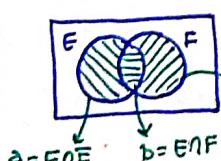
$P(A) = \frac{n(A)}{n}$  ← When every outcome is equally likely

### The Inclusion-Exclusion Principle

For any 2 events, E and F:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Proof



$$\text{Now, } P(E \cup F) = P(a \cup b \cup c) = LHS = P(a) + P(b) + P(c)$$

$$P(E) = P(a \cup b) \quad P(F) = P(b \cup c) \quad P(E \cap F) = P(b)$$

$$\text{RHS: } P(a \cup b) + P(b \cup c) - P(b)$$

$$= P(a) + P(b) + P(c) - P(b) = RHS \Leftrightarrow$$

Hence proved.

→ For 3 events E, F & G, that formula mathematically expands to,

$$\begin{aligned} P(E \cup F \cup G) &= P(E) + P(F) + P(G) \\ &\quad - P(E \cap F) - P(E \cap G) - P(F \cap G) \\ &\quad + P(E \cap F \cap G) \end{aligned}$$

## De Morgan's Law for Probability

For any 2 events E and F :

$$P(\overline{E \cap F}) = P(\overline{E} \cup \overline{F})$$

OR

$$P(\overline{E \cup F}) = P(\overline{E} \cap \overline{F})$$

$$\Rightarrow P(E \cap F) = 1 - P(\overline{E \cap F}) \quad (P(A) + P(\overline{A}) = 1)$$
$$= 1 - P(\overline{E} \cup \overline{F}) \quad (\text{DeMorgan's law})$$

$$\Rightarrow P(E \cup F) = 1 - P(\overline{E \cup F}) \quad (P(A) + P(\overline{A}) = 1)$$
$$= 1 - P(\overline{E} \cap \overline{F}) \quad (\text{DeMorgan's law})$$

GATE 1997 Ques. The prob. that it will rain today is 0.5, and the prob. that it will rain tomorrow is 0.6. The prob. that it will rain either today or tomorrow is 0.7. What is the prob. that it will rain today and tomorrow?

$$\hookrightarrow P(\text{today} \cup \text{tomorrow}) = P(\text{today}) + P(\text{tomorrow}) - P(\text{today} \cap \text{tomorrow})$$

$$0.7 = 0.5 + 0.6 - P(\text{today} \cap \text{tomorrow})$$

$$P(\text{today} \cap \text{tomorrow}) = 0.6 + 0.5 - 0.7 = 0.4$$

GATE 08 Ques. A sample space has 2 events, A and B such that probabilities

$$P(A \cap B) = \frac{1}{2}, \quad P(A) = \frac{1}{3}, \quad P(\overline{B}) = \frac{1}{3}. \quad P(A \cup B) = ?$$

$$\hookrightarrow P(A) = 1 - P(\overline{A}) = 1 - \frac{1}{3} = \frac{2}{3} \quad P(B) = 1 - P(\overline{B}) = 1 - \frac{1}{3} = \frac{2}{3}$$

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

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

# Card, Coin and dice in probability



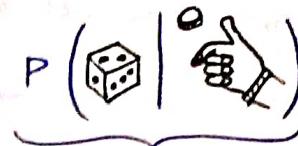
A photograph of a man with dark hair and a beard, looking directly at the camera with a surprised or shocked expression. He is wearing a light-colored, textured shirt. The background is a plain, light-colored wall.

Permanent hoon sir

# Conditional Probability

Probability of an event based on some partial information.

"change in Belief"



(ex) INDIA vs Australia

(assume fairplay conditions & equally good teams)

if the tossing of coin has happened, then what is the prob. of occurring 2 on the dice

Before start of play: What is the chance of India winning? = 0.5

India scores 395 batting first: what is the chance of India winning?  $> 0.5$

- What exactly happened here?

A: event that India will win

$$P(A) = 0.5$$

B: India scores 395 runs.

$P(A)$  changes once we know that event B has occurred

$$P(A|B) > 0.5$$

$$P(A|B) \neq P(A)$$

**NOTE** →

$$\begin{array}{c} P(A|B) > P(A) \\ \leq \\ = \end{array} \quad \left. \begin{array}{l} \text{there is} \\ \text{no relation b/w} \\ P(A) \& P(A|B) \end{array} \right.$$

anything can happen

## Example of conditioning

Dice tossing : We consider the following situation

- We throw 2 dice.
- We look for  $P(\text{sum of 2 faces is } 9)$

$$(2,6), (6,2), (4,5), (5,4)$$

• Without prior info :  $P(\text{sum of 2 faces is } 9) = \frac{4}{36} = \frac{1}{9}$

- with additional info :

If first face is = 4. Then

- Only 6 possible results : (4,1), (4,2), (4,3), (4,5), (4,6)
- Among them, only (4,5) give sum = 9
- Prob. of having sum = 9 becomes  $\frac{1}{6}$

Infected by covid : Suppose 1% of population has covid

$$\frac{1}{100}$$

$\therefore P(\text{someone is having covid}) = \frac{1}{100} \leftarrow \text{without prior info}$

$P(\text{person is having covid} | \text{Person is having symptoms of covid}) \approx 1 \leftarrow \text{with prior info}$

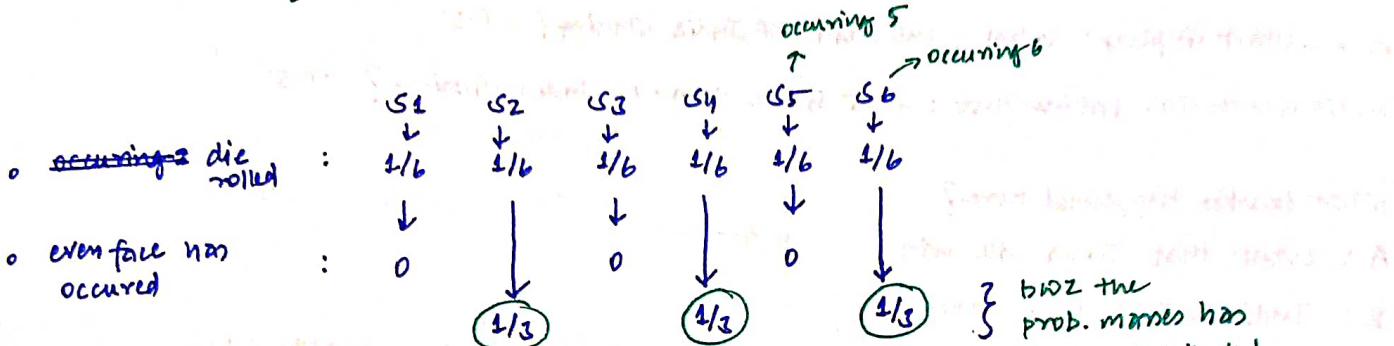
Ques. A die has been rolled where additional info is given that the experiment has produced an even face.

What is the prob. of that 2 occurred?

$$P(\text{occurring } 2) \rightarrow \frac{1}{6}$$

$$P(\text{occurring } 2 | \text{even face}) = \frac{1}{3}$$

$$\Omega = \{2, 4, 6\}$$



old sample space  
 $\Omega = \{1, 2, 3, 4, 5, 6\}$

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

New sample space  
 $\Omega' = \{2, 4, 6\}$

$$P(2) = \frac{1}{3}$$

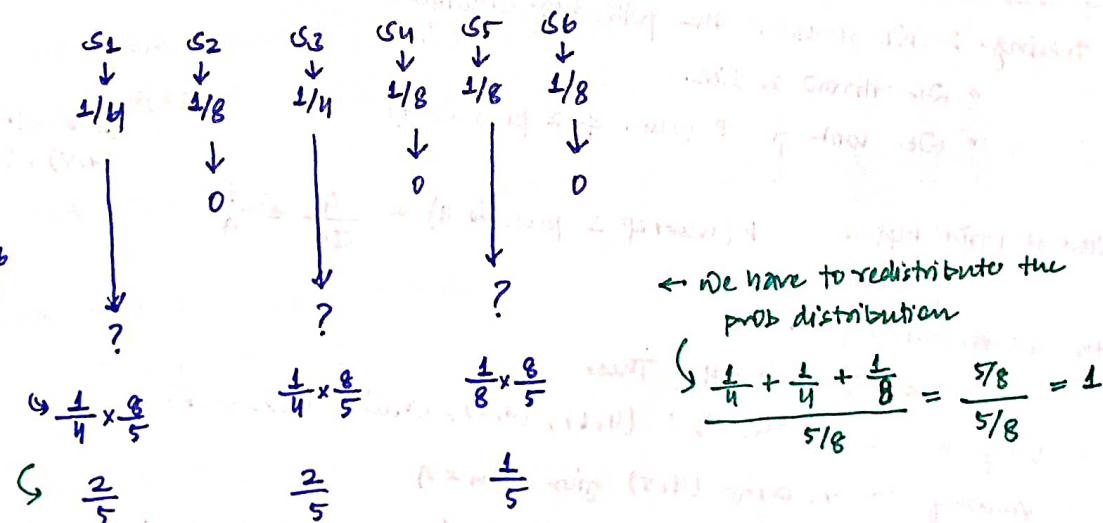
$$\frac{\frac{1}{6} + \frac{1}{6} + \frac{1}{6}}{3/2} = \frac{1/2}{3/2} \Rightarrow$$

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

$$P(2 | \text{even face}) = \frac{1}{3}$$

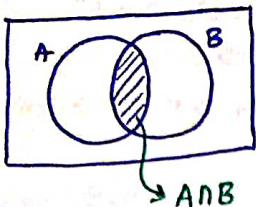
Ques. If we have the sample space as:

Initial prob. distribution  
Now, additional info is that,  $S_2, S_4 \& S_6$  has not occurred



$$\text{Now, therefore } P(S_1 | \{S_1, S_2, S_3\}) = \frac{2}{5}$$

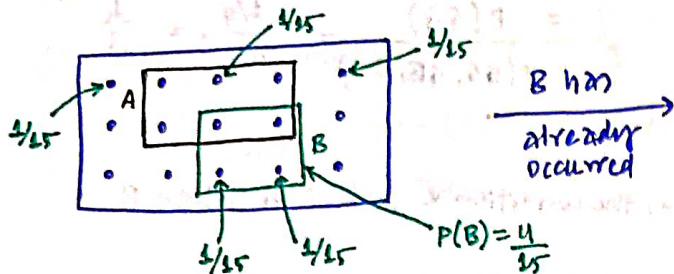
$$P(S_2 | \{S_1, S_2, S_3\}) = 0$$



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

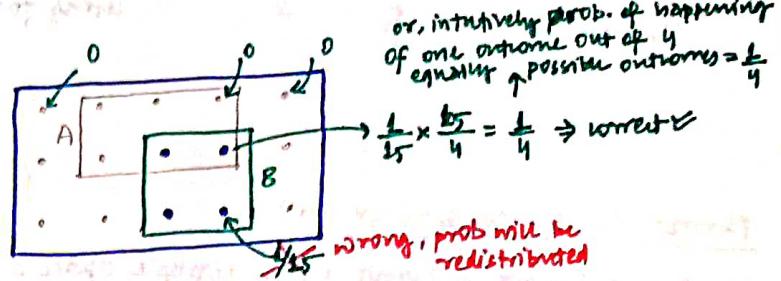
prob. that A has occurred  
given that B has already occurred.

lets understand this intuitively,



every outcome in the sample space is equally possible

we always work here



outcomes when event B does not happen is not possible

we dont need  
this new sample space  
This is just for  
intuition.

$$\frac{\frac{1}{15} + \frac{1}{15} + \frac{1}{15} + \frac{1}{15}}{4/15} = \frac{1/15}{4/15} = \frac{1}{60}$$

using intuition:

$$P(A|B) = \text{prob}(A) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

- Using intuition is only for understanding we wont use it to solve questions as it is a bit confusing sometimes.  
Always use formula.

$\hookrightarrow$  we the prob mass distribution is not equally among outcomes.

ques. A die has been rolled where additional info is given that the experiment has produced no. less than 4.

What is the prob that 2 occurred?

$$\hookrightarrow \{1, 2, 3\} \leftarrow \text{given} \quad \text{Intuitively, } P(2 | \{1, 2, 3\}) = \frac{1}{3}$$

$$\text{Formula : } P(2 | \{1, 2, 3\}) = \frac{P(\{2\} \cap \{1, 2, 3\})}{P(\{1, 2, 3\})} = \frac{P(\{2\})}{P(\{1, 2, 3\})} = \frac{1/6}{1/2} = \frac{1}{3}$$

ques. consider a family with 2 children. We are interested in the children's genders. Our sample space is  $S = \{(G,G), (G,B), (B,G), (B,B)\}$ . Assume all 4 possible outcomes are equally likely.

- What is the prob. that both children are girls given that the 1st is a girl?

$$\text{Intuitively: } \{(G,G), (G,B)\} \rightarrow \frac{1}{2}$$

$$\text{Formulae : } P(\{GG\} | \{GG, GB\}) = \frac{P(GG)}{P(GG, GB)} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{4}} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

(b) We ask the father: "Do you have atleast one daughter?". He responds "Yes".

Given this extra info, what is the prob that both children are girls?

In other words, what is the prob that both children are girls given that we know atleast one of them is girl?

↪  $\{GG, GB, BG\}$

Intuitively  $\rightarrow \frac{1}{3}$

Using formula:  $P(GG | \{GG, GB, BG\})$

$$= \frac{P(GG)}{P(GG, GB, BG)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Formal def of conditional prob:

If A and B are 2 events in a sample space S, then the conditional prob of A given B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ when } P(B) > 0$$

$$P(A|B) = \frac{\text{no. of elements of } A \cap B}{\text{no. of elements of } B} \cdot \left( \begin{array}{l} \text{when all outcomes are} \\ \text{equally likely} \end{array} \right)$$

$$\frac{\frac{n(A \cap B)}{n(S)}}{\frac{n(B)}{n(S)}} = \frac{n(A \cap B)}{n(B)}$$

→ conditional prob can also be viewed as a prob law on a new universe B, bcoz all of the conditional prob is concentrated on B.

Ques. A fair dice is tossed twice

E = {The sum of the face values of these 2 tosses is 5.},

(a) Calculate  $P(E)$ .

(b) given that the event E happens, what is the prob that the face value of the first toss is less than that of the 2nd toss?

↪ The sample space S,  $S = \{(i,j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}$

in total =  $6 \times 6 = 36$  elements.

(a) The event E is  $E = \{(1,4), (2,3), (3,2), (4,1)\}$

$$\text{and } P(E) = \frac{4}{36} = \frac{1}{9}$$

(b) The conditional prob.

intuitively,  $S' = \{(1,4), (2,3), (3,2), (4,1)\}$

$$\therefore \text{Prob} = \frac{2}{4} = \frac{1}{2}$$

using formulae,

$$P(\{ \text{ } \} | \{(1,4), (2,3), (3,2), (4,1)\})$$

$$= P\left(\frac{\{(1,4), (2,3)\}}{4/36}\right)$$

$$= \frac{2/36}{4/36} = \frac{1}{2} \quad \checkmark$$

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

Ques. Suppose we flip 3 fair coins. Let A stand for the event that 1<sup>st</sup> 2 coins are both heads. Let B stands for the event that 3<sup>rd</sup> coin is different from the 2<sup>nd</sup> coin.

(a) Find  $P(A)$ ,  $P(B)$ .

$$P(A) = P(HHT) + P(HHH) = \frac{2}{8} = \frac{1}{4}$$

$$P(B) = P(HHT) + P(HTH) + P(THT) + P(THH) = \frac{4}{8} = \frac{1}{2}$$

(b) Find the conditional probability  $P(A|B)$ .

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \text{ and } B)}{P(B)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

$$P(A \text{ and } B) = P(HHT) = \frac{1}{8}$$

### Multiplication Rule (or Product Rule)

(directly follows from conditional probability definition)

$$P(A \cap B) = P(A) P(B|A) = P(B) P(A|B)$$

$$P(A \cap B) = \frac{P(AB)}{P(B)}$$

$$P(B|A) = \frac{P(AB)}{P(A)}$$

o  $P(A \cap B) = P(A, B) = P(AB)$  (notation)

$$P(A, B) = P(A) \cdot P(B|A)$$

$$P(B, A) = P(B) \cdot P(A|B)$$

$$P(A \cap B \cap C) = P(A, B, C) = P(C) \cdot P(B|C) \cdot P(A|B, C)$$

Ques. Which of the following is correct.  $P(A, B, C) = ?$

(a)  $P(C) \cdot P(B|C) \cdot P(A|B, C) \checkmark \rightarrow P(A, B, C)$

(b)  $P(A) \cdot P(B|A) \cdot P(C|A, B) \checkmark \rightarrow P(C, B, A)$

(c)  $P(B) \cdot P(A|B) \cdot P(C|A, B) \checkmark \rightarrow P(C, A, B)$

(d)  $P(B) \cdot P(C|B) \cdot P(A|B, C) \checkmark \rightarrow P(A, C, B)$

All are correct, as  $P(A, B, C) = P(C, B, A) = P(C, A, B) = P(A, C, B)$

$$\Rightarrow P(A, B, C, D) = P(D) \cdot P(C|D) \cdot P(B|C, D) \cdot P(A|B, C, D)$$

## TREE DIAGRAM

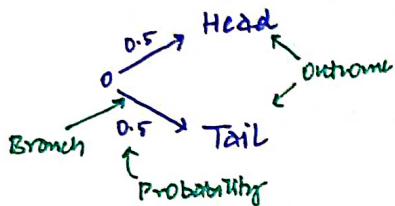
(for conditional probability)

→ useful in sequential models

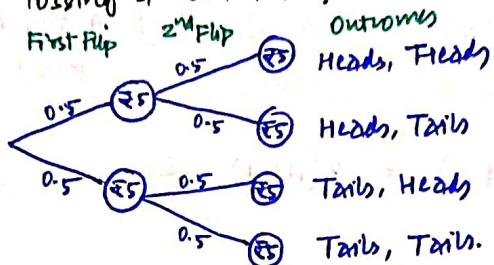
Many experiments have an inherently sequential character, such as for ex :-

- (i) Tossing a coin 3 times.
- (ii) Observing the value of a stock on 5 successive days.
- (iii) Receiving 8 successive digits at a communication receiver

→ Tree diagram for coin toss



→ Tossing a coin twice!

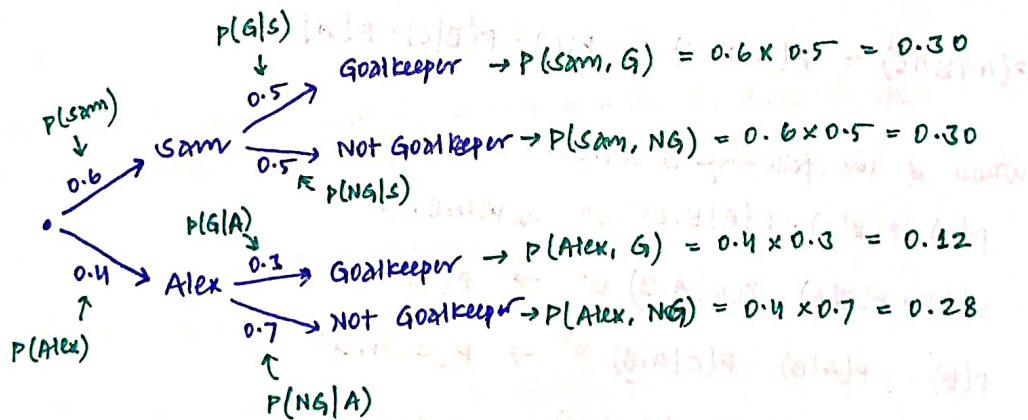


Ques. You are off to soccer, & love being the Goalkeeper, but that depends who is the coach today:

- with coach Sam the prob of being Goalkeeper is 0.5
- — " — Alex      ||      0.3

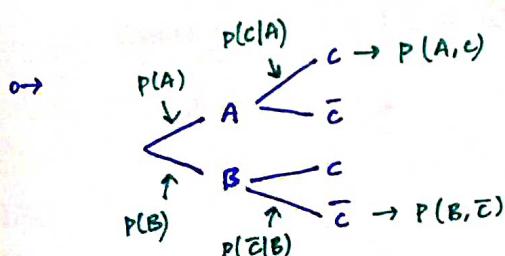
probability of Sam being coach = 0.6.

So, what's the prob you'll be a goalkeeper today?



$$\therefore P(G) = P(\text{Sam}, G) + P(\text{Alex}, G)$$

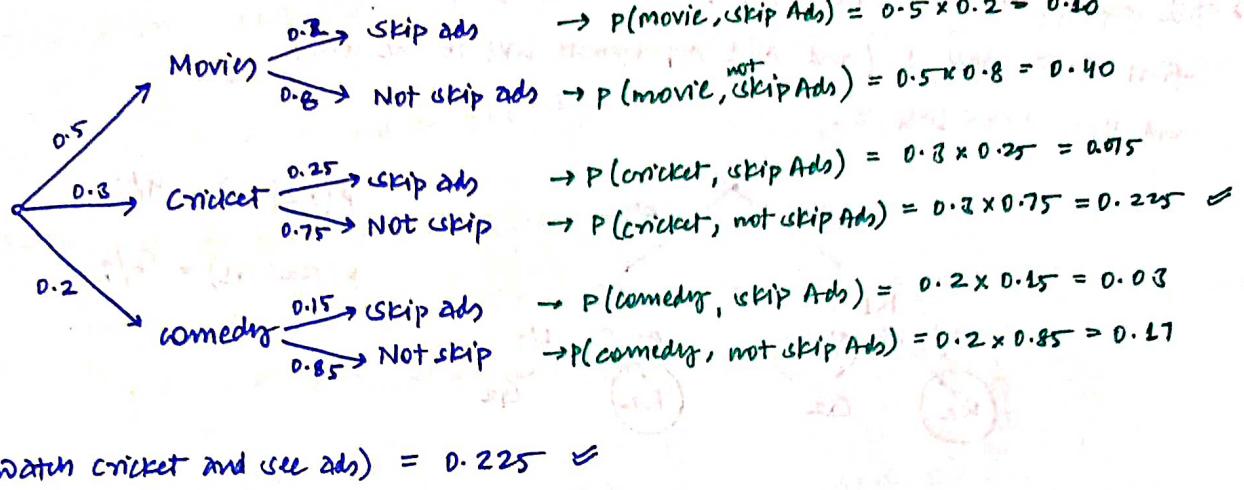
$$= 0.30 + 0.12 = \boxed{0.42}$$



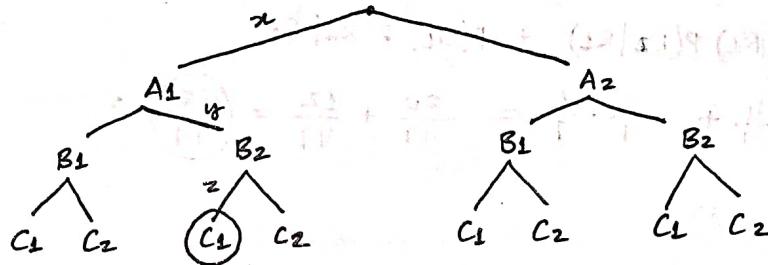
$$P(A, C) = P(A) \cdot P(C|A) \quad \leftarrow \text{so this tree diagram makes sense mathematically too}$$

Ques. On a Sunday evening, 50% watch a movie, 30% watch a cricket match and 20% watch comedy show. The % of people skipping ads are 20%, 25% & 15% respectively.

$$P(\text{watch cricket and see ads}) = ?$$



Conceptual ques.



(i) The prob ' $x$ ' represents

- (a)  $P(A_1)$  ✓
- (b)  $P(A_1 | B_2)$
- (c)  $P(B_2 | A_1)$
- (d)  $P(C_1 | B_2 \cap A_1)$

(ii) The prob ' $z$ ' represents

- (a)  $P(C_1)$
- (b)  $P(B_2 | C_1)$
- (c)  $P(C_1 | B_2)$
- (d)  $P(C_1 | B_2 \cap A_1)$  ✓

(ii) The prob ' $y$ ' represents

- (a)  $P(B_2)$
- (b)  $P(A_2 | B_2)$
- (c)  $P(B_2 | A_1)$  ✓
- (d)  $P(C_1 | B_2 \cap A_1)$

(iv) The circled node represents the event

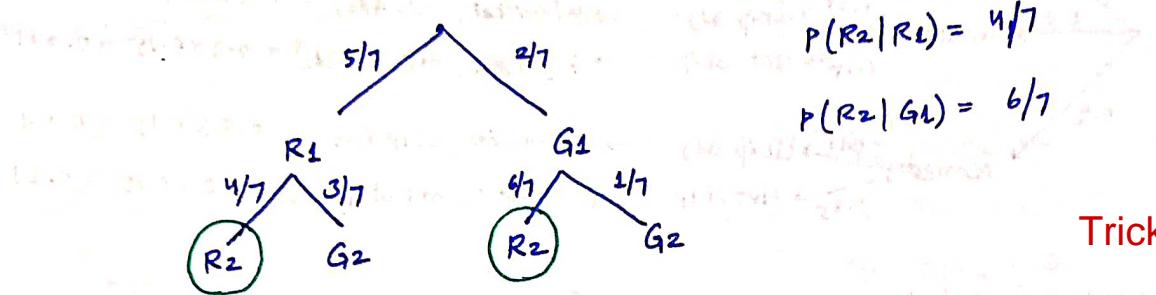
- (a)  $C_1$
- (b)  $B_2 \cap C_1$
- (c)  $A_1 \cap B_2 \cap C_1$  ✓
- (d)  $C_1 | B_2 \cap A_1$ .

Ques. An urn contains 5 red balls and 2 green balls. A ball is drawn. If it's green, a red ball is added to the urn & if its red a green ball is added to the urn. (The original ball is not returned to the urn). Then a second ball is drawn. What is the prob that the 2nd ball is red?

↪ The sequence of the actions are:

first draw ball 1 (and add appropriate ball to the urn)

and then draw ball 2.



$$\begin{aligned} \therefore P(R_2) &= P(R_2, R_1) + P(R_2, G_1) \\ &= P(R_1) P(R_2|R_1) + P(G_1) P(R_2|G_1) \\ &= \frac{5}{7} \cdot \frac{4}{7} + \frac{2}{7} \cdot \frac{6}{7} = \frac{20}{49} + \frac{12}{49} = \boxed{\frac{32}{49}} \end{aligned}$$

Answers and markings done below (A) to (D)

(A) 4/9

(B) 5/9

(C) 6/9

(D) 12/9

## Conditional Probabilities satisfy the 3 Axioms :

(i) Non Negative :  $P(A) \geq 0 \rightarrow P(A|B) \geq 0 \rightarrow \frac{P(A \cap B)}{P(B)} \geq 0$ .  
 conditional prob is always non-ve.

(ii) Normalization : earlier we used to say that prob of sample space = 1.  
 i.e.,  $P(\Omega) = 1$ .  $\rightarrow$  prob of sample space given some event  $B$ .

Now, it will be like  $P(\Omega|B) = 1 \rightarrow P(\Omega \cap B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$ .

(iii) Additivity : if  $A_1$  and  $A_2$  are disjoint then,

earlier :  $P(A_1 \cup A_2) = P(A_1) + P(A_2)$

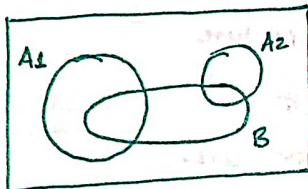
now :  $P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$

proof :  $P(A_1 \cup A_2 | B) = \frac{P((A_1 \cup A_2) \cap B)}{P(B)}$

$$= \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)}$$

$$= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)}$$

$$= P(A_1 | B) + P(A_2 | B)$$



distributive

disjoint sets

## Axioms of Probability

1. Non negativity :

$$P(A) \geq 0$$

2. Normalization :

$$P(\Omega) = 1$$

3. Additivity :

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

1. Non negativity :

$$P(A|B) \geq 0$$

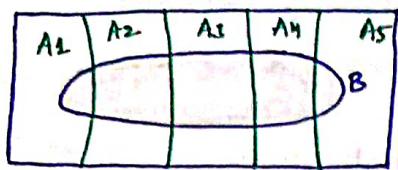
2. Normalization :

$$P(\Omega|B) = 1$$

3. Additivity :

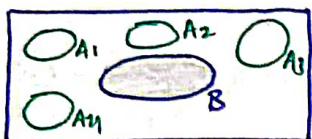
$$P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$$

## Divide & conquer



$$P(B) = P((B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup (B \cap A_4) \cup (B \cap A_5))$$

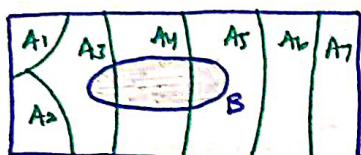
o→



$$\Rightarrow P(B) = P((B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup (B \cap A_4)) \quad \text{X}$$

We cannot write like  
this in this case

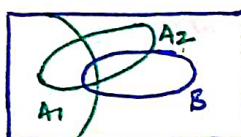
[NOTE] → To write this, we should have all the  $A_i$ 's as mutually exhaustive.



$$\Rightarrow A_1 \cup A_2 \cup \dots \cup A_n = \Omega \quad \& \text{ all are disjoint}$$

$$\text{then, } B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n) \quad \text{✓}$$

o→



$(B \cap A_1) \& (B \cap A_2)$  are not disjoint here

bcz  $A_1 \& A_2$  are not disjoint.

so can't use divide & conquer here

[NOTE] → if  $A_1, A_2, A_3$  are making partition then,

for any event  $B$ :

$$P(B) = P((B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3))$$

$$= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)$$

- 1.  $A_1, A_2, A_3$  are mutually exhaustive
- 2.  $A_1, A_2, A_3$  are mutually disjoint.

## Marginalization (also called Total Probability)

If  $A_1, A_2$  and  $A_3$  partition the sample space then for any event  $B$  :

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)$$

Very easy to prove using venn diagram:



$$\Rightarrow P(B) = P((B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3))$$

## Rules in Probability

- If outcomes are equally likely,

$$P(A) = \frac{\text{no. of elements in } A}{\text{total no. of sample points}}$$

- Inclusion Exclusion principle :→

for any 2 events  $A, B$  :

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

- De-Morgan's Law :→

for any 2 events  $A, B$  :

$$P(\overline{A \cap B}) = P(\overline{A} \cup \overline{B})$$

$$\therefore P(A) + P(\overline{A}) = 1$$

- Total probability :→

If  $A_1, A_2, A_3$  partition the sample space then for any event  $B$ :

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)$$

- Bayes' Theorem :→

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

$$= \frac{P(A \cap B)}{P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)}$$

## Rules in conditional Probability

$$P(A|B) = \frac{\text{no. of elements of } A \cap B}{\text{no. of elements of } B}$$

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

$$P(\overline{(A \cup B)}|E) = P(\overline{A} \cap \overline{B}|E)$$

$$P(A|B) + P(\overline{A}|B) = 1$$

$$P(B|A) = P(B \cap A_1|A) + P(B \cap A_2|A) + P(B \cap A_3|A)$$

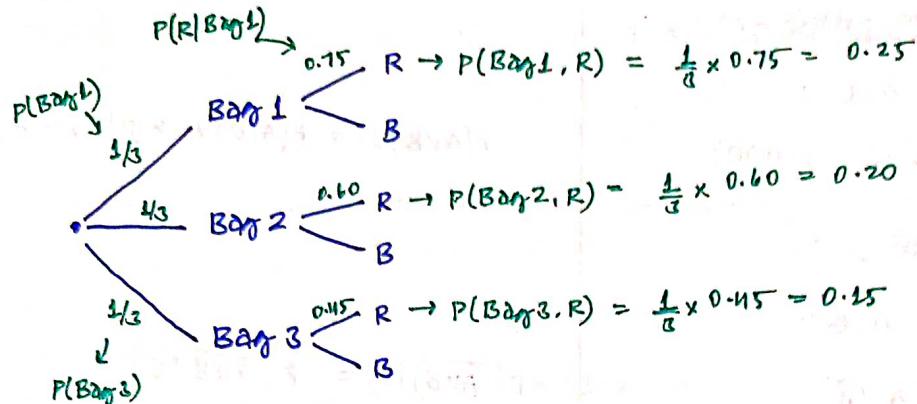
$$P(A|B,E) = \frac{P(A \cap B|E)}{P(B|E)}$$

$$= \frac{P(A \cap B|E)}{P(B \cap A_1|E) + P(B \cap A_2|E) + P(B \cap A_3|E)}$$

Ques. I have 3 bags that each contain 100 marbles :

- Bag 1 has 75 red and 25 blue marbles.
- Bag 2 60 R 40 B
- Bag 3 45 R 55 B

I chose one of the bag at random & then pick a marble from the chosen bag, also at random. What is the prob. that the chosen marble is red?



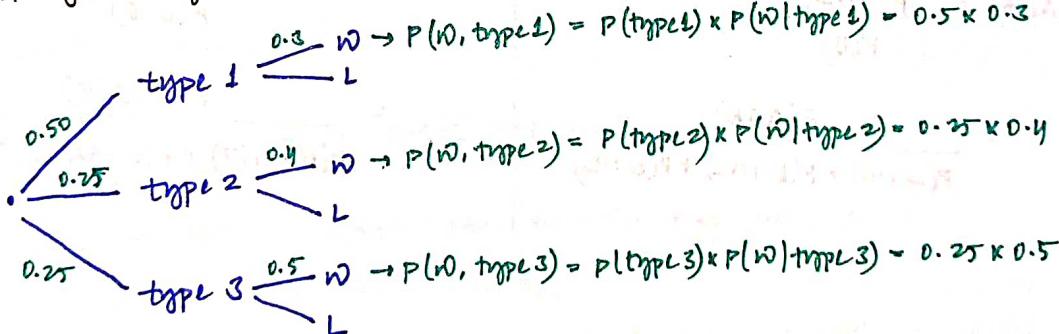
$$P(R) = 0.25 + 0.20 + 0.15 = 0.60$$

$$P(R|Bag_1) = 0.25$$

Ques. In bags there we'll learn to answer  $P(Bag_1|R)$  → The picked marble is red, what is the prob it is from Bag 1

Ques. You enter a chess tournament where your probability of winning a game is 0.3 against half the players (call them type 1), 0.4 against a quarter of the players (call them type 2), and 0.5 against the remaining quarter of the players (call them type 3).

You play a game against a random chosen opponent. What is the prob of winning?

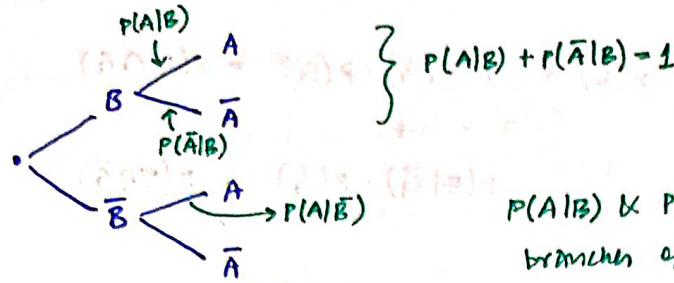


$$\begin{aligned} P(W) &= 0.5 \times 0.3 + 0.25 \times 0.4 + 0.25 \times 0.5 \\ &= 0.375 \end{aligned}$$

## True / False

(i)  $P(A|B) + P(\bar{A}|B) = 1$  True

(ii)  $P(A|B) + P(A|\bar{B}) = 1$  False



$P(A|B)$  &  $P(A|\bar{B})$  are of complete different branches of different parents.  
They are not related, hence  $P(A|B) + P(A|\bar{B}) = 1$  is false

(iii)  $P(A|\bar{B}) + P(\bar{A}+B) = 1$  False

for the similar reason ↳

Ques. Which of the following is true?

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

(b)  $P(\bar{B}|A) = P(B) - P(B|A)$

↪  $P(B) + P(\bar{B}) = 1$  (in prob)

↪  $P(\bar{B}|A) + P(B|A) = 1$

Now,  $P(B|A) + P(\bar{B}|A) = 1$  (in conditional prob)

⇒  $P(B) = 1$  ↳ This need not be True

⇒ TRUE

(c)  $P(E,F) + P(\bar{E}\bar{U}\bar{F}) = 1$

(d)  $P(E\bar{U}F) = 1 - P(\bar{E}\bar{U}\bar{F}) P(F)$

↪  $P(E\bar{U}F) + P(\bar{E}\bar{U}\bar{F})$

↪  $1 - P(\bar{E}\bar{U}\bar{F})$

Now,  $P(E\bar{U}F) = 1 - P(\bar{E}\bar{U}\bar{F})$  demorgan's law

⇒  $1 - P(\bar{E}\bar{U}\bar{F}) \cdot P(F)$

⇒ TRUE

↪  $P(\bar{E}\bar{U}\bar{F}) = P(\bar{E}\bar{U}\bar{F})$  factorization

=  $P(F) \cdot P(\bar{E}|F)$

(e)  $P(E, F|E) = P(E, F|F)$

(f)  $P(E|F) P(F) + P(\bar{E}|F) P(F) = P(F)$

↓  
 $P(E\bar{U}F|E)$

↓  
 $P(E, F) + P(\bar{E}, F)$

=  $\frac{P(E\bar{U}F|E)}{P(E)}$

TRUE

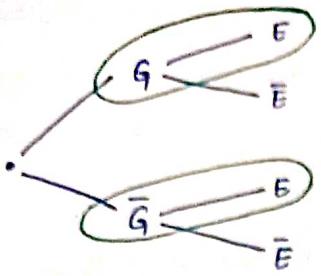
↪  $P(E\bar{U}F|F)$

=  $\frac{P(E\bar{U}F|F)}{P(F)}$

=  $\frac{P(E\bar{U}F)}{P}$

LHS ≠ RHS

FALSE



→ Total probability for  $E$  to happen,

$$P(E) = P(G \cap E) + P(\bar{G} \cap E)$$

Another intuition to this is,

$$P(E) = P(E|G) \cdot P(G) = P(E \cap G)$$

+

$$P(E|\bar{G}) \cdot P(\bar{G}) = P(E \cap \bar{G})$$

Ques (T/F)

$$P(E|F) = P(E|F, G) P(G|F) + P(E|F, \bar{G}) P(\bar{G}|F)$$

↳ in RHS, there is an extra variable ' $G$ ', so we can use concept of total prob for it.

$$\Rightarrow \text{in normal prob: } P(E) = P(E \cap G) + P(E \cap \bar{G})$$

$$\text{in conditional prob: } P(E|F) = P(E \cap G|F) + P(E \cap \bar{G}|F)$$

$$\hookrightarrow \text{in normal prob: } P(E) = P(E \cap G) + P(E \cap \bar{G}) \quad \xrightarrow{\text{factorization}}$$

$$= P(G) \cdot P(E|G) + P(\bar{G}) \cdot P(E|\bar{G})$$

$$\text{if in conditional prob: } P(E|F) = P(G|F) \cdot P(E|G, F) + P(\bar{G}|F) \cdot P(E|\bar{G}, F)$$

Hence proved

(TRUE)

Method 2:

$$P(E|F) = P(E|F, G) P(G|F) + P(E|F, \bar{G}) P(\bar{G}|F)$$

$$\text{LHS: } \frac{P(E \cap F)}{P(F)}$$

$$\text{RHS: } \frac{P(E, F, G)}{P(F, G)} \cdot \frac{P(F, G)}{P(F)} + \frac{P(E, F, \bar{G})}{P(F, \bar{G})} \cdot \frac{P(F, \bar{G})}{P(F)}$$

$$= \frac{P(E, F, G) + P(E, F, \bar{G})}{P(F)} = \frac{P(E, F)}{P(F)}$$

LHS = RHS

→ (TRUE)

Method 3:

$$P(E|F) = \underbrace{P(E|F, G)}_{\frac{P(E, G|F)}{P(G|F)}} P(G|F) + \underbrace{P(E|F, \bar{G})}_{\frac{P(E, \bar{G}|F)}{P(\bar{G}|F)}} P(\bar{G}|F)$$

$$P(E|G) = \frac{P(E, G)}{P(G)}$$

$$= \frac{P(E, G|F)}{P(G|F)} + \frac{P(E, \bar{G}|F)}{P(\bar{G}|F)}$$

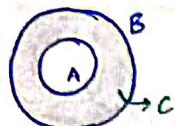
$$= P(E, G|F) + P(E, \bar{G}|F)$$

$$= P(E|F)$$

Ques. True/False

If A and B are events and  $A \subseteq B$  then  $P(A) \leq P(B)$ .

↪



$$B = A \cup C$$

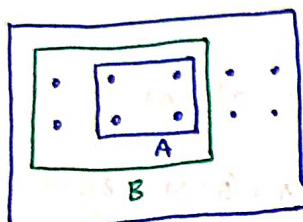
$$P(B) = P(A \cup C) = P(A) + P(C)$$

$$P(B) = P(A) + P(C) \quad (\because P(C) > 0)$$

We know that  $P(C) > 0$ , therefore

$$\Rightarrow P(B) \geq P(A)$$

(e)



$$P(B) = \frac{10}{10}$$

$$P(A) = \frac{4}{10}$$

$$A \subseteq B$$

$$P(A) \leq P(B)$$

Ques. True/False

$$P(E, F, G) \leq \min\{P(E), P(F), P(G)\}$$

$$\hookrightarrow E \cap F \cap G \subseteq E \Rightarrow P(E \cap F \cap G) \leq P(E)$$

$$E \cap F \cap G \subseteq F$$

$$P(E, F, G) \leq P(F)$$

$$E \cap F \cap G \subseteq G$$

$$P(E, F, G) \leq P(G)$$

Ques. True/False

If the occurrence of event F makes event E more likely, then the occurrence of E necessarily makes F also more likely.

$$\hookrightarrow i.e. P(E|F) \geq P(E) \Rightarrow P(F|E) \geq P(F)$$

Tricky

$$\text{given: } P(E|F) \geq P(E)$$

$$\Rightarrow \frac{P(E,F)}{P(F)} \geq P(E)$$

$$\Rightarrow \frac{P(E,F)}{P(E)} \geq P(F)$$

$$\Rightarrow P(F|E) \geq P(F) \quad \text{Hence Proved}$$

TRUE

Ques. Which of the following is equal to  $P(B \cap C|A)$ ?

(a)  $P(B|C \cap A)$

(c)  $P(B|C \cap A) P(C|A)$

(b)  $\frac{P(B|C)}{P(A)}$

(d)  $P(B|C) P(C|A)$

$\hookrightarrow P(B \cap C|A) = P_A(B \cap C)$

 $= P_A(B|C) P_A(C)$ 
 $= P(B|C \cap A) P(C|A)$

Thus, option (c) is correct

Simpler Method  $\rightarrow$  We can extend the normal formulae

$$P(B, C) = P(C) \cdot P(B|C)$$

NOW, putting conditional prob. formulae

$$P(B, C|A) = P(C|A) \cdot P(B|C, A)$$

$\therefore$  option (c) is correct

Ques. Which of the following is true?

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

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

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

$\Rightarrow$  option (a) is correct



$$A = (B \cap A) \cup (\bar{B} \cap A)$$

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

Ques. If we wish to express  $P(B|A)$  in terms of only  $B$  and  $\bar{A}$ , show that

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

$\hookrightarrow P(B|A) = \frac{P(A, B)}{P(A)}$

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

$$= \frac{P(B) - P(B|\bar{A}) P(\bar{A})}{1 - P(\bar{A})} \quad \left\{ P(A, B) + P(\bar{A}, B) = P(B) \right\}$$

Ques. (T/F)  $P(A, B|C, D, E) = \frac{P(A, B, C, E|D)}{P(C, E|D)}$

TRUE

in RHS  
if it is given 'D'  $\rightarrow$  so, lets skip 'D' first  
from LHS & then evaluate.

$\hookrightarrow P(A, B|C, \bar{D}) = \frac{P(A, B, C, E)}{P(C, E)}$  {writing formula without 'D'}

Adding 'given D' to both the sides now  $\Rightarrow$

$$\Rightarrow P(A, B|C, E, D) = \frac{P(A, B, C, E|D)}{P(C, E|D)}$$

$\Rightarrow$  If we can say that:

$$P(A, B|C, D, E) = \frac{P(A, B, E|C, D)}{P(E|C, D)}$$

$$\text{then } P(A, B|E) = \frac{P(A, B, E)}{P(E)}$$

$$P(A, B|C, D, E) = \frac{P(A, B, C, D, E)}{P(C, D, E)}$$

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

$$P(A|B, C) = \frac{P(A, B|C)}{P(B|C)}$$

## Live Session 1 : GATE PYQs Tree Method

ques. A person starts his journey from Bangalore with 0.1 probability he goes to Hyderabad & with 0.9 prob he goes to Chennai.  
If he is in Hyderabad then 0.3 probability he goes to Mumbai and with 0.7 prob he goes to Pune.

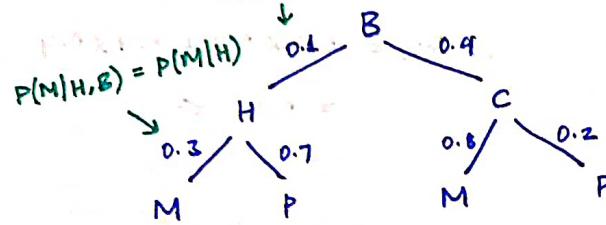
If he is in Chennai then with 0.8 prob he goes to Mumbai and with 0.2 prob he goes to Pune.

Prob that he <sup>will</sup> goes to Pune?

Prob that he will go to Mumbai?

If person is in Mumbai what is the prob that he came from Hyderabad? bcoz Bang  
is just starting point

$$P(H|B) = P(H)$$



$$P(M|H, B) = P(M|H) \quad \text{&} \quad P(H|B) = P(H)$$

bcoz according to the ques, once you reach to Hyd. you don't care from where you have come

$$\therefore P(M|H) = P(M|H) P(H)$$

$$= 0.3 \times 0.1 = 0.03$$

$$(i) P(\text{Pune}) = 0.1 \times 0.7 + 0.9 \times 0.2 = 0.25$$

$$(ii) P(\text{Mumbai}) = 0.1 \times 0.3 + 0.9 \times 0.8 = 0.75$$

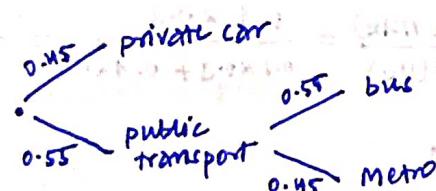
$$(iii) P(H|M) = ?$$

$$\hookrightarrow P(H|M) = \frac{P(H \cap M)}{P(M)} = \frac{0.03}{0.75} = 0.04$$

ISRO 2016

ques. A person on a trip has a choice between a private car and public transport. The prob of using a private car is 0.45. While using public transport, the further choices available are bus & metro. Out of which the probability of commuting by a bus is 0.55. In such a situation, the prob of using a car, bus and metro respectively would be?

- (a) 0.45, 0.30 & 0.25
- (b) 0.45, 0.25 and 0.30
- (c) 0.45, 0.55 & 0
- (d) 0.45, 0.35 and 0.20



$$P(\text{private car}) = 0.45$$

$$P(BM) = 0.55 \times 0.55 = 0.3025$$

$$P(\text{Metro}) = 0.45 \times 0.55 = 0.2475$$

$$P(\text{Metro, public})$$

GATE 08

Ques. Aishwarya studies either CSE or Maths everyday. If she studies CSE on a day, then the prob that she'll study Maths the next day is 0.6.

If she studies Maths on a day, then the prob that she studies CSE the next day is 0.4.

Given that she studies CSE on Monday, what is the prob that she studies CSE on Wednesday?

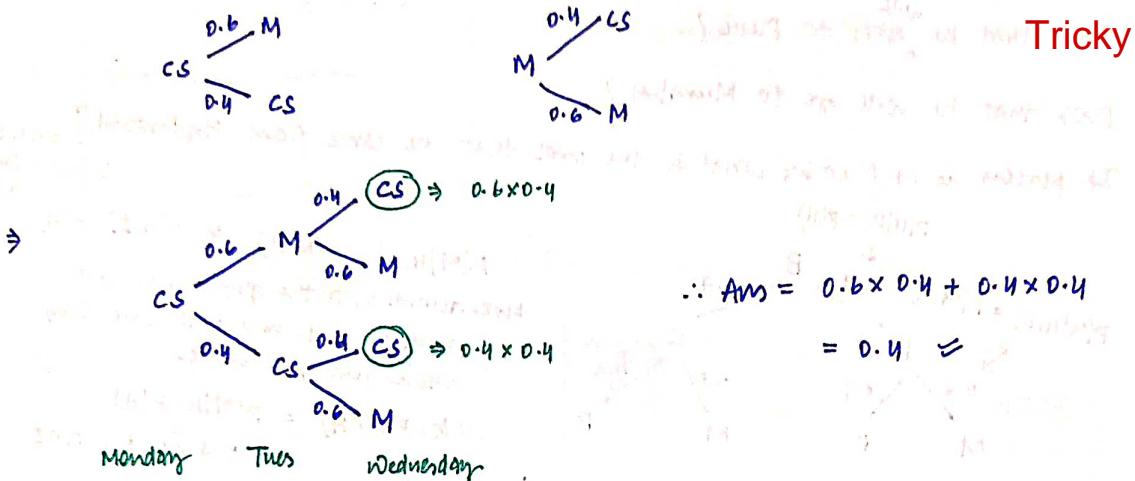
(A) 0.24

(B) 0.36

(C) 0.4

(D) 0.6

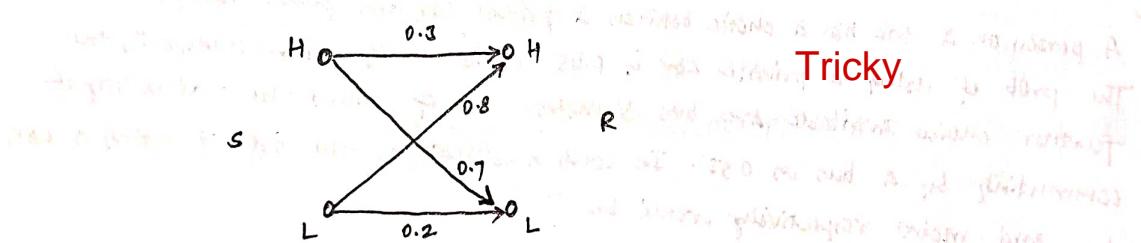
6



GATE 21

Ques. A sender (S) transmits a signal, which can be one of the 2 kinds : H & L with probabilities 0.1 and 0.9 respectively to a Receiver (R).

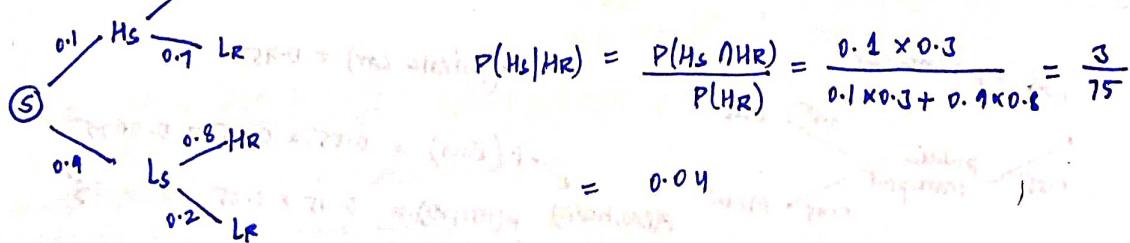
In the graph below, the wedge of the edge  $(u, v)$  is the prob of receiving  $v$  when  $u$  is transmitted, where  $u, v \in \{H, L\}$ . For ex, the prob that the received signal is L, given the transmitted signal was H, is 0.7.



If the received signal is H, the prob that the transmitted signal was H is \_\_\_\_.

$$P(H_R | H_S) \rightarrow 0.3 / H_R$$

We want to find:  $P(H_S | H_R)$



### GATE IT 2006

Ques. In a certain town, the prob that it'll rain in the afternoon is known to be 0.6. Moreover, meteorological data indicates that if the temp at noon is less than or equal to  $25^{\circ}\text{C}$ , the prob that it'll rain in the afternoon is 0.4. The temp at noon is equally likely to be above  $25^{\circ}\text{C}$  or at/below  $25^{\circ}\text{C}$ . What is the prob that it'll rain in the afternoon on a day when the temp at noon is above  $25^{\circ}\text{C}$ ?

$$(A) 0.4$$

$$(B) 0.6$$

$$(C) 0.8$$

$$(D) 0.9$$

$$P(\text{rain}) = 0.6$$

$$P(>25^{\circ}\text{C}) = P(\leq 25^{\circ}\text{C}) = 0.5$$

$$P(\text{rain} | \leq 25^{\circ}\text{C}) = 0.4$$

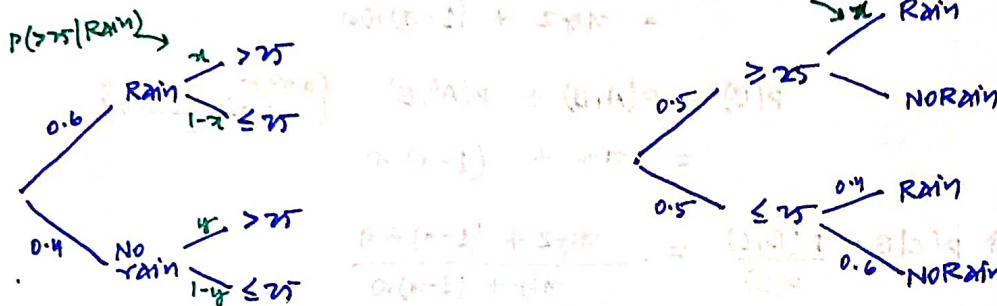
$$P(\text{rain} | > 25^{\circ}\text{C}) = ?$$

Most IMP

Trickier

Change the Perspective

→ We can make 2 tree diagrams from the given info:



$$P(>25^{\circ}\text{C}) = 0.5 = P(\leq 25^{\circ}\text{C})$$

$$P(\text{rain}) = 0.6$$

$$\begin{aligned} \Rightarrow 0.6x + 0.4y &= 0.5 && \text{Using total probability} \\ 0.6(1-x) + 0.4(1-y) &= 0.5 && \end{aligned}$$

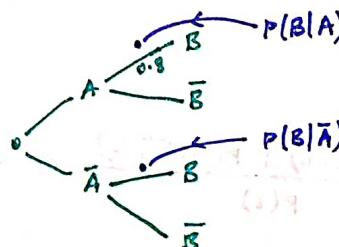
But using these eqns we can't find  $x$  &  $y$  so we gotta use the other tree diagram.

$$\begin{aligned} & 0.5x + 0.5 \times 0.4 = 0.6 \\ & x = 1.2 - 0.4 = 0.8 \end{aligned}$$

$$P(\text{Rain} | > 25^{\circ}\text{C}) = 0.8$$

$$\Rightarrow P(\text{Rain} | > 25^{\circ}\text{C}) = 0.8$$

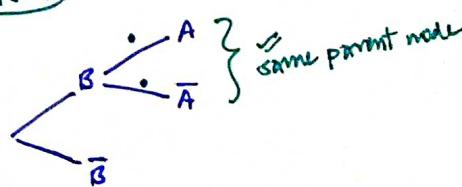
$$\text{then } P(\text{Rain} | \leq 25^{\circ}\text{C}) = 0 \quad \text{you don't know} \rightarrow \text{bcz } P(B|A) + P(B|\bar{A}) \neq 1$$



ques T/F

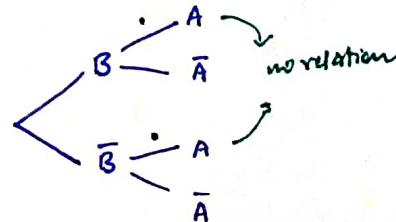
$$(i) P(A|B) + P(\bar{A}|B) = 1$$

TRUE



$$(ii) P(A|B) + P(A|\bar{B}) = 1$$

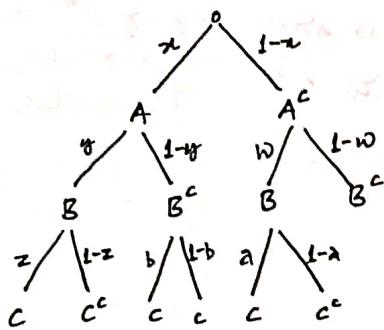
FALSE



## Live Session 2 : GATE PYQs and Bayes's Theorem

Someone asked this ques.

$$P(C|B) = ?$$



$$P(C|B, A) = ?$$

$$P(A, B, C) = xyz$$

$$\text{Now, } P(C|B) = \frac{P(B, C)}{P(B)}$$

$$P(B, C) = P(A, B, C) + P(A^c, B, C)$$

$$= xyz + (1-x)zw$$

{using marginalization}

$$P(B) = P(A, B) + P(A^c, B)$$

$$= xy + (1-x)w$$

$$\Rightarrow P(C|B) = \frac{P(B, C)}{P(B)} = \frac{xyz + (1-x)zw}{xy + (1-x)w}$$

$$P(C|A) = ?$$

$$\boxed{\text{Method 1}} \rightarrow P(C|A) = P(C, B|A) + P(C, B^c|A) \quad \cancel{P(C|A) = P(C|A, B) + P(C|A, B^c)}$$

we need to do marginalization wrt C.

$$= \frac{P(A, B, C)}{P(A)} + \frac{P(A, B^c, C)}{P(A)}$$

$$= \frac{xyz}{x} + \frac{x(1-y)b}{x} = yz + (1-y)b$$

$$\boxed{\text{Method 2}} \rightarrow P(C|A) = \frac{P(C, A)}{P(A)} = \frac{P(C, A, B) + P(C, A, B^c)}{x} = \frac{xyz + x(1-y)b}{x}$$

$$= yz + (1-y)b$$

$$P(A|C) = ?$$

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(A, B, C) + P(A, B^c, C)}{P(C)} = \frac{xyz + x(1-y)b}{P(C)}$$

$$P(C) = \underbrace{P(C, A)}_{=} + \underbrace{P(C, A^c)}_{=} \\ = P(A, B, C) + P(A, B^c, C) + P(A^c, B, C) + P(A^c, B^c, C)$$



## BAYES THEOREM

Posterior probability	Likelihood	Prior probability
$p(A B)$	$p(B A) p(A)$	$p(B)$

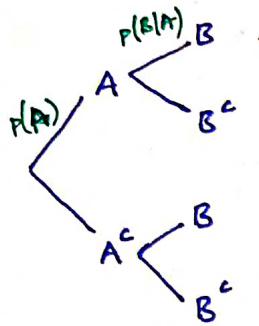
## BAE'S THEOREM

$$P(chill|Netflix) = \frac{P(Netflix|chill)P(chill)}{P(Netflix)}$$

HaaHaHAhAHaaa

## BAYE'S THEOREM

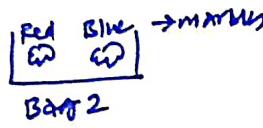
→ It is for the backward reasoning, not for the forward reasoning which have done  
 ex.  $P(A|B)$ ,  $P(A^c|B)$  etc.  $\downarrow$   
 $P(B|A)$ ,  $P(B|A^c)$  etc until now.



$$P(A|B) = \frac{P(A, B)}{P(B)}$$

$$= \frac{P(A, B)}{P(A, B) + P(A^c, B)}$$

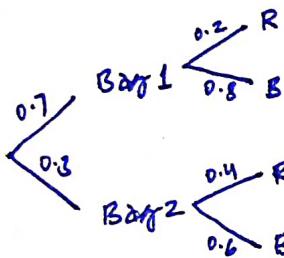
(ex)



$$P(\text{Bag 1}) = 0.7 \quad P(\text{Bag 2}) = 0.3$$

$$P(R|\text{Bag 1}) = 0.2$$

$$P(R|\text{Bag 2}) = 0.4$$



$$P(\text{Bag 1}|R) = \frac{P(R \wedge \text{Bag 1})}{P(R)}$$

$$= \frac{0.2 \times 0.7}{0.7 \times 0.2 + 0.3 \times 0.4} = \frac{7}{13} = 0.54$$

Reverse type of question

$$P(\text{Bag 1}|R) = ?$$

you have chosen the red marble what is the prob it has come from Bag 1

→ Formal def of Baye's Theorem:

Let  $A_1, A_2, \dots, A_n$  be disjoint events that form a partition of the sample space, and assume that  $P(A_i) > 0$ , for all  $i$ . Then, for any event  $B$  such that  $P(B) > 0$ ,

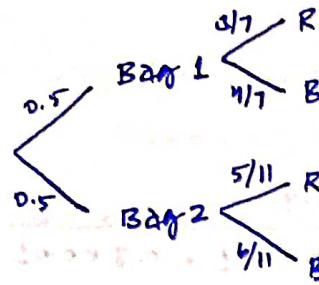
we have  $P(A_i|B) = \frac{P(A_i) P(B|A_i)}{P(B)}$  {or  $\frac{P(A_i, B)}{P(B)}$ }

$$= \frac{P(A_i) P(B|A_i)}{P(A_1) P(B|A_1) + P(A_2) P(B|A_2) + \dots + P(A_n) P(B|A_n)}$$

A<sub>i</sub>'s are making a partition so we have marginalized (Total prob)

## Questions

(i)

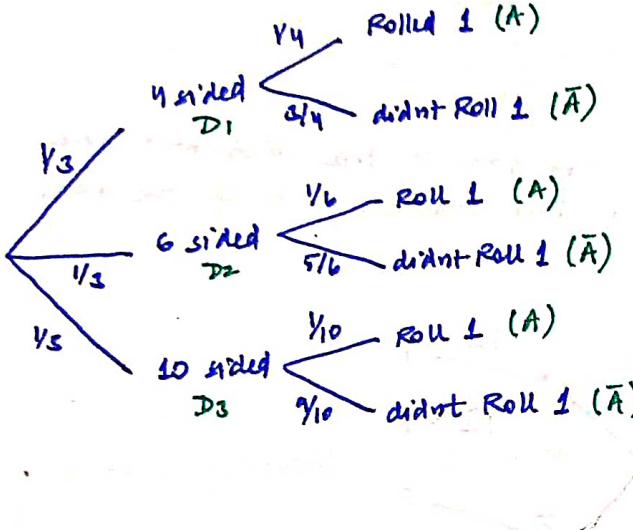


$$P(Bag 2 | R) = ?$$

$$P(Bag 2 | R) = \frac{P(R, Bag 2)}{P(R)}$$

$$= \frac{0.5 \times \frac{5}{11}}{0.5 \times \frac{3}{7} + 0.5 \times \frac{5}{11}} = 0.37$$

(ii)



Let A be event that rolled 1  
B dies be D1, D2 & D3.

$$\therefore P(D_1 | A) = ?$$

$$P(D_1 | A) = \frac{P(A, D_1)}{P(A)}$$

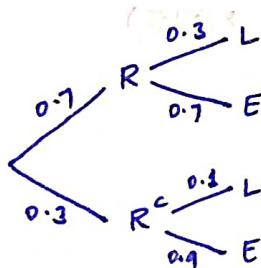
$$= \frac{\frac{1}{3} \times \frac{1}{4}}{\frac{1}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{1}{6} + \frac{1}{5} \times \frac{1}{10}} = \frac{15}{31}$$

(iii)

L = Late, R = Rain, E = Early,  $P(L|R) = 0.3$

$$P(L|R^c) = 0.1$$

$$P(R) = 0.7$$



$$(a) \quad P(E) = P(R, E) + P(R^c, E) \quad \text{Marginalization}$$

$$= P(R) P(E|R) + P(R^c) P(E|R^c)$$

$$= 0.7 \times 0.7 + 0.3 \times 0.1 = 0.76$$

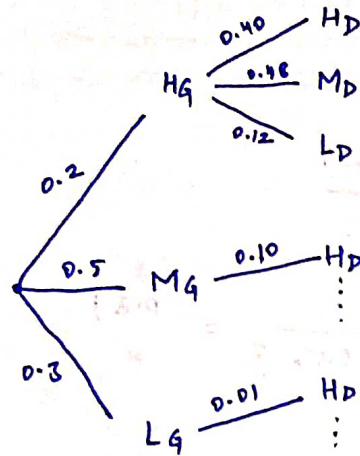
$$(b) \quad P(R|E) = \frac{P(R, E)}{P(E)}$$

$$= \frac{P(R, E)}{P(R, E) + P(R^c, E)} = \frac{0.7 \times 0.7}{0.7 \times 0.7 + 0.3 \times 0.1} = \frac{49}{76} = 0.6447$$

$\boxed{P(E) = \frac{P(R,E) + P(R^c,E)}{P(R) + P(R^c)}}$

$\boxed{P(R|E) = \frac{P(R,E)}{P(E)}}$

(iv)



$$P(H_G | H_D) = ?$$

$$P(H_G | H_D) = \frac{P(H_G, H_D)}{P(H_D)}$$

$$= \frac{0.2 \times 0.4}{0.2 \times 0.1 + 0.5 \times 0.1 + 0.3 \times 0.1}$$

$$= \frac{8}{13.3} \approx 0.60$$

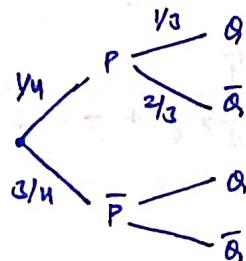
(v)

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

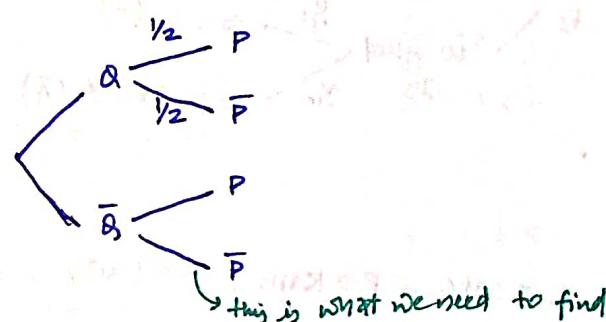
$$P(B|P) = \frac{1}{3}$$

$$P(P|B) = \frac{1}{2}$$

$$P(\bar{P}|\bar{B}) = ??$$



We can have 2 tree diagrams here  
as our understanding of the question



we need to find out  $P(B)$  first;

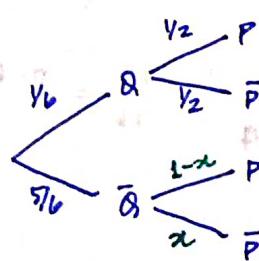
$$\Rightarrow P(P, B) = P(P) \cdot P(B|P) = P(B) \cdot P(P|B)$$

$$\downarrow \quad \downarrow \quad \downarrow$$

$$\frac{1}{4} \quad \frac{1}{3} \quad \frac{1}{2}$$

$$\Rightarrow P(B) = \frac{\frac{1}{4} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{6}$$

Now, the 2nd tree will become



$$P(\bar{P}) = \frac{3}{4}$$

$$P(\bar{P}) = P(Q, \bar{P}) + P(\bar{Q}, \bar{P})$$

$$\frac{3}{4} = \frac{1}{6} \times \frac{1}{2} + \frac{5}{6} \times x \Rightarrow x = \frac{4}{5} = 0.8$$

$$\therefore P(\bar{P}|\bar{B}) = 0.8$$

$$(vi) P(\text{odd}) = 0.9 \times P(\text{even})$$

$\hookrightarrow \{2, 4, 6\}$

$$P(\text{even}) > 3 = 0.75 \quad P(>3) = ?$$

Let,  $P(\text{even}) = x$   $P(2) = P(4) = P(6) = x$

$$\therefore P(\text{even}) = 3x$$

$$\Rightarrow P(\text{odd}) = 0.9 \times 3x = 0.27 \quad 2.7x$$

$$\left. \begin{array}{l} P(\text{odd}) + P(\text{even}) = 1 \\ 0.27 + 3x = 1 \end{array} \right\} x = \frac{1}{5.7} = P(2) = P(4) = P(6)$$

$$\text{Now, } P(\text{even} | > 3) = 0.75$$

$$\Rightarrow \frac{P(\text{even}, > 3)}{P(>3)} = \frac{2x}{2.7} = 0.75$$

Didnt get it

$$\Rightarrow P(>3) = \frac{0.75}{2x} = \frac{0.75}{\frac{2}{5.7}} = 0.468$$

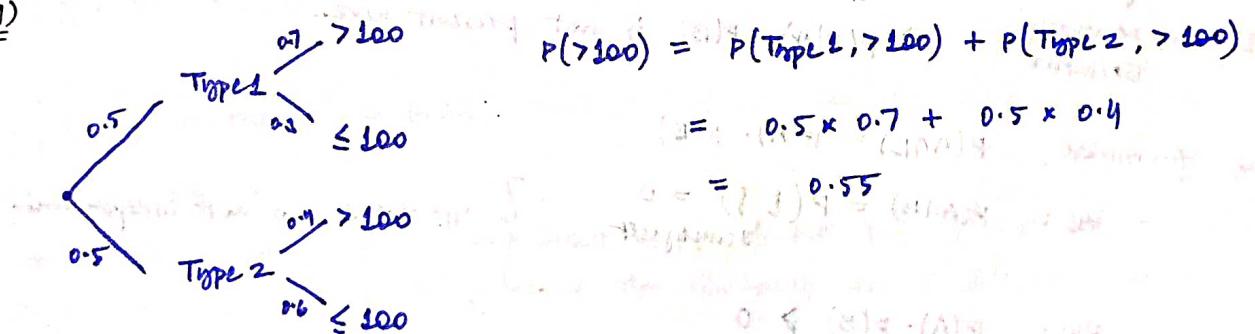
OR

$$P(\text{odd}) + P(\text{even}) = 0.9x + x = 1 \Rightarrow x = \frac{1}{1.9} \quad P(\text{even}) = \frac{1}{1.9}$$

$$P(\text{even} | > 3) = \frac{P(\text{even}, > 3)}{P(>3)} = \frac{P(>3 | \text{even}) P(\text{even})}{P(>3)}$$

$$\Rightarrow 0.75 = \frac{\frac{2}{3} \times \frac{1}{1.9}}{P(>3)} \Rightarrow P(>3) = 0.468$$

(vii)



## Independence of Events

till now we have been seeing  $P(A|B) = \frac{P(A, B)}{P(B)}$   
 given that B has happened, the prob of A.

the occurrence of B updates  
 the prob of occurrence of A  
 our belief are changing

- Do we always update our beliefs?

A: I had a sandwich for breakfast.

B: It will rain today.

If A occurs, will you update your belief about B?

↪ NO, occurrence of A provides no info about B.

$$\Rightarrow P(B|A) = P(B)$$

There is no impact of occurrence of A over B.

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

⇒ In general,

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

For independent events,

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

comparing both  $\Rightarrow P(B) = P(B|A)$

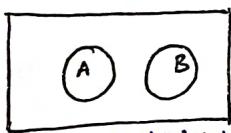
$\hookrightarrow P(A) > 0$

bcz A is in denominator

Ques. Are these events Independent? from the diag. they may look independent  
 but they are not

NO

(ii)



A, B are disjoint

if A has occurred, the prob of occurrence of B is 0, B can never occur  
 so, they are extremely dependent.

$$P(B|A) = 0 \quad \text{Hence, } P(A|B) = 0.$$

so, A clearly conveys info about B

They are Mutually Exclusive

$$\Rightarrow P(B|A) = P(B) \text{ is not present here.}$$

using formulae,  $P(A \cap B) = P(A) \cdot P(B)$

$$\text{LHS: } P(A \cap B) = P(\{\}) = 0$$

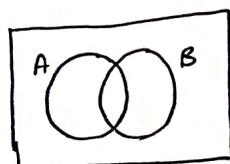
↳ empty set

} RHS  $\neq$  RHS, so not independent

$$\text{RHS: } P(A) \cdot P(B) \geq 0$$

$$\begin{matrix} \downarrow \\ > 0 \end{matrix} \quad \begin{matrix} \downarrow \\ > 0 \end{matrix}$$

(iii)



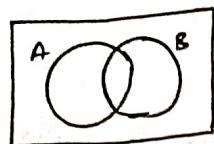
for indep:  $P(A \cap B) = P(A) \cdot P(B)$

→ we can't say anything here.

iii)

$$P(A) = 0.30$$

$$P(B) = 0.40 \quad P(A \cap B) = 0.12$$



$$P(A \cap B) = 0.12$$

$$P(A) \cdot P(B) = 0.30 \times 0.40 = 0.12$$

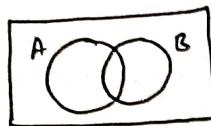
$P(A \cap B) = P(A) \cdot P(B) \rightarrow \text{Independent} \Leftrightarrow$

iv)

$$P(A) = 0.30$$

$$P(B) = 0.40$$

$$P(A \cap B) = 0.20$$



$$P(A) \cdot P(B) = 0.12 \neq P(A \cap B)$$

$\Rightarrow \text{Not Independent}$

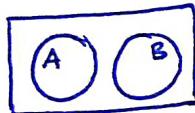
$\rightarrow$  Use  $P(A \cap B) = P(A) \cdot P(B)$  to check Indep. events.

Venn diag. doesn't reveal anything

### Mutually Exclusive Events and Independent Events.

Mutually Exclusive events:

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



Independent Events:

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

$\rightarrow$  Two ME events are indep. iff  $P(A) = 0$  or  $P(B) = 0$

Inclusion-Exclusion principle:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

for Indep. Events:

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

for ME events:

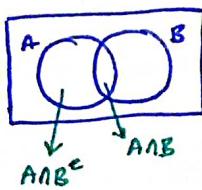
$$P(A \cup B) = P(A) + P(B)$$

**NOTE** → If A and B are indep. then  $A$  and  $B^c$  are also indep.

Intuitive Arg →

If A does not change your belief about the likelihood that B occurred then A should not change your belief about the likelihood that B does not occur.

Formal Proof → we want to prove if  $P(A \cap B) = P(A)P(B)$ , then  $P(A \cap B^c) = P(A)P(B^c)$



$$\text{Proof: } A = (A \cap B) \cup (A \cap B^c)$$

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

$$P(A) = P(A)P(B) + P(A)P(B^c)$$

$$P(A \cap B^c) = P(A)[1 - P(B)]$$

$$P(A \cap B^c) = P(A) \cdot P(B^c) \Leftrightarrow \text{Hence proved}$$

If Present does not change your belief about Future that past happened then Present should not change your belief about Future that in past that didn't happen

**NOTE**

$$\begin{array}{c} A, B \text{ are indep.} \Leftrightarrow A^c, B \text{ are indep.} \\ \Downarrow \\ A, B^c \text{ are indep.} \quad A^c, B^c \text{ are indep.} \end{array}$$

### Independence of 3 events

indep of 2 events:  $P(A \cap B) = P(A) \cdot P(B)$

events A, B and C are said to be independent

if all of the following eq's hold:

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

$$P(A \cap C) = P(A) \cdot P(C),$$

$$P(B \cap C) = P(B) \cdot P(C),$$

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

### Indep. of 4 events A, B, C, D

any pair of 2:  $P(A \cap B) = P(A) \cdot P(B)$

any pair of 3:  $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

$$P(A \cap B \cap C \cap D) = P(A) \cdot P(B) \cdot P(C) \cdot P(D)$$

Ques. True/False

i) If collection of events are independent then they are pairwise independent.

**TRUE**

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

$$P(A \cap C) = P(A) \cdot P(C)$$

$$P(B \cap C) = P(B) \cdot P(C)$$

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

if this is given  $\rightarrow$  then A, B

A, C

B, C

are gonna be  
independent

ii) If collection of events are pairwise independent then they are independent.

**FALSE**

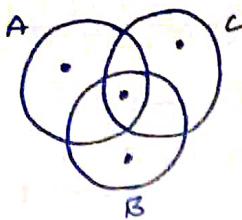
$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A \cap C) = P(A) \cdot P(C)$$

$$P(B \cap C) = P(B) \cdot P(C)$$

but  $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$  is not given, so not independent  
only pairwise independent

(ex)



⇒ A, B, C pairwise indep? YES

⇒ A, B, C independent? NO

$$P(A \cap B) = P(A) \cdot P(B) \Leftrightarrow P(B \cap C) = P(B) \cdot P(C) \Leftrightarrow P(A \cap C) = P(A) \cdot P(C) \Leftrightarrow$$

$$\frac{1}{4} = \frac{2}{4} \cdot \frac{2}{4} \quad \frac{1}{4} = \frac{2}{4} \cdot \frac{2}{4} \quad \frac{1}{4} = \frac{2}{4} \cdot \frac{2}{4}$$

But  $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$  → does not hold.

$$\frac{1}{4} = \frac{2}{4} \cdot \frac{2}{4} \cdot \frac{2}{4}$$

### Independence and Pair wise Independence

- Two independent fair coin tosses

Event A : First toss is H

Event B : Second toss is H.

Event C : 1st and 2nd toss gives same result.

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

$$- P(C) = \frac{2}{4} = \frac{1}{2}$$

$$\Rightarrow P(A \cap B) = P(A) \cdot P(B) \quad \frac{1}{4} \quad \frac{2}{4} \quad \frac{2}{4}$$

$$P(A \cap C) = P(A) \cdot P(C) \quad \frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{2}$$

$$P(B \cap C) = P(B) \cdot P(C) \quad \frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{2}$$

$$\Rightarrow P(A \cap B \cap C) \neq P(A) \cdot P(B) \cdot P(C) \quad \frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2}$$

⇒ A, B, C are pairwise indep.

But not indep.

The point is that just knowing about A or just knowing about B tells us nothing about C, but knowing what happened with both A & B gives us info about C. (in fact, in this case gives us perfect info about C)

⇒ Pairwise Indep doesn't imply Independence.

Pairwise

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

3-wise

$$P(A \cap C) = P(A) \cdot P(C)$$



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

k-wise

$$P(B \cap C) = P(B) \cdot P(C)$$

Ques. T/F

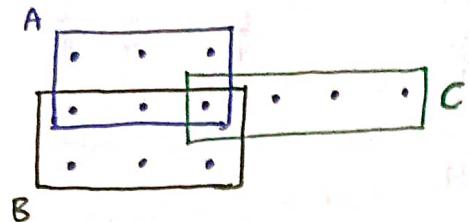
If we have  $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$  then we can also say pairwise indep.

i.e.

$$P(A \cap B \cap C) = P(A) P(B) P(C) \Rightarrow P(A \cap B) = P(A) P(B) \text{ and}$$

$$P(A \cap C) = P(A) P(C) \text{ and}$$

$$P(B \cap C) = P(B) P(C)$$



$$P(A \cap B \cap C) = P(A) P(B) P(C) \Leftrightarrow$$

$$\frac{1}{12} \quad \frac{6}{12} \times \frac{6}{12} \times \frac{4}{12}$$

$$P(A \cap B) \neq P(A) P(B) \text{ if } P(A \cap C) \leq P(B \cap C)$$

$$\frac{3}{12} \quad \frac{6}{12} \quad \frac{6}{12}$$

**[NOTE]** Neither pairwise independence implies independence nor implied by independence

2-wise

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



3-wise

$$P(A \cap B \cap C) = P(A) P(B) P(C)$$



o

$k$ -wise  $\not\Rightarrow$   $k'$ -wise

**(2)** consider 2 events  $A \& B$  such that  $P(A \cap B) \neq P(A) P(B)$  i.e. they are dependent events.

Now consider a 3rd event  $C = \emptyset$ .

$$P(A \cap B \cap C) = 0 \text{ since } C = \emptyset \text{ and } P(A) \cdot P(B) \cdot P(C) = 0 \text{ since } P(C) = 0.$$

(So, Although every set of 3 events in this ~~collection~~ collection (there is only one set) of 3 events has the independence property, this collection is not pairwise independent)

Ques. Are  $A$  and  $A^c$  independent if  $0 < P(A) < 1$ ?

**Answer 1:** Intuitively, no. If you know  $A$  happens, then you know  $A^c$  does not happen

**Answer 2:** Formally,  $P(A \cap A^c) = P(\emptyset) = 0$ .

if  $0 < P(A) < 1$ , then  $P(A) P(A^c) \neq 0$

o  $A \& A^c$  are extremely dependent

## CONDITIONAL INDEPENDENCE

Two events A and B are independent given C :

$$P(A \cap B | C) = P(A|C) \cdot P(B|C)$$

(OR)

$$P(B|C) = P(B|A, C)$$

→ A and B are independent given C.

Earlier Normally, it was like  $P(B) = P(B|A)$ .

$$\text{Independence: } P(A \cap B) = P(A) \cdot P(B)$$

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

Now,

$$P(A \cap B | C) = P(A|C) \cdot P(B|A, C)$$

Comparing them

Earlier, Independence

$$P(A \cap B) = P(A) \cdot P(B) \quad \text{--- ①}$$

$$P(A \cap B) = P(A) \cdot P(B|A) \quad \text{--- ②}$$

$$= P(B) \cdot P(A|B) \quad \text{--- ③}$$

$$1 \& 2 \Rightarrow P(B) = P(B|A)$$

$$1 \& 3 \Rightarrow P(A) = P(A|B)$$

conditional Independence

$$P(A \cap B | C) = P(A|C) \cdot P(B|C) \quad \text{--- ①}$$

$$P(A \cap B | C) = P(A|C) \cdot P(B|A, C) \quad \text{--- ②}$$

$$= P(B|C) \cdot P(A|B, C) \quad \text{--- ③}$$

$$1 \& 2 \Rightarrow P(B|C) = P(B|A, C)$$

$$1 \& 3 \Rightarrow P(A|C) = P(A|B, C)$$

NOTE →

if A and B are indep then

A and B' are indep

A' and B are indep

A' and B' are indep

if A and B are indep given C then

A and B' are indep given C

A' and B are indep given C

A' and B' are indep given C

ans. T/F.

$$P(A \cap B) = P(A) \cdot P(B) \Rightarrow P(A \cap B | C) = P(A|C) \cdot P(B|C) ?$$

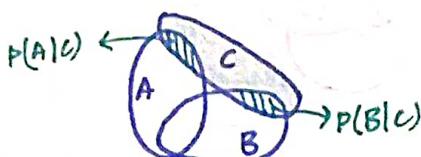
**FALSE**

(e) Assume A and B are indep i.e.  $P(A \cap B) = P(A) \cdot P(B)$

$$\text{here, } P(A \cap B | C) = 0$$

$$\text{K } P(A|C) \cdot P(B|C) > 0 \\ > 0 \quad > 0$$

$$\} \Rightarrow P(A \cap B | C) \neq P(A|C) \cdot P(B|C)$$



so, A and B are not independent given C.

earlier A and B were indep but after C they got dependent.

(Ex) Independence doesn't imply conditional Independence

Suppose that my friends Alice & Bob are the only 2 people who ever call me. Each day, they decide independently whether to call me : letting A be the event that Alice calls and B be the event that Bob calls

A & B are unconditionally independent

But suppose that I hear the phone ringing now.

conditional on this observation, A & B are no longer independent :

if the phone call isn't from Alice, it must be from Bob. In other words, letting R be the event that the phone is Ringing,

we have  $P(B|R) < 1$ ,  $P(B|A^c, R) = 1$

so, B and  $A^c$  are not conditionally indep given R, and likewise for A & B.

→ suppose there are only 2 people Alice & Bob who call me.

And both of them calls me independently

A : Alice's call

$$P(A \cap B) = P(A) P(B) \quad (\text{given that both called})$$

B : Bob's call

R : Phone Ring

Suppose that phone rings then -

$P(A|R)$  be the prob that Alice has called given the phone rings

we can say that

$$P(A|B^c, R) = 1 \neq P(A|R)$$

so,  $P(A|R) \neq P(A|B^c, R) \Rightarrow$  not conditional Indep

→ earlier A & B were indep,

but after C they just got dependent.

$$P(A \cap B) = P(A) P(B) \Rightarrow P(A \cap B|C) = P(A|C) P(B|C)$$



(Ex) consider rolling a die & let,

$$A = \{1, 2\} \quad B = \{2, 4, 6\} \quad C = \{1, 4\}$$

$$\text{Thus, we have } P(A) = \frac{1}{3}, \quad P(B) = \frac{1}{2}$$

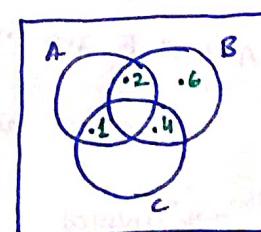
$$P(A \cap B) = \frac{1}{6} = P(A) \cdot P(B)$$

Thus, A & B are indep., but we have

$$P(A|C) = \frac{1}{2}, \quad P(B|C) = \frac{1}{2}; \quad P(A \cap B|C) = P(\{2\}|C) = 0$$

$$\text{Thus, } P(A \cap B|C) \neq P(A|C) P(B|C)$$

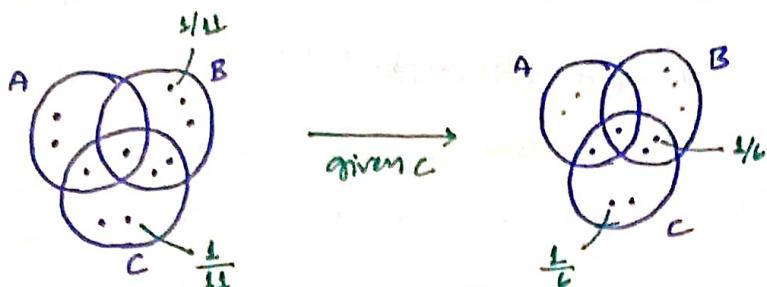
which means A & B are not conditionally indep given C.



(ex)

## Conditional Independence doesn't imply Independence

i.e.  $P(A \cap B|C) = P(A|C)P(B|C) \Rightarrow P(A \cap B) = P(A)P(B)$



$$P(A \cap B|C) = P(A|C)P(B|C)$$

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

even though  
They are conditionally  
independent

$$P(A \cap B) \neq P(A)P(B)$$

$$\frac{1}{12} \neq \frac{1}{3} \times \frac{1}{2}$$

they are not indep.

Ques. T/F . Given the following factorization →

$$P(A, B, C) = P(C)P(B|C)P(A|C)$$

can we say A & B are independent given C ?

TRUE

↪  $P(A, B, C) = P(C)P(B|C) \underbrace{P(A|B, C)}_{\downarrow}$  (general formula)

but here we have,  $P(A|C)$

comparing both,  $P(A|C) = P(A|B, C) \Rightarrow$  A & B are conditional independent.

$$\Rightarrow A \perp B | C$$

Ques. Suppose A and B are conditionally independent given C then which of the following is/are true?

a)  $P(A \cap B) = P(A)P(B)$  ✗

b)  $P(A, B, C) = P(A|B, C)P(B|C)P(C)$  ✗

c)  $P(A, B, C) = P(A|C)P(B|C)P(C)$  ✓

d)  $P(A, B|C) = P(A|C)P(B|C)$  ✓

↪  $P(A, B, C) = P(C) \cdot P(B|C) \cdot P(A|B, C) \rightarrow$  if A & B are indep given C, then  
 $P(A|B, C) = P(A|C)$

$\underline{P(A, B, C)} = P(C) \cdot P(A, B|C)$

$$\rightarrow P(A, B, C) = P(C)P(B|C)P(A|C)$$

↓  
can consider whole thing  
as one event

- Ques. Mark True to all the expressions that are equal to  $P(A|B)$ , given no independence assumption.
- (i)  $P(A, C|B) + P(A, C^c|B)$  TRUE
  - ↳  $P(A|B) = P(A, C|B) + P(A, C^c|B)$  Using marginalization
  - (ii)  $\frac{P(A, C|B)}{P(C|B)}$  FALSE
  - ↳ 
$$\frac{P(A, C|B)}{P(C|B)} = \frac{\frac{P(A, B, C)}{P(B)}}{\frac{P(C, B)}{P(B)}} = \frac{P(A, B, C)}{P(B, C)} = P(A|B, C) \neq P(A|B)$$
  - (iii)  $P(A|B, C) + P(A|B, C^c) \neq P(A|B)$  FALSE  
 ↳ this is not equal to  $P(A|B)$ , bcz marginalization doesn't work on the given that side.  
 if it was  $P(A, C|B) + P(A, C^c|B)$  then it were true.
  - (iv)  $\frac{P(A, B, C) + P(A, B, C^c)}{P(B, C) + P(B, C^c)}$  TRUE  
 ↳ Using marginalization
  - (v)  $\frac{P(B|A) * P(A|C)}{P(B, C) + P(B, C^c)}$  FALSE  
 ↳ 
$$\frac{P(B|A) P(A|C)}{P(B)} \neq P(A|B)$$
  - (vi)  $\frac{P(A|C, B) P(C|A, B)}{P(C|B)}$  FALSE  
 ↳ 
$$\frac{\frac{P(A, B, C)}{P(B, C)} \cdot \frac{P(A, B, C)}{P(A|B)}}{\frac{P(B, C)}{P(B)}} \neq P(A|B)$$

Ques. Mark all expressions that are equal to  $P(A, B, C)$ , given that  $A \perp B$  ( $A$  &  $B$  are indep.).

(i)  $P(A) P(B) P(C|A, B)$  TRUE

$$\hookrightarrow P(A, B, C) = P(A) \underbrace{P(B|A)}_{\downarrow} P(C|A, B)$$
$$= P(A) P(B) P(C|A, B)$$

(ii)  $P(C) P(A|C) P(B|C)$  not equal

$$\hookrightarrow P(A, B, C) = P(C) \cdot P(A|C) \cdot P(B|A, C)$$

now  $P(B|A, C) \neq P(B|C)$   
so given expression is FALSE

(iii)  $P(A) P(B|A) P(C|A, B)$  TRUE

$$\hookrightarrow P(A, B, C) = P(A) P(B|A) P(C|A, B)$$

(iv)  $\underbrace{P(A|C) \cdot P(C|B)}_{\text{not equal}} \cdot P(B)$  not equal

$$\hookrightarrow P(A, B, C) = P(B) P(C|B) \underbrace{P(A|B, C)}_{\text{not equal}}$$

now,  $P(A|B, C) \neq P(A|C)$   
so given expression is FALSE

(v)  $P(A) P(C|A) \underbrace{P(B|C)}_{\text{not equal}}$  not equal

$$\hookrightarrow P(A, B, C) = P(A) P(C|A) P(B|A, C)$$

$P(B|C) \neq P(B|A, C)$   
so given exp is FALSE

(vi)  $P(A, C) P(B|A, C)$  TRUE

$$\hookrightarrow P(B, A, C) = P(A, C) P(B|A, C)$$

take this as  
single event

Ques. If  $A \perp B|C$ . then  $P(A, B|C) = ?$

(i)  $P(A|C) P(B|C)$  TRUE

$$P(A, B|C) = P(A|C) P(B|C)$$

(ii)  $\frac{P(C) P(B|C) P(A|C)}{P(C|A, B)}$  FALSE

$$\hookrightarrow \frac{P(C) \cdot \frac{P(B|C)}{P(C)} \cdot \frac{P(A|C)}{P(C)}}{\frac{P(C|A, B)}{P(A, B)}} \neq \frac{P(A, B, C)}{P(C)}$$

$A \perp B$  means

$$P(A, B) = P(A) \cdot P(B)$$

$$\text{or } P(A|B) = P(A)$$

$$\text{or } P(B|A) = P(B)$$

general formula

(iii)  $\frac{P(C, A|B) P(B)}{P(C)}$  TRUE

$$\hookrightarrow P(A, B|C) = \frac{P(A, B, C)}{P(C)} = \frac{P(B) P(C, A|B)}{P(C)}$$

(iv)  $P(A|B) P(B|C)$  FALSE

$\hookrightarrow$  it should have been  $P(A|C) P(B|C)$

# RANDOM VARIABLES INTRO

Even though this chapter may appear to cover a lot of new ground, but this is not really the case.

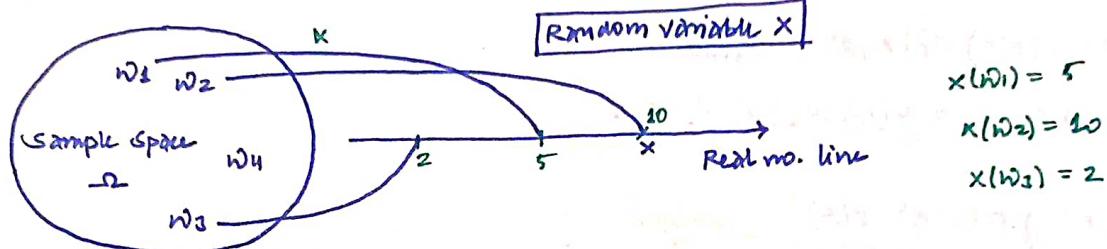
In this chapter, we'll simply take concepts from Chapter 1 (probabilities, conditioning, independencies etc) & apply them on Random variables, rather than events.

The only genuine new concepts relate to means & variances.

## Random Variable

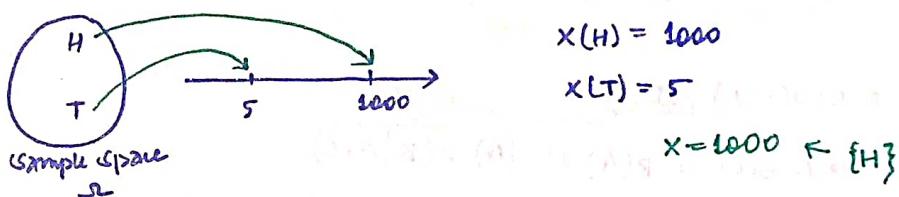
Function that maps an outcome to the Real no.

$w_1, w_2, w_3, w_4$   
are outcomes.



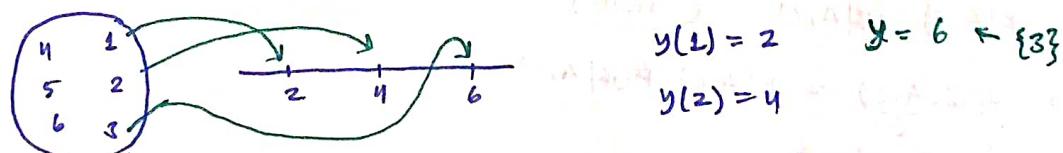
### (ex) tossing a coin

possible outcomes : {H, T}



### (ex) Rolling a die

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



Q Why are we calling it "Random variable" when it is a func?

↳ 'Random' bcoz we are assigning some func to Random experiment

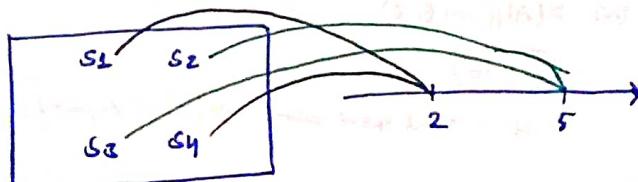
'Variable' bcoz we want to treat this as variable.

We don't want to write  $y(1)=2$ , but  $y=2$

$x(H)=1000$ , but  $x=1000$

} mathematical convenience

### (ex)



$$x \leq 5 = \{s_1, s_2, s_3, s_4\}$$

$$x < 5 = \{s_1, s_4\}$$

$$x > 0 = \{s_1, s_2, s_3, s_4\}$$

$x=5$  is a set.

$$\{s_2, s_3\}$$

$x=2$  is a set

$$\{s_1, s_3\}$$

$x=5 \leftarrow \{s_2, s_3\} \rightarrow$  associated set of outcomes related to  $x=5$

$$x=5 = \{w \mid x(w)=5\}$$

(e) tossing a coin

	$x(\cdot)$
H	5
T	10

{func" mapping  
Head to 5,  
Tail to 10.

$$\text{now, } P(X=5)$$

$$\begin{aligned} &= P(\{\omega \mid x(\omega)=5\}) \\ &= P(\{H\}) \\ &= \frac{1}{2} \end{aligned}$$

(e) Rolling a dice

	$P(\cdot)$	$x(\cdot)$
1	$\frac{1}{6}$	1
2	$\frac{1}{6}$	1
3	$\frac{1}{6}$	1
4	$\frac{1}{6}$	2
5	$\frac{1}{6}$	2
6	$\frac{1}{6}$	2

$$\circ P(\text{outcome is 3}) = \frac{1}{6}$$

$$\begin{aligned} \circ P(X=1) &= P(\{\omega \mid x(\omega)=1\}) \\ &= P(\{1, 2, 3\}) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2} \end{aligned}$$

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

	$P(\cdot)$	$x(\cdot)$
1	$\frac{1}{12}$	1
2	$\frac{1}{6}$	2
3	$\frac{1}{12}$	2
4	$\frac{4}{12}$	1
5	$\frac{4}{24}$	2
6	$\frac{4}{24}$	1

$$\begin{aligned} \circ P(X=1) &= P(1) + P(3) + P(4) + P(6) \\ &= \frac{1}{12} + \frac{1}{12} + \frac{4}{12} + \frac{4}{24} \\ &= \frac{8}{12} = 0.67 \end{aligned}$$

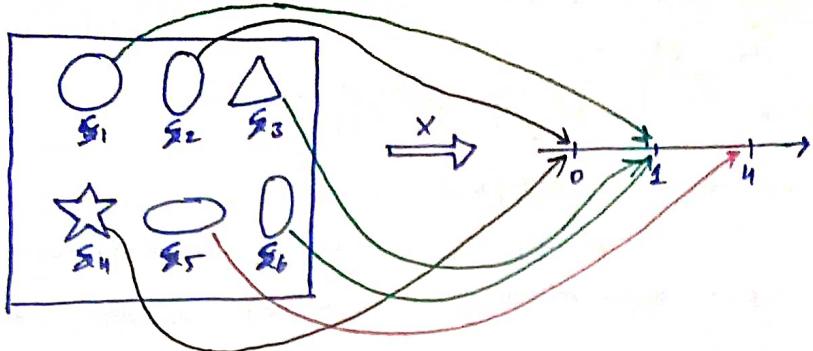
(e) Possible outcomes in Sample space  $\equiv \{1, 3, 7, 12\}$

	$P(\cdot)$	$x(\cdot)$
1	$\frac{1}{2}$	1
3	$\frac{1}{4}$	1
7	$\frac{1}{8}$	1
12	$\frac{1}{8}$	1

$$P(X=1) \rightarrow P(\{\omega \mid x(\omega)=1\}) = 1$$

$$P(X=2) = 0$$

@2



Assume  $\{s_1, s_2, \dots, s_6\}$   
is equally likely  $\forall s_i$   
 $P(s_1) = P(s_2) = \dots = P(s_6) = \frac{1}{6}$

$$\circ P(X < 1) = P(\{s_2, s_4\}) = P(s_2) + P(s_4) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$$\circ P(X < 2) = \frac{5}{6}$$

$$\circ P(1 < X < 2) = 0$$

$$\circ P(X \leq 4) = 1$$

$$\rightarrow \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1 \Leftarrow$$

**[NOTE]** → Any condition on RV is an event.

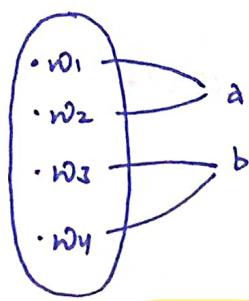
@2  $x < 2 \equiv \{w | X(w) < 2\}$   
 $\uparrow$  set of some outcomes  $\equiv$  EVENT  
 $x = 3 \equiv \{w | X(w) = 3\}$

$x=2$  is an event  $\Leftarrow$   
 $x < 2$  is an event  $\Leftarrow$   
 $x = 3$  is an event  $\Leftarrow$

Ques. T/F.

Suppose we have 2 events  $x=a$ ,  $x=b$ . We can say that both of the events are mutually exclusive.  $\rightarrow$  there is no common outcome.

Let.



$$x = a \rightarrow \{w_1, w_2\}$$

$$x = b \rightarrow \{w_3, w_4\}$$

**RV is a "funct"**. A "funct" cannot give multiple outputs to an input & every one input must have an output

$$f(x) = x^2 \quad f: R \rightarrow R^+$$

$$f(2) = 4$$

$$f(3) = 9$$

$$f(-2) = 4$$

$$\begin{array}{l} \cdot 2 \rightarrow 4 \\ \cdot -2 \rightarrow 4 \end{array} \quad \text{funct} \Leftarrow$$

$$\begin{array}{l} \cdot 2 \rightarrow 4 \\ \cdot 2 \rightarrow 5 \end{array} \quad \text{not even a funct}$$

Ques. T/F

$X$  maps any outcome to either  $a, b$  or  $c$

Thus  $x=a, x=b$  and  $x=c$ , these 3 events are mutually exclusive & exhaustive.

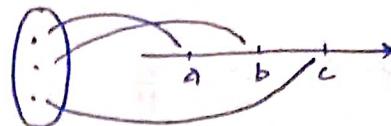
covers whole sample space

TRUE

According to the func' def it is True

Any outcome is either mapped to  $x=a$  or  $x=b$  or  $x=c$ .

so all the outcomes are getting covered.



→  $x=a$  is an event

→  $x=a, x=b$  are mutually exclusive

→  $x=a, x=b, x=c$  are mutually exhaustive

(suppose only 3 possibilities)

→  $x=a, x=b, x=c$  are partitioning the sample space

②  $P(y=2) = P(y=2, x=a) + P(y=2, x=b) + P(y=2, x=c)$  Marginalization

this is nothing  
just events  
as seen before

{ earlier we use to marginalize with events.  
②  $P(y=2) = P(y=2, A_1) + P(y=2, A_2) + P(y=2, A_3)$  }

events

Ques. T/F .  $P(x=5)$  is same as  $\sum_k P(x=5, y=k)$

TRUE

② Coin Toss.

let  $x$  be a RV which takes value 1 on Head and 0 on tail.

what is the prob of  $x$  being 1?

	x	
H	1/2	1
T	1/2	0

$$P(x=1) = 1/2$$

Coin Toss Twice

$x = \text{no. of H after both flips}$

$x_1 = \text{no. of H on 1st flip}$

$x_2 = \text{no. of H on 2nd flip}$

$$P(x=2) = ?$$

$$P(x_1 + x_2 < 2) = ?$$

$$P(x=2) = 3/4$$

$$P(x_1 + x_2 < 2) = 1/4 + 1/4 + 1/4 = 3/4$$

it is also a RV but we call

it as func' of RV

$$y = x_1 + x_2$$

$$\text{i.e. } y(w) = x_1(w) + x_2(w)$$

$P(\cdot)$		x	$x_1$	$x_2$	y
$y_1$	HH	2	1	1	2
$y_2$	HT	1	1	0	1
$y_3$	TH	1	0	1	1
$y_4$	TT	0	0	0	0

outcomes  
when  $y < 2$

## (ex) Two Dice Tossed

$X = \text{"sum of the scores on the 2 dice"}$

Possible values of  $X$ ?

↳ 2 to 12

$$\text{now } P(X=2) = \frac{1}{36} \quad \text{P}(X=2) = P((1,1)) = \frac{1}{36}$$

$$P(X=3) = \frac{2}{36} = \frac{1}{18}$$

$\downarrow$   
 $(1,2), (2,1)$

Ques.  $P(X=x) = \begin{cases} \frac{1}{9}, & \text{if } x \text{ is an integer in the range } [-4, 4] \\ 0, & \text{otherwise} \end{cases}$

let  $y = |x|$ . Find out  $P(y=2)$ . let  $Z = X^2$ . Find  $P(Z=16)$

<u>PMF for <math>X</math></u>		$y$	$Z$
$x$	$P(X=k)$		
-4	$\frac{1}{9}$	4	16
-3	$\frac{1}{9}$	3	9
-2	$\frac{1}{9}$	2	4
-1	$\frac{1}{9}$	1	1
0	$\frac{1}{9}$	0	0
1	$\frac{1}{9}$	1	1
2	$\frac{1}{9}$	2	4
3	$\frac{1}{9}$	3	9
4	$\frac{1}{9}$	4	16

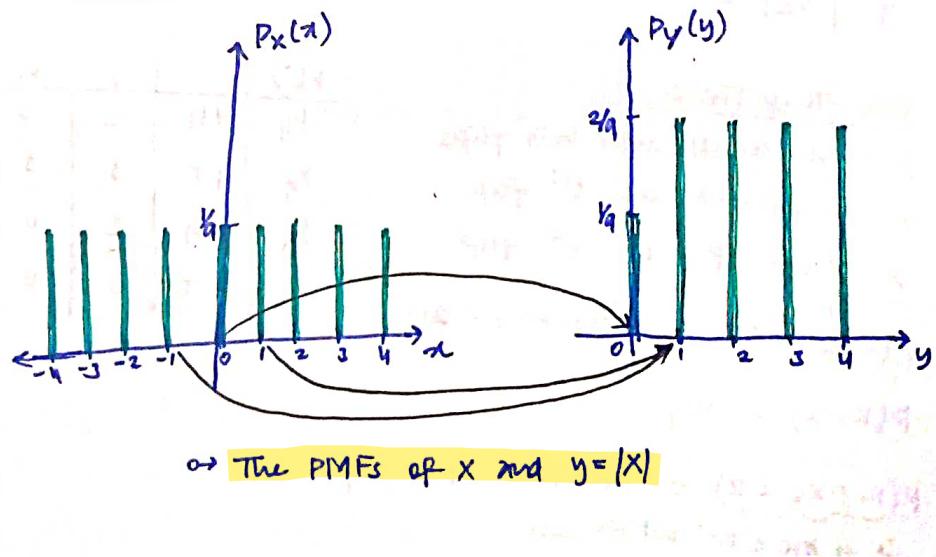
- $P(X<0) = \frac{4}{9}$
- $P(Y=2) = \frac{2}{9} \quad \checkmark \quad P(Y=2) = P(X=2 \text{ or } X=-2)$
- $P(Y=2) = P(X=2) + P(X=-2) = \frac{1}{9} + \frac{1}{9} = \frac{2}{9}$
- $P(Z=16) = P(X=4 \text{ or } X=-4) = \frac{1}{9} + \frac{1}{9} = \frac{2}{9}$

(X=2) → y=2  
(X=-2) → y=2  
(X=4) → z=16  
(X=-4) → z=16

→ Prob. Mass Function  
PMF for  $Z$

$Z$	$P(\cdot)$
$x=0 \rightarrow 0$	$\frac{1}{9}$
$x=1 \rightarrow 1$	$\frac{2}{9}$
$x=-1 \rightarrow -1$	$\frac{2}{9}$
$x=2 \rightarrow 4$	$\frac{2}{9}$
$x=-2 \rightarrow 4$	$\frac{2}{9}$
$x=3 \rightarrow 9$	$\frac{2}{9}$
$x=-3 \rightarrow 9$	$\frac{2}{9}$
$x=4 \rightarrow 16$	$\frac{2}{9}$
$x=-4 \rightarrow 16$	$\frac{2}{9}$

→ in graphical representation



Ques. Let  $X$  and  $Z$  are 2 RV.

$y = X \oplus Z$ . Find  $P(y=y)$

	$x=0$	$x=1$
$P(x=x)$	$p$	$1-p$

	$z=0$	$z=1$
$P(z=z)$	$a$	$1-a$

$$\hookrightarrow \begin{array}{ll} x : P(x=0) = p & P(x=1) = 1-p \\ z : P(z=0) = a & P(z=1) = 1-a \end{array} \quad y = X \oplus Z \quad P(y=0) = ? \quad P(y=1) = ?$$

**Method 1** →

$x$	$z$	$y = x \oplus z$
0	0	0 $\text{PAV}$
0	1	1 $P(1-a)$
1	0	1 $(1-p)a$
1	1	0 $(1-p)(1-a)$

$$P(y=0) = pa + (1-p)(1-a)$$

$$P(y=1) = p(1-a) + (1-p)a$$

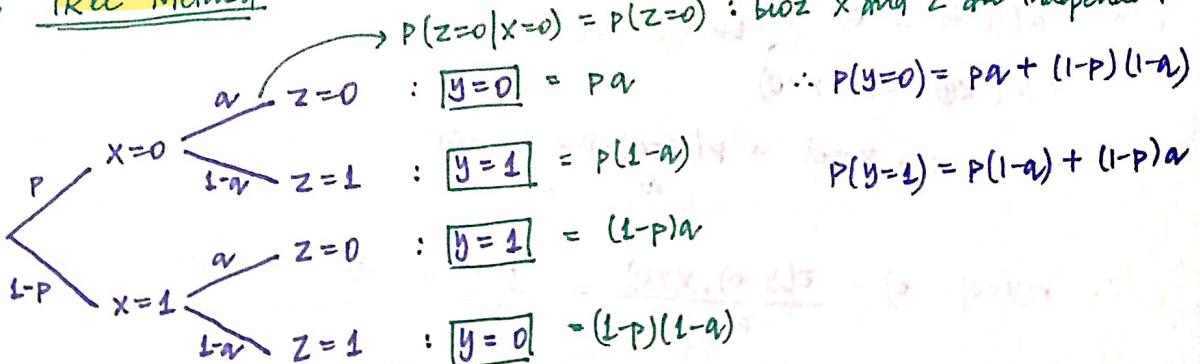
**Method 2** →

$$\begin{aligned} P(y=0) &= P((x=0, z=0) \text{ or } (x=1, z=1)) \quad \text{bcoz both events are disjoint} \\ &= P(x=0, z=0) + P(x=1, z=1) \\ &= P(x=0) \cdot P(z=0) + P(x=1) \cdot P(z=1) \\ &= pa + (1-p)(1-a) \end{aligned}$$

$$\text{Hence } P(y=1) = p(1-a) + (1-p)a$$

**Method 3** →

Tree Method



Ques. Variation :

$$P(y=0 | X=0) = ?$$

**Method 1** →

$x$	$z$	$y$
0	0	0
0	1	1
1	0	1
1	1	0

$$P(y=0 | X=0)$$

$$= P(X \oplus Z=0 | X=0)$$

$$= P(Z=0 | X=0) \rightarrow \frac{P(X=0, Z=0)}{P(X=0)} = \frac{pa}{p} = a$$

$$= a$$

### Method 2 :

$$\underline{x=0, z=0}$$

↓

$$pa$$

↓

$$\underline{x=0, z=1}$$

↓

$$p(1-a)$$

↓

$$\underline{x=1, z=0}$$

↓

$$(1-p)a$$

↓

$$0$$

$$\underline{x=1, z=1}$$

↓

$$(1-p)(1-a)$$

↓

$$0$$

Earlier :  
prob.

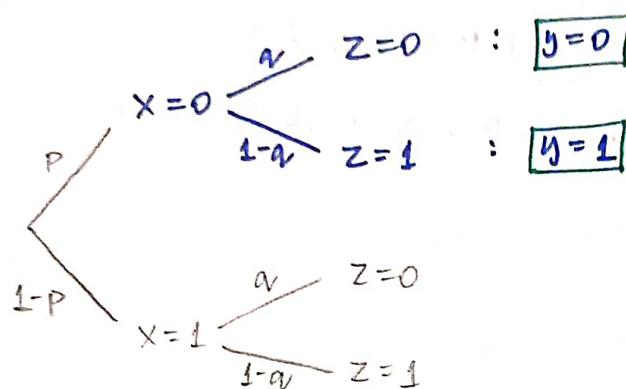
updated  
when  
prob given  
 $x=0$ :

$$\frac{pa}{p} = a$$

$$\frac{p(1-a)}{p} = (1-a)$$

prob has been redistributed :  $a + (1-a) = 1$

### Method 3 : TREE Method



$$P(y=0 | x=0) = a$$

$$P(y=1 | x=0) = 1-a$$

whatever  
written with pencil  
is unnecessary and has  
been removed.

### Method 4 : Using Plain Formula

$$P(y=0 | x=0) = \frac{P(y=0, x=0)}{P(x=0)}$$

$x \& z$  are indep.  
 $y \& x$  are not indep.  
so, we can't write

$$P(y=0, x=0) = P(y=0) \cdot P(x=0)$$

$$\begin{aligned} \text{now, } P(y=0, x=0) &\rightarrow \text{bcoz } y=x \oplus z \\ &= P(x \oplus z=0, x=0) \\ &= P(z=0, x=0) = \underset{a}{P(z=0)} \cdot \underset{P}{P(x=0)} = aP \end{aligned}$$

$$\therefore P(y=0 | x=0) = \frac{P(y=0, x=0)}{P(x=0)} = \frac{Pa}{P} = a$$

Ques. Consider 2 RV  $X$  and  $Y$  with joint prob distribution given in the table below:

	$y=0$	$y=1$	$y=2$
$x=0$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$
$x=1$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$

$$(i) P(X=0, Y=1) = \frac{1}{4}$$

$$(ii) P(X=0, Y \leq 1) = \frac{1}{6} + \frac{1}{4} = \frac{5}{12}$$

$$(iii) \left\{ \begin{array}{l} P(X=0) = \frac{1}{6} + \frac{1}{4} + \frac{1}{8} = \frac{13}{24} \\ \text{PMF of } X \end{array} \right.$$

$$(iv) \left\{ \begin{array}{l} P(X=1) = \frac{1}{8} + \frac{1}{6} + \frac{1}{6} = \frac{11}{24} \\ \text{PMF of } X \end{array} \right.$$

$$P(X=0) = P(X=0, Y=0) + P(X=0, Y=1) + P(X=0, Y=2)$$

$$\frac{1}{6} + \frac{1}{4} + \frac{1}{8}$$

$$(v) \left\{ \begin{array}{l} P(Y=0) = \frac{1}{6} + \frac{1}{8} = \frac{7}{24} \\ \text{PMF of } Y \end{array} \right.$$

$$(vi) \left\{ \begin{array}{l} P(Y=1) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8} \\ \text{PMF of } Y \end{array} \right.$$

$$(vii) \left\{ \begin{array}{l} P(Y=2) = \frac{1}{8} + \frac{1}{6} = \frac{7}{24} \\ \text{PMF of } Y \end{array} \right.$$

$$(viii) P(Y=1 | X=0) =$$

Using the intuition (Reduced domain)

	$y=0$	$y=1$	$y=2$
$x=0$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$
$x=1$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$

← these are previous prob, they will be redistributed now

↓

$P(Y=1 | X=0) = \frac{6}{13}$

(  $\frac{\frac{1}{4}}{\frac{1}{6} + \frac{1}{4} + \frac{1}{8}} = \frac{1}{4} \cdot \frac{12}{24} = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$  )

↳ redistribution of prob.

it is removed  
bcoz conditional prob is given

Using Formula directly:

$$P(Y=1 | X=0) = \frac{P(Y=1, X=0)}{P(X=0)} = \frac{\frac{1}{4}}{\frac{13}{24}} = \frac{6}{13}$$

(ix) Are  $X$  and  $Y$  independent?

We need to check for all the values there is no other way.

$$\bullet P(X=0, Y=0) = P(X=0) \cdot P(Y=0) \times \text{no need to further check for other values.} \rightarrow X \& Y are dependent$$

$\frac{1}{6} \neq \frac{13}{24} \times \frac{1}{6}$  one counter example is sufficient

$$\bullet P(X=0, Y=1) = P(X=0) \cdot P(Y=1)$$

$$\bullet P(X=0, Y=2) = P(X=0) \cdot P(Y=2)$$

Ques. I toss a coin twice and define  $X$  to be the no. of heads I observe. Then, I toss the coin 2 more times and define  $Y$  to be the no. of heads that I observe this time  $\rightarrow$  last 2 tosses. Find  $P(X \leq 2)$  and  $(Y > 1)$ .

	$X$	$Y$
HH	2	2
HT	1	1
TH	1	1
TT	0	0

for simplicity

$$\begin{aligned} P(X \leq 2, Y > 1) &= P(X \leq 2) \cdot P(Y > 1) \quad \text{bcoz} \\ &= \frac{3}{4} \cdot \frac{1}{4} \\ &= \frac{3}{16} \end{aligned}$$

Actually it would be like this

Sequence	$X$	$Y$
HHHH	2	2
HHHT	2	1
HHTH	2	1
HHTT	2	0
HTHH	1	2
HTHT	1	1
HTTH	1	1
HTTT	0	0

$X$  &  $Y$  are independent.  
bcoz what happens in 1<sup>st</sup> 2 tosses  
doesn't effect what happens in 2<sup>nd</sup> 2 tosses.  
Both are independent tosses.

Ques. Given the joint prob. distribution.

(a) Find distribution of  $X$  and  $Y$ .

		$X$			
		1	2	3	4
1		$\frac{1}{16}$	0	$\frac{1}{8}$	$\frac{1}{16}$
2		$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{4}$	0
3		0	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
4		$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{16}$	$\frac{1}{32}$

$\left. \begin{array}{l} \text{PMF of } Y \\ \text{distribution of } X \end{array} \right\} P(Y=1) = P(Y=1, X=1) + P(Y=1, X=2) + P(Y=1, X=3) + P(Y=1, X=4)$

$$= \frac{1}{16} + 0 + \frac{1}{8} + \frac{1}{16} = \frac{1}{4}$$

$$P(Y=2) = \frac{1}{32} + \frac{1}{32} + \frac{1}{4} = \frac{5}{16}$$

$$P(Y=3) = \frac{1}{8} + \frac{1}{16} + \frac{1}{16} = \frac{1}{4}$$

$$P(Y=4) = \frac{1}{16} + \frac{1}{32} + \frac{1}{16} + \frac{1}{32} = \frac{3}{16}$$

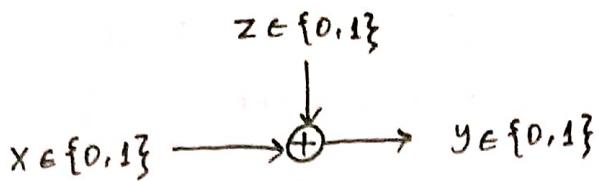
If we can find for  $X$ :

$P(X=1)$	$= \frac{1}{16} + \frac{1}{32} + 0 + \frac{1}{16} = \frac{5}{32}$
$P(X=2)$	$= 0 + \frac{1}{32} + \frac{1}{8} + \frac{1}{32} = \frac{3}{16}$
$P(X=3)$	$= \frac{1}{8} + \frac{1}{4} + \frac{1}{16} + \frac{1}{16} = \frac{1}{2}$
$P(X=4)$	$= \frac{1}{16} + 0 + \frac{1}{16} + \frac{1}{32} = \frac{5}{32}$

(b) Find  $P(Y=1 | X=1)$

$$= \frac{P(Y=1, X=1)}{P(X=1)} = \frac{\frac{1}{16}}{\frac{5}{32}} = \frac{2}{5}$$

Ques. consider the following binary channel



$$P(x=0) = p$$

$$P(z=0) = q$$

$y = (x+z) \bmod 2 = x \oplus z$ , and  $x$  and  $z$  are independent. Find  $P(y=0)$ .

$$P(x=0) = p \quad y = x \oplus z, \quad P(y=0) = P(x \oplus z = 0)$$

$$P(z=0) = q \quad = P(x=0, z=0) + P(x=1, z=1) \quad \text{ } \begin{matrix} x \oplus z \text{ are} \\ \text{independent} \end{matrix}$$

$$= pq + (1-p)(1-q)$$

or using the table:

x	z	y
0	0	0
0	1	1
1	0	1
1	1	0

$P(y=0) = pq + (1-p)(1-q)$

$$P(y=0|x=0) = P(x \oplus z = 0 | x=0), \text{ putting } x=0 \text{ in } x \oplus z = 0$$

$$= P(0 \oplus z = 0) = P(z=0) = q$$

$$P(y=1|x=0) = P(x \oplus z = 1 | x=0) = P(0 \oplus z = 1) = P(z=1) = 1-q$$

$$P(y=0|x=1) = P(x \oplus z = 0 | x=1) = P(1 \oplus z = 0) = P(z=1) = 1-q$$

$$P(y=1|x=1) = P(x \oplus z = 1 | x=1) = P(1 \oplus z = 1) = P(z=0) = q$$

GATE 17

Ques.

Suppose  $X_i$  for  $i=1, 2, 3$  are independent & identically distributed RV whose prob PMF are  $\Pr[X_i=0] = \Pr[X_i=1] = \frac{1}{2}$  for  $i=1, 2, 3$ .

Define another RV,  $y = X_1 X_2 \oplus X_3$ , where  $\oplus$  denotes XOR.

Then,  $\Pr[y=0 | X_3=0] = ?$

$$\hookrightarrow y = X_1 X_2 \oplus X_3 \quad \Pr[X_i=1] = \Pr[X_i=0] = \frac{1}{2}$$

**Method 1**  $\rightarrow \Pr(y=0 | X_3=0)$

$$= \Pr(X_1 X_2 \oplus X_3 = 0 | X_3=0) \xrightarrow{\text{OR}} \Pr(X_1 X_2 \oplus 0 = 0) = \Pr(X_1 X_2 = 0)$$

$$= \Pr(X_1 X_2 = 0 | X_3=0) = \Pr(X_1 X_2 = 0)$$

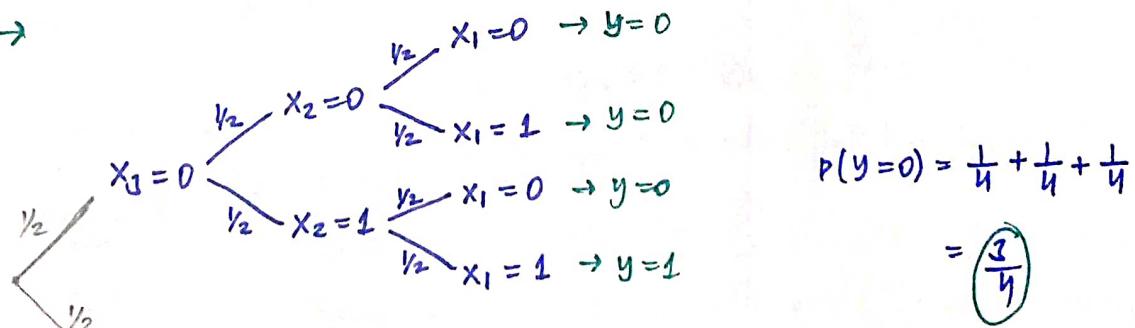
$$\begin{aligned} \text{Now, } \Pr(X_1 X_2 = 0) &= \Pr(X_1=0, X_2=0) + \Pr(X_1=0, X_2=1) + \Pr(X_1=1, X_2=0) \\ &= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \quad \leftarrow \text{bcoz } X_1, X_2 \text{ are indep.} \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} \end{aligned}$$

**Method 2**  $\rightarrow$  Reduced Domain

$X_1$	$X_2$	$X_3$	$y$	initial prob.	prob redistributed
0	0	0	0	$\frac{1}{8}$	$\frac{1}{4}$
0	0	1	1	$\frac{1}{8}$	$\frac{1}{4}$
0	1	0	0	$\frac{1}{8}$	$\frac{1}{4}$
0	1	1	1	$\frac{1}{8}$	$\frac{1}{4}$
1	0	0	0	$\frac{1}{8}$	$\frac{1}{4}$
1	0	1	1	$\frac{1}{8}$	$\frac{1}{4}$
1	1	0	1	$\frac{1}{8}$	$\frac{1}{4}$
1	1	1	0		

$\therefore \Pr(y=0 | X_3=0) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$

**Method 3**  $\rightarrow$



whatever written with pencil is unnecessary and has been removed.

**Method 4**

$$P(y=0 | X_3=0) = \frac{P(y=0, X_3=0)}{P(X_3=0)}$$

$$= \frac{P(X_1 X_2 \oplus X_3 = 0, X_3=0)}{P(X_3=0)} =$$

$$= \frac{P(X_1 X_2 = 0, X_3=0)}{P(X_3=0)}$$

$$= \frac{P(X_1 X_2 = 0) \cdot P(X_3=0)}{P(X_3=0)}$$

$$= P(X_1 X_2 = 0) = \frac{3}{4}$$

**Method 5**

$X_1$	$X_2$	$X_3$	$y$
0	0	0	0 $\rightarrow \frac{1}{8}$
0	0	1	1 $\rightarrow \frac{1}{8}$
0	1	0	0 $\rightarrow \frac{1}{8}$
0	1	1	1 $\rightarrow \frac{1}{8}$
1	0	0	0 $\rightarrow \frac{1}{8}$
1	0	1	1 $\rightarrow \frac{1}{8}$
1	1	0	1 $\rightarrow \frac{1}{8}$
1	1	1	0 $\rightarrow \frac{1}{8}$

$$P(y=0 | X_3=0)$$

$$= \frac{P(y=0, X_3=0)}{P(X_3=0)}$$

$$= \frac{\frac{1}{8} + \frac{1}{8} + \frac{1}{8}}{\frac{1}{2}} = \frac{3}{4}$$

# TYPES OF RANDOM VARIABLES

- Discrete RV → takes discrete values.

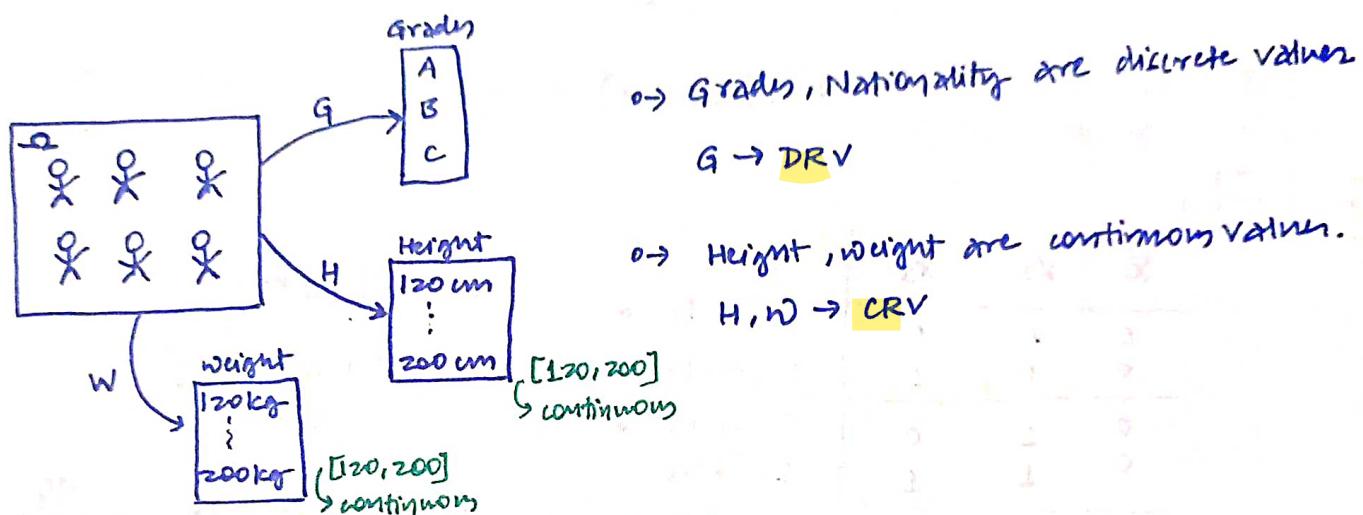
A RV is called discrete if takes either finite or countably infinite no. of values.

- (ex) The no. of sides in the two rolls  $\rightarrow \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix}$

- Continuous RV → takes continuous values.

A RV that can take continuous and uncountably infinite no. of values.

- (ex) consider an experiment of choosing a point from the interval  $[-1, 1]$ .



→ When you think of a RV, immediately ask yourself

- what are the possible values?
- what are their probabilities?

## (Intuitive) question

Write down the type of RV in the following table.

<u>RV description</u>	<u>Range</u>	<u>DRV or CRV ?</u>
(i) $X$ , the no. of heads in $n$ flips of a fair coin.	$\{0, 1, \dots, n\}$	DRV
(ii) $N$ , the no. of people born this year.	$\{0, 1, 2, \dots\}$	DRV
(iii) $F$ , the no. of flips of a fair coin upto and including my first head.	$\{1, 2, 3, \dots\}$ <small>H TH THH TTT...H</small>	DRV
(iv) $B$ , the amount of time I wait for the next bus in seconds.	$[0, \infty)$ <small>there could be partial seconds waited.</small>	CRV
(v) $C$ , the temp. in celsius of liquid water.	$(0, 100)$ <small>the temp can be any real no. in this range</small>	CRV

# DISCRETE RANDOM VARIABLE

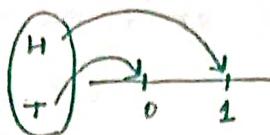
## PROBABILITY MASS FUNCTION (PMF)

listing down prob. of each value for a DRV is called PMF.

Expt : Flipping the coin once.

$X = \text{no. of heads in a flip.}$

	$x=0$	$P(\cdot)$
T		$\frac{1}{2}$
H	$x=1$	$\frac{1}{2}$



Ques. Suppose that a coin is tossed twice. Let  $X$  represent the no. of heads that can come up. Write down PMF of  $X$ .

Ans :  $X : 0, 1, 2$

	$x$	$P(\cdot)$
TT	0	$\frac{1}{4}$
HT, TH $\rightarrow$	1	$\frac{2}{4} = \frac{1}{2}$
HH	2	$\frac{1}{4}$

$\left. \begin{array}{l} \text{PMF of } X \\ \text{every prob. should be } \geq 0 \text{ and } \leq 1 \\ \text{sum of all prob. should be 1.} \end{array} \right\}$

$x$	
HH	2
HT	1
TH	1
TT	0

Ques 2. Check whether the func' given by

$$P(X=x) = \frac{x+2}{25} \text{ for } x=1, 2, 3, 4, 5 \text{ is valid PMF of RV } X.$$

every prob. should be  $\geq 0$  and  $\leq 1$

sum of all prob. should be 1.

$$\hookrightarrow P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$$

$$= \frac{3}{25} + \frac{4}{25} + \frac{5}{25} + \frac{6}{25} + \frac{7}{25}$$

$$= \frac{25}{25} = 1 \Leftrightarrow$$

Ques 3. Suppose that a pair of fair dice are to be tossed, & let the RV  $X$  denote the sum of the points. Obtain the Prob distribution of  $X$ .

<u>PMF for <math>X</math></u>	$x$	2	3	4	5	6	7	8	9	10	11	12
$P(\cdot)$		$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

NOTE → calculation of the PMF of a RV  $X$ .

For each possible value  $x$  of  $X$  :

1. collect all the possible outcome that give rise to the event  $\{X=x\}$ .
2. Add their probabilities to obtain  $P(X=x)$ .

Ques 4. The RV  $X$  has prob. funcn

$$P(X=x) = \begin{cases} kx & , x=1,2 \\ k(x-2) & , x=3,4,5 \end{cases}$$

where  $k$  is constant.

Find the value of  $k$ .

$$\hookrightarrow P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5) = 1$$

$$k + 2k + k + 2k + 3k = 1$$

$$9k = 1$$

$$\Rightarrow \boxed{k = \frac{1}{9}}$$

Ques 5. Suppose that  $B, C$  are DRV each taking values  $1, 2, 3$ , with joint PMF  $P(B, C)$  given by the following table.

Note that one of the entries in the table is missing.

$B \setminus C$	1	2	3
1	$\frac{1}{20}$	$\frac{2}{20}$	0
2	$\frac{2}{20}$	$\frac{5}{20}$	$\frac{1}{20}$
3	$\frac{3}{20}$	$\frac{2}{20}$	K → let it be 'K'

Determine the value of the missing entry  $P(3,3)$

$$\hookrightarrow P(C=1) + P(C=2) + P(C=3) = 1$$

$$\left(\frac{1}{20} + \frac{2}{20} + \frac{3}{20}\right) + \left(\frac{2}{20} + \frac{5}{20} + \frac{2}{20}\right) + \left(0 + \frac{1}{20} + K\right) = 1$$

$$\frac{6}{20} + \frac{9}{20} + \frac{1}{20} + K = 1$$

$$K = 1 - \frac{16}{20} = \frac{4}{20}$$

$$\boxed{K = \frac{4}{20}}$$

Ques 6 Let  $X$  be a DRV taking the values  $\{1, 2, \dots, n\}$  all with equal prob.

Let  $Y$  be another DRV taking values in  $\{1, 2, \dots, n\}$ .

Assume that  $X, Y$  are independent. Find  $P(X=Y)$ .

$$X = 1, 2, \dots, n$$

$$Y = 1, 2, \dots, n$$

nothing is said about prob of  $y$

$$P(X=Y) = P(X=Y=1) + P(X=Y=2) + P(X=Y=3) + \dots + P(X=Y=n)$$

Tricky

$$= P(X=1, Y=1) + P(X=2, Y=2) + \dots + P(X=n, Y=n)$$

$$= P(X=1) \cdot P(Y=1) + P(X=2) \cdot P(Y=2) + \dots + P(X=n) \cdot P(Y=n)$$

$$= \frac{1}{n} P(Y=1) + \frac{1}{n} P(Y=2) + \dots + \frac{1}{n} P(Y=n)$$

$$= \frac{1}{n} [P(Y=1) + P(Y=2) + \dots + P(Y=n)] = \frac{1}{n} \times 1 = \boxed{\frac{1}{n}}$$

(Q) simplify

$$P(X=Y) = \sum_{k=1}^n P(X=k, Y=k) = \sum_{k=1}^n P(X=k) P(Y=k)$$

$$= \frac{1}{n} \sum_{k=1}^n P(Y=k) = \frac{1}{n}$$

(Q)

$$P(X=Y) = \sum_{k=1}^n P(X=Y | Y=k) P(Y=k)$$

$$= \sum_{k=1}^n \frac{P(X=Y, Y=k)}{P(Y=k)} P(Y=k)$$

$$= \sum_{k=1}^n P(X=Y, Y=k) = \sum_{k=1}^n P(X=Y) P(Y=k) = \frac{1}{n} \sum_{k=1}^n P(Y=k) = \frac{1}{n}$$

Ques 7 Let  $X$  and  $Y$  be DRV with  $\text{Range}(X) = \{0, 1, 2\}$  and  $\text{Range}(Y) = \{1, 2\}$  with joint distribution given by the chart below.

	$X=0$	$X=1$	$X=2$
$Y=1$	0.1	0.2	0.1
$Y=2$	0.3	0.2	0.1

What is the value of  $P(X=1 | Y=2)$ ?

$$P(X=1 | Y=2) = \frac{P(X=1, Y=2)}{P(Y=2)} = \frac{0.2}{0.3 + 0.2 + 0.1} = \frac{0.2}{0.6} = 0.333.$$

Ques 8 A DRV,  $Y$ , has probability mass func'

$$P(Y=y) = c(y-3)^2, \quad y = -2, -1, 0, 1, 2.$$

Find the value of the constant  $c$ .

Writing the PMF in table we have

$y$	-2	-1	0	1	2
$P(Y=y)$	$25c$	$16c$	$9c$	$4c$	$c$

The sum of PMF is 1.

$$\Rightarrow \sum_{y=-2}^2 P(Y=y) = 55c = 1$$

$$\Rightarrow \boxed{c = \frac{1}{55}}$$

## PMF of a RV which is func<sup>n</sup> of RV

Ques 9 Suppose we roll a dice and let  $X$  be the value that is showing. Find PMF of  $X^2$ ?

↪ PMF of  $X$

$X$	$P(\cdot)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

PMF of  $X^2$

possible values of $X^2$	$X^2$	$P(\cdot)$
$x=1 \rightarrow 1$	1	$\frac{1}{6}$
$x=2 \rightarrow 4$	4	$\frac{1}{6}$
$x=3 \rightarrow 9$	9	$\frac{1}{6}$
$x=4 \rightarrow 16$	16	$\frac{1}{6}$
$x=5 \rightarrow 25$	25	$\frac{1}{6}$
$x=6 \rightarrow 36$	36	$\frac{1}{6}$

$$P(X^2=1) = P(X=1) = \frac{1}{6}$$

$$P(X^2=4) = P(X=2) = \frac{1}{6}$$

Ques 10 Suppose  $X$  has PMF  $P(X=k) = \frac{k}{10}$  for  $k = 1, 2, 3, 4$ .

Find the PMF of  $X^2 - 4X + 3$ .

PMF of  $X$

$X$	$P(\cdot)$
1	$\frac{1}{10}$
2	$\frac{2}{10}$
3	$\frac{3}{10}$
4	$\frac{4}{10}$

$$y = X^2 - 4X + 3$$

$$x=1 \Rightarrow y=0$$

$$x=2 \Rightarrow y=-1$$

$$x=3 \Rightarrow y=0$$

$$x=4 \Rightarrow y=3$$

PMF of  $y$

$y$	$P(\cdot)$
0	$\frac{1}{10}$
-1	$\frac{2}{10}$
3	$\frac{4}{10}$

$$P(y=0) = P(x=1) + P(x=3)$$

$$= \frac{1}{10} + \frac{3}{10} = \frac{4}{10}$$

Ques 11 PMF of  $X$  is given below:

$$P(X=-2) = \frac{1}{8}, P(X=-1) = \frac{1}{4}, P(X=1) = \frac{3}{8}, P(X=2) = \frac{1}{4}$$

$$\text{let } y = X^2.$$

Find out PMF of  $y$ .

PMF of  $X$

$X$	$P(\cdot)$
-2	$\frac{1}{8}$
-1	$\frac{1}{4}$
1	$\frac{3}{8}$
2	$\frac{1}{4}$

$$y = X^2$$

$$x=-2 \Rightarrow y=4$$

$$x=2 \Rightarrow y=4$$

$$x=-1 \Rightarrow 1$$

$$x=1 \Rightarrow 1$$

PMF of  $y$

$y$	$P(\cdot)$
1	$\frac{1}{4} + \frac{3}{8} = \frac{5}{8}$
4	$\frac{1}{8} + \frac{1}{4} = \frac{3}{8}$

Ques 12  $P(X=x) = \frac{x^2}{a}$ ,  $x \in \{-3, -2, -1, 1, 2, 3\}$

(a) Find  $a$ .

The sum of PMF is 1

$$\Rightarrow P(X=-3) + P(X=-2) + P(X=-1) + P(X=1) + P(X=2) + P(X=3) = 1$$

$$\frac{9}{a} + \frac{4}{a} + \frac{1}{a} + \frac{1}{a} + \frac{4}{a} + \frac{9}{a} = 1$$

$$\frac{28}{a} = 1 \Rightarrow a = 28$$

(b) Find PMF of  $Z = X^2$

$Z = X^2$	$P(\cdot)$
9	$\frac{18}{28} [P(X=3) + P(X=-3)]$
4	$\frac{8}{28} [P(X=2) + P(X=-2)]$
1	$\frac{2}{28} [P(X=1) + P(X=-1)]$

$$P(Z=X^2) = \begin{cases} \frac{18}{28}, & Z=9. \\ \frac{8}{28}, & Z=4. \\ \frac{2}{28}, & Z=1. \end{cases}$$

## EXPECTATION OF DRV

- A single no to summarize PMF.
- Weighted average (in proportion to probabilities) of the possible values of  $x$ .

→ The PMF of a RV  $X$  provides us with several nos., the probabilities of all the possible values of  $x$ . It would be desirable to summarize this info in a single representative no. This is accomplished by the expectation of  $x$ , which is a weighted (in proportion to probabilities) avg of the possible values of  $x$ .

Expectation of DRV :  $E(x) = \sum k \cdot P(x=k)$

Ques. If we toss a coin and  $X$  is 1 if we have heads and 0 if we have tails, what is the expectation of  $X$ ?

↳  $x = \begin{cases} 1 & H \\ 0 & T \end{cases}$   
 $E[x] = ?$

$x$	$P(\cdot)$
1	$\frac{1}{2}$
0	$\frac{1}{2}$

$$E[x] = \sum k \cdot P(x=k)$$

$$= 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \boxed{\frac{1}{2}}$$

$$E[x] = 1 \cdot P(x=1) + 0 \cdot P(x=0)$$

Ques. A fair blue die has faces numbered 1, 1, 2, 3, 5 and 5. The RV  $B$  represents the score when the die is rolled. Find  $E(B)$ .

↳

$B$	$P(\cdot)$
1	$\frac{2}{6}$
2	$\frac{1}{6}$
3	$\frac{2}{6}$
5	$\frac{2}{6}$

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

$$= \frac{2}{6} (1+3+5) = \frac{2}{6} \cdot 9 = \boxed{3}$$

↳  $P(B = E[B]) = \frac{2}{6}$  } but it is not useful.  
we may or may not get the prob of expected value

Ques. In a game a fair die is rolled, and the amount of money won and lost is determined by the no. rolled. The amount are given in the table below:

Number	1	2	3	4	5	6
winnings	-8	3	1	-8	1	5

Find the expected amount of money won or lost  $\rightarrow E[x]$

↳  $x = \text{amount of money won or lost}$

	$x$	$P(\cdot)$
1	-8	$\frac{1}{6}$
2	3	$\frac{1}{6}$
3	1	$\frac{1}{6}$
4	-8	$\frac{1}{6}$
5	1	$\frac{1}{6}$
6	5	$\frac{1}{6}$

$$E[x] = \frac{1}{6}(-8) + \frac{1}{6}(3) + \frac{1}{6}(1) + \frac{1}{6}(-8) + \frac{1}{6}(1) + \frac{1}{6}(5)$$

$$= \frac{1}{6}(-8+3+1-8+1+5)$$

$$= -2$$

Ques. Let  $S$  and  $B$  are 2 RV. Their joint prob distribution is given below:

	$S=0$	$S=1$	$S=2$	
$B=0$	0.1	0.1	0	0.2 $\rightarrow P(B=0) = 0.2$
$B=1$	0.1	0.2	0	0.3
$B=2$	0	0	0.5	0.5 Find the $E[B]$ ?

PMF for  $B$ :

$B$	$P(\cdot)$
0	0.2
1	0.3
2	0.5

$$\begin{aligned} E(B) &= 0 \times 0.2 + 1 \times 0.3 + 2 \times 0.5 \\ &= 0.3 + 1 \\ &= (1.3) \end{aligned}$$

Ques. A biased die with 6 faces is rolled. The DRV  $X$  represents the score on the uppermost face. The prob. distribution of  $X$  is shown in the table below:

$x$	1	2	3	4	5	6
$P(X=x)$	a	a	a	b	b	0.3

Given that  $E(X) = 4.2$  Find the value of  $a$  and  $b$ .

↳ Sum of all probabilities = 1

$$\Rightarrow 3a + 2b + 0.3 = 1$$

$$3a + 2b = 0.7 \quad \text{--- (1)}$$

$$E(X) = 4.2$$

$$\Rightarrow 1 \cdot a + 2 \cdot a + 3 \cdot a + 4 \cdot b + 5 \cdot b + 6 \cdot 0.3 = 4.2$$

$$6a + 9b = 4.2 - 1.8$$

$$6a + 9b = 2.4 \quad \text{--- (2)}$$

$$6a + 9b = 2.4$$

$$\begin{array}{r} -6a + 4b = 2.4 \\ \hline 5b = 1 \end{array} \Rightarrow b = 1/5 = 0.20 \Rightarrow a = \frac{0.7 - 0.4}{3} = 0.1$$

$$a = 0.1, b = 0.2$$

Biostatistics

Statistical Methods

## Revisiting prob definition

prob of head =  $\frac{1}{2}$ , then what does this mean?

Do we mean that if we toss 2 times then we'll definitely be getting one head?

2 times - 1 Head NO

→ it means,

4 times - 2 Heads NO

if you toss enough no. of times then  
you'll be getting half of times head.

8 times - 4 Heads NO

2000 times toss  $\approx$  996 heads. {A mathematician did it}

## SAMPLE probabilities v/s Real probabilities

Let's suppose we have a biased coin,

$$P(H) = \frac{1}{3} \quad \left. \begin{array}{l} \text{underlying prob. distribution} \\ \text{Real Probabilities} \end{array} \right\}$$

$$P(T) = \frac{2}{3} \quad \left. \begin{array}{l} \text{(fix, Absolute, no one can change it)} \\ \text{we collected some samples.} \end{array} \right\}$$

Now, we start tossing the coin and we get: (This process is called sampling the data)

$$H, H, T, T, T, H, T, T, T, H, H, T \quad \rightsquigarrow P(H) = \frac{5}{12} \quad P(T) = \frac{7}{12} \quad \left. \begin{array}{l} \text{sample distribution} \end{array} \right\}$$

$$H, H, T, T, T, T, H, T, T, T, H, H, T \quad \rightsquigarrow P(H) = \frac{6}{15} \quad P(T) = \frac{9}{15} \quad \left. \begin{array}{l} \text{sample distribution} \end{array} \right\}$$

→ The sample distribution will keep on changing

but if we collect enough samples, we'll keep getting closer to the real underlying distribution

and this is exactly what we desire, to approach the real probabilities.

underlying distribution :  $P(H) = \frac{1}{3}$   
(no body knows)  $P(T) = \frac{2}{3}$

## Sample distribution

(calculated by sampling the data)  
i.e. by doing the experiment and then calculating

tossed the coin 9000 times.  $\begin{cases} \rightarrow 2990 \text{ Heads} \\ \rightarrow 6010 \text{ Tails} \end{cases} \quad P(H) = \frac{2990}{9000} \approx \text{Real prob.}$

→ Why most companies are hungry for more and more data?

↪ You have data → calculate (sample) prob.

with more & more data

Sample prob  $\rightsquigarrow$  Real prob.

- in Real world, no one knows underlying distribution.
- we can just collect more & more data
- we calculate probabilities using data.
- if you have more data, you get better probabilities.

## EXPECTATION V/S AVERAGE

What does expectation really mean? How Expectation is different from Average?

Suppose you are playing a soccer game & the #goals you scored in successive matches are:

$$\rightarrow 1, 3, 6, 1, 3, 6, 5, 1, 5, 1, 3, 6$$

$$\text{Avg of all goals} = \frac{1+3+6+1+3+6+5+1+5+1+3+6}{12} \quad \text{rewriting}$$

$$= 1\left(\frac{4}{12}\right) + 3\left(\frac{3}{12}\right) + 6\left(\frac{3}{12}\right) + 5\left(\frac{2}{12}\right)$$

if we take closer look, these are nothing but probabilities  
not real probabilities, but sample probabilities.

this is nothing but  
prob of scoring 5 goals in a  
match.

is it real prob? NO  
it is sample prob.

So, the Avg is nothing but, expectation, calculated with sample probabilities.

$$\rightarrow 1, 3, 6, 1, 3, 6, 5, 1, 5, 1, 3, 6, 5, 6, 1$$

$$\text{Avg of all goals} = \frac{1+3+6+1+3+6+5+1+5+1+3+6+5+6+1}{15} \quad \text{rewriting}$$

$$= 1\left(\frac{1}{15}\right) + 3\left(\frac{3}{15}\right) + 6\left(\frac{4}{15}\right) + 5\left(\frac{3}{15}\right)$$

Again this is avg, but if you look at it in rewritten form, it is  
expectation, calculated with sample probabilities.

$$\rightarrow \text{for 12 sample : } 1\left(\frac{4}{12}\right) + 3\left(\frac{3}{12}\right) + 6\left(\frac{3}{12}\right) + 5\left(\frac{2}{12}\right) \rightarrow \underline{\underline{\text{Avg}}}$$

$$\text{15 sample : } 1\left(\frac{1}{15}\right) + 3\left(\frac{3}{15}\right) + 6\left(\frac{4}{15}\right) + 5\left(\frac{3}{15}\right) \rightarrow \underline{\underline{\text{Avg}}}$$

if you collect more and more samples  
(the sample probabilities will tend to real probabilities)

$$1 \cdot P(1) + 3 \cdot P(3) + 6 \cdot P(6) + 5 \cdot P(5) \Rightarrow \underline{\underline{\text{Expectation}}}$$

Sample probabilities  $\xrightarrow{\text{more more data}}$  Real probabilities      Avg  $\xrightarrow{\text{enough data}}$  Expectation

$\rightarrow$  Expectation is just Average on enough data.

but for our purpose, we just say that Expectation is just Avg.

$\uparrow$   
to avoid getting  
into details as  
we assume we have enough data.

(ex) If we roll a die,  $X$  = Number appears on top

$$E[X] = ?$$

$$X = 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6$$

$$P(\cdot) = \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6}$$

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

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

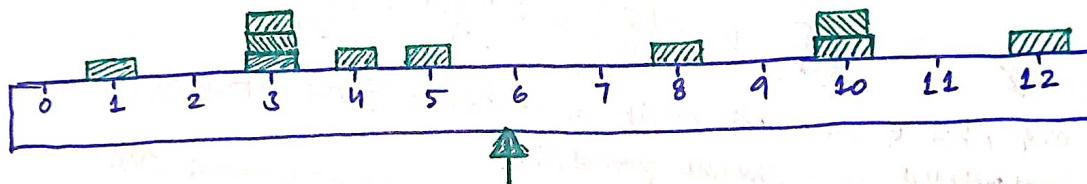
This 3.5 intuitively means that

if you roll a die, on avg you'll get 3.5.  
ie if you roll a die and take avg of all the outputs  
it will come out to be 3.5.

if you roll a die enough no. of times,  
the avg of all the outputs will be 3.5.

- you won't even get 3.5 ever, you'll just get integers {1, 2, ..., 6}  
but if you take avg of all, then you'll get 3.5.

### → EXPECTATION as CENTRE OF MASS (Physics)



$$P(1) = \frac{1}{10} \quad P(3) = \frac{3}{10} \quad P(4) = \frac{1}{10} \quad P(5) = \frac{1}{10} \quad P(7) = \frac{1}{10} \quad P(8) = \frac{2}{10} \quad P(10) = \frac{1}{10} \quad P(12) = \frac{1}{10}$$

$$E[X] = 1 \times \frac{1}{10} + 3 \times \frac{3}{10} + 4 \times \frac{1}{10} + 5 \times \frac{1}{10} + 7 \times \frac{1}{10} + 8 \times \frac{2}{10} + 10 \times \frac{1}{10} + 12 \times \frac{1}{10}$$

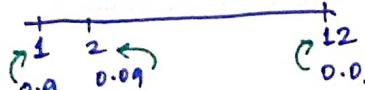
$$= \frac{1+3+4+5+7+8+10+12}{10} = \frac{59}{10} = 5.9$$

ques. T/F if  $a \leq X \leq b$  then  $a \leq E[X] \leq b$ .

**TRUE**

expectation is nothing but avg.

Avg will always be somewhere b/w the Range of nos  
& never outside that Range

(ex)  → intuitively, expectation will be closer to 1, bcoz 1 is very heavy.

$$E[X] = 1 \times 0.9 + 2 \times 0.09 + 12 \times 0.01 \quad \text{calculated.}$$

$$= 1.20$$

$X$	$P(\cdot)$
1	0.9
2	0.09
12	0.01

very close to 1.  
bcoz  $P(1)$  is very high.

### MIT Book Question

Consider a quiz where a person is given 2 questions & must decide which question to answer 1st. Ques 1 will be answered correctly with a prob. of 0.8 and the person will then receive a prize \$100, while ques. 2 will be answered correctly with prob 0.5, & the person will then receive a prize \$200.

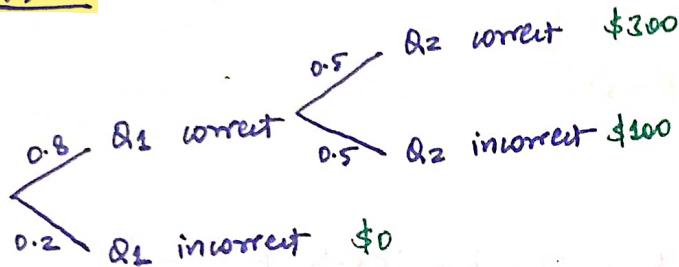
If the 1st question attempted is answered incorrectly, the quiz terminates, i.e., the person is not allowed to attempt the 2nd ques. If the 1st ques is answered correctly, the person is allowed to attempt the 2nd ques.

which ques to be answered 1st to maximize the expected value of the total prize received?

$$\begin{array}{ll} Q_1 : & 0.8 \quad 100 \\ & 0.2 \quad 0 \end{array}$$

$$Q_2 : \quad 0.5 \quad 200$$

Case 1:



• Q1 answered 1st

X: Amount won.

PMF of X:

X	P(X)
0	0.2
100	$0.8 \times 0.5 = 0.4$
300	$0.8 \times 0.5 = 0.4$

$$\therefore E[X] = 0 \times 0.2 + 100 \times 0.4 + 300 \times 0.4 \\ = \$160.$$

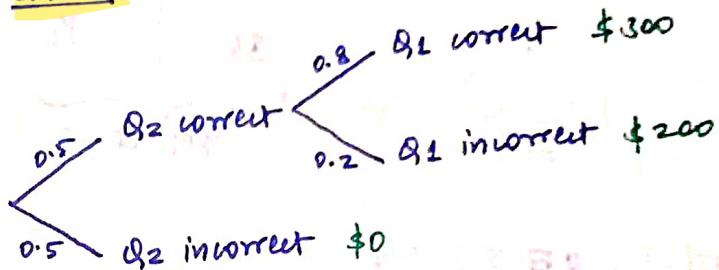
∴ if we start with ques 1, expected money won = \$160.

∴ It is better to start with Q1. ← Answer

\* A very similar question is asked by GATE in 2021.  
take a look at it also.

Tricky Hai Boss... Think  
se padho

case 2:



• Q2 answered 1st

PMF of X:

X.	P(X)
0	0.5
200	$0.5 \times 0.2 = 0.1$
300	$0.5 \times 0.8 = 0.4$

$$\begin{aligned} E[X] &= 0 \times 0.5 + 200 \times 0.1 + 300 \times 0.4 \\ &= \$140. \end{aligned}$$

if we start with ques 2,  
expected money won = \$140.

## Expectation of RV which is a func<sup>n</sup> of RV

Ques. Suppose we roll a die & let  $X$  be the value that is showing. We want to find the expectation  $E[X^2]$ ?

$x$	$P(X=k)$	$x^2$	$P(X^2=k)$
1	$\frac{1}{6}$	1	$\frac{1}{6}$
2	$\frac{1}{6}$	4	$\frac{1}{6}$
3	$\frac{1}{6}$	9	$\frac{1}{6}$
4	$\frac{1}{6}$	16	$\frac{1}{6}$
5	$\frac{1}{6}$	25	$\frac{1}{6}$
6	$\frac{1}{6}$	36	$\frac{1}{6}$

PMF of  $X$

PMF of  $X^2$

$$E[y] = \sum y \cdot P(y)$$

$$E[X^2] = \sum k P(X^2=k) = 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6} + 16 \times \frac{1}{6} + 25 \times \frac{1}{6} + 36 \times \frac{1}{6}$$

$$E[X^2] = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$$

Method 2 →

$$E[X^2] = \sum_k x^2 \cdot P(X=k)$$

$$\Rightarrow E[X^2] = \sum_k x^2 \cdot P(X=k)$$

$$= 1^2 \cdot P(X=1) + 2^2 \cdot P(X=2) + 3^2 \cdot P(X=3)$$

$$4^2 \cdot P(X=4) + 5^2 \cdot P(X=5) + 6^2 \cdot P(X=6)$$

$$= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6}$$

$$= \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$$

$$E[y] = \sum_k y \cdot P(X=k)$$

$$E[\sin x] = \sum_k \sin x \cdot P(X=k)$$

$$E[X^2 + 3X + 2] = \sum_k (x^2 + 3x + 2) \cdot P(X=k)$$

$$E[g(x)] = \sum_k g(x) \cdot P(X=k)$$

this stays the same  
no need to modify for  
every func<sup>n</sup>.

i.e. we don't need to find the  
PMF of  $g(x)$  to plug in the values  
in calculating  $E[g(x)]$ .

Ques.

x	P(x)
-2	1/5
-1	1/5
0	1/5
1	1/5
2	1/5

$$E[x^2] = ?$$

Method 1 → possible values of  $x^2 = 0, 1, 4$

finding PMF of  $x^2$

$x^2$	P(x)
0	1/5
1	2/5
4	2/5

$$E[x^2] = 0 \cdot \frac{1}{5} + 1 \cdot \frac{2}{5} + 4 \cdot \frac{2}{5}$$
$$= (2)$$

Method 2 →  $E[x^2] = \sum_k x^2 P(x=k)$

$$= (-2)^2 \cdot P(x=-2) + (-1)^2 \cdot P(x=-1) + 0^2 \cdot P(x=0)$$
$$+ (1)^2 \cdot P(x=1) + 2^2 \cdot P(x=2)$$

$$= 4 \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} + 0 \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} + 4 \cdot \frac{1}{5} = (2)$$

NOTE → Expected value Rule for Func's of RV

let  $X$  be a RV with PMF  $P_X(x)$ , and  $g(x)$  be a real valued func' of  $X$ .  
Then, the expected value of the RV  $g(x)$  is given by

$$E[g(x)] = \sum_x g(x) P_X(x)$$

Ques. PMF of  $X$  is given below:

$$P(X=-2) = \frac{1}{8} \quad P(X=-1) = \frac{1}{4} \quad P(X=1) = \frac{3}{8} \quad P(X=2) = \frac{1}{4}.$$

let  $y = x^2$ . Find out expectation of  $y$ .

$$\hookrightarrow E[x^2] = \sum_k x^2 P(x=k)$$

$$= (-2)^2 \cdot \frac{1}{8} + (-1)^2 \cdot \frac{1}{4} + (1)^2 \cdot \frac{3}{8} + (2)^2 \cdot \frac{1}{4}$$
$$= 4 \cdot \frac{1}{8} + \frac{1}{4} + \frac{3}{8} + 4 \cdot \frac{1}{4} = \left(\frac{17}{8}\right)$$

Ques. Let  $X$  be a DRV with  $P_X(k) = \frac{1}{5}$  for  $k = -1, 0, 1, 2, 3$ . Let  $y = 2|X|$ . Find the range and PMF of  $y$ .

$$y = 2|X| \Rightarrow \text{Range of } y : \\ \{0, 2, 4, 6\}$$

PMF of  $y$

	$y$	$P(\cdot)$
$x = -1$	0	$\frac{1}{5}$
$x = 2$	2	$\frac{1}{5} + \frac{1}{5} = \frac{2}{5}$
	4	$\frac{1}{5}$
	6	$\frac{1}{5}$

$$\begin{aligned} x = -1 &\Rightarrow y = 2|-1| = 2 \\ x = 0 &\Rightarrow y = 2|0| = 0 \\ x = 1 &\Rightarrow y = 2|1| = 2 \\ x = 2 &\Rightarrow y = 2|2| = 4 \\ x = 3 &\Rightarrow y = 2|3| = 6 \end{aligned}$$

## EXPECTATION Tricky Questions

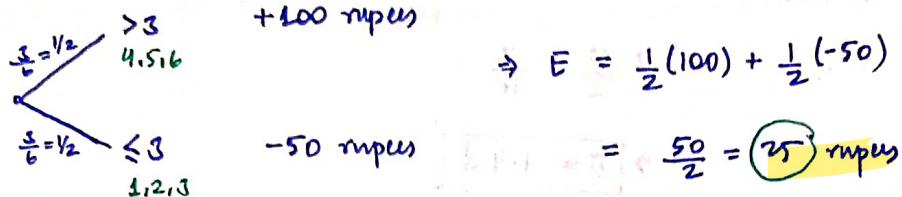
### LAW OF TOTAL EXPECTATION :

$$E[X] = E_1 P(A) + E_2 P(A^c)$$

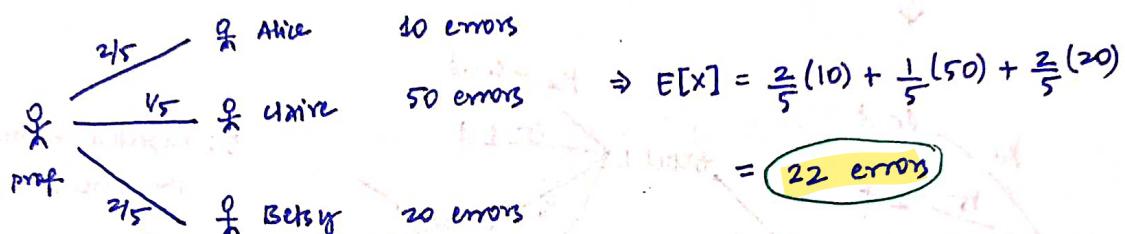
We can break the expectation into 2 parts just like 1 and 0 of total probability.

$A \rightarrow E_1$  if A happens expectation is  $E_1$   
 $A^c \rightarrow E_2$  if A doesn't happen expectation is  $E_2$

ques 1



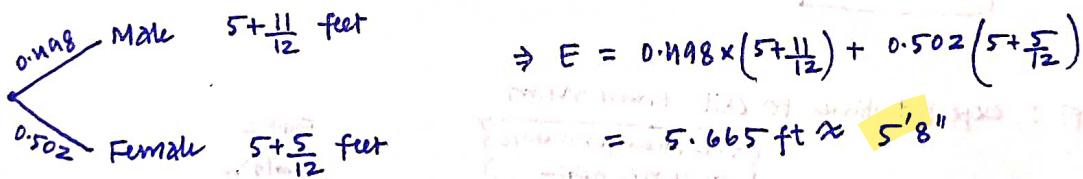
ques 2



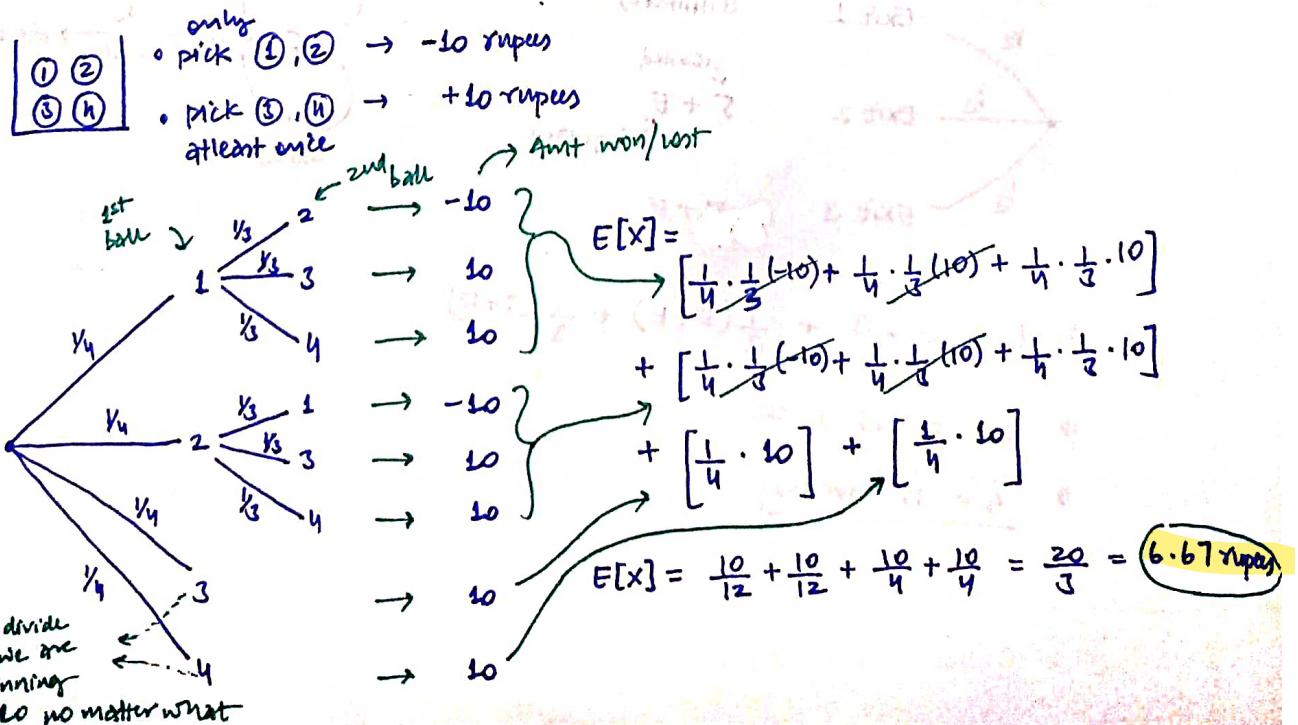
ques 3 Expected height of randomly chosen male =  $5'11'' = 5 + \frac{11}{12}$  feet

female =  $5'5'' = 5 + \frac{5}{12}$  feet

$$P(\text{male}) = \frac{49.8}{100} = 0.498 \quad P(\text{female}) = 0.502$$

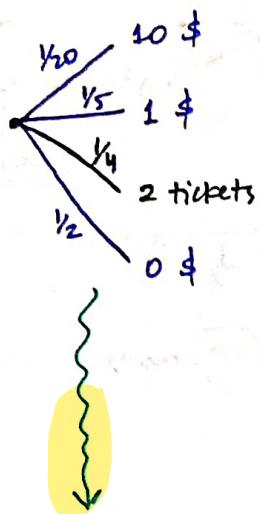


ques 4

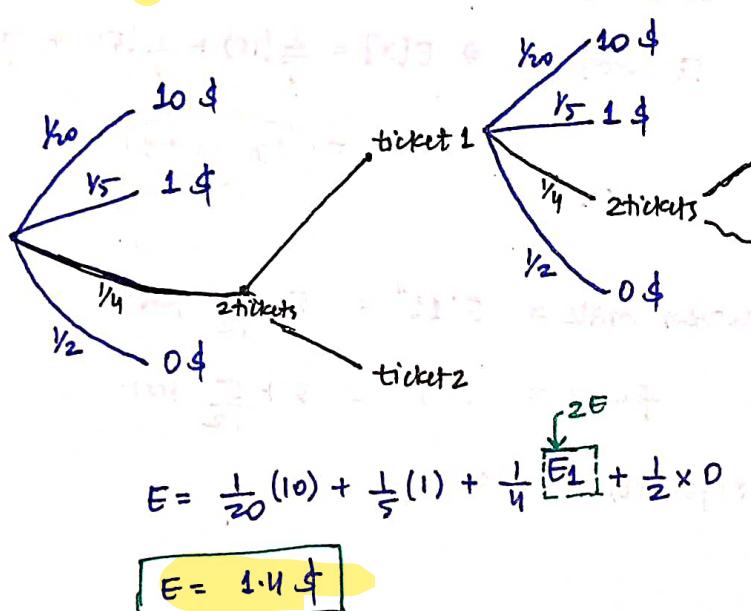


## Recursive way to Find Expectation

ques 5.  $E \rightarrow$  expected no. of prize from one ticket



$$\begin{aligned}
 E &= \frac{1}{20} \times 10 + \frac{1}{5} \times 1 + \frac{1}{4} \times \underline{2E} + \frac{1}{2} \times 0 \\
 \Rightarrow E &= \frac{1}{2} + \frac{1}{5} + \frac{E}{2} + 0 \\
 \Rightarrow \frac{E}{2} &= \frac{7}{10} \\
 \Rightarrow E &= 1.4 \text{ $}
 \end{aligned}$$



$E$ : expected money with the one ticket

if we are winning 2 tickets itself then we are winning  $2E$  \$.

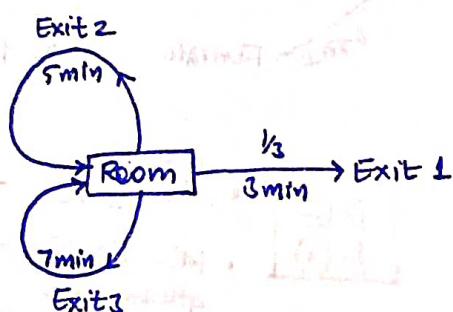
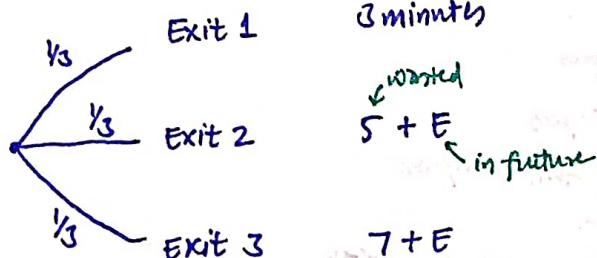
$$E = \frac{1}{20} (10) + \frac{1}{5} (1) + \frac{1}{4} [E_1] + \frac{1}{2} \times 0$$

$$E_1 = 2E$$

$$E = 1.4 \text{ $}$$

ques 6.  $E$ : expected time to exit from room

$\left\{ \begin{array}{l} \text{time taken for movie} \\ \text{to leave maze} \end{array} \right\}$



$$\Rightarrow E = \frac{1}{3} \times 3 + \frac{1}{3} (5+E) + \frac{1}{3} (7+E)$$

$$\Rightarrow 3E = 15 + 2E$$

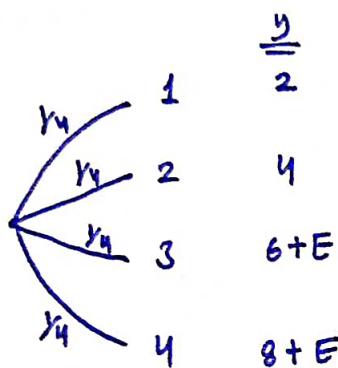
$$\Rightarrow E = 15 \text{ minutes}$$

Ques7 Consider the following recursive func<sup>n</sup>.

```
int nearl()
{
    int b = randomInteger(1,4);
    if (b==1) return 2;
    else if (b==2) return 4;
    else if (b==3) return (6 + nearl());
    else return (8 + nearl());
}
```

Let,  $y$  = value returned by  $\text{nearl}$ .

What is  $E[y]$ ?



$$E = \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 4 + \frac{1}{4} \cdot (6+E) + \frac{1}{4} \cdot (8+E)$$

$$E = \frac{2 + 4 + (6+E) + (8+E)}{4}$$

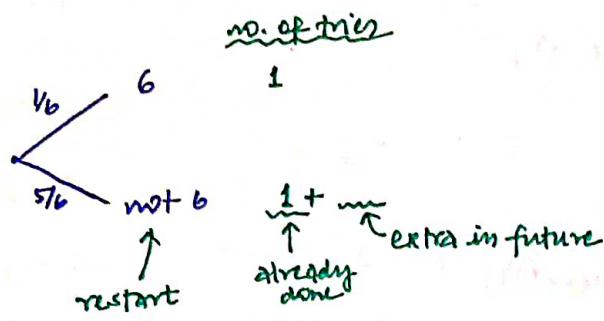
$$4E = 2E + 20$$

$$\boxed{E=10}$$

ques 8.  $X$ : Number of tries to get 6

$$E[X] = ?$$

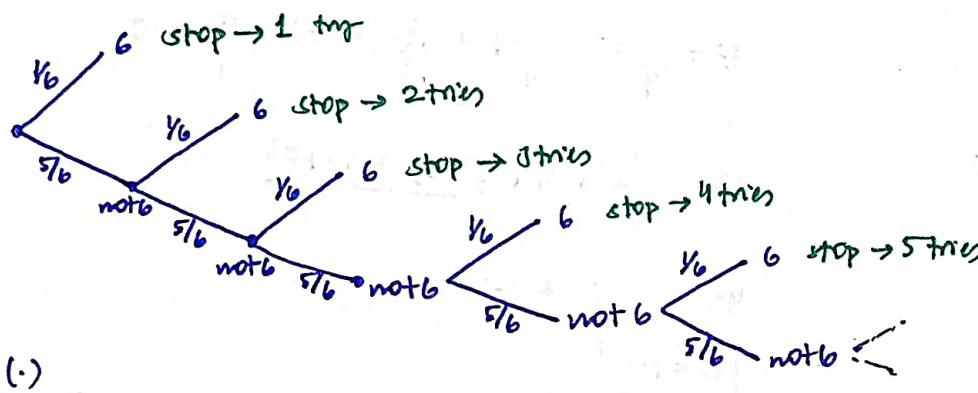
Method 1:



$$\Rightarrow E = \frac{1}{6} \cdot 1 + \frac{5}{6} (1+E)$$

$$E = \frac{1}{6} + \frac{5}{6} + \frac{5}{6} E \Rightarrow \boxed{E=6}$$

Method 2:



$X$	$P(X)$
1	$\frac{1}{6}$
2	$\frac{5}{6} \cdot \frac{1}{6}$
3	$\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$
4	$\left(\frac{5}{6}\right)^3 \cdot \frac{1}{6}$
⋮	⋮

$$\Rightarrow E = 1 \cdot \frac{1}{6} + 2 \cdot \frac{5}{6} \cdot \frac{1}{6} + 3 \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} + 4 \cdot \left(\frac{5}{6}\right)^3 \cdot \frac{1}{6} + \dots$$

$$E = \frac{1}{6} \left[ 1 + 2 \cdot \frac{5}{6} + 3 \cdot \left(\frac{5}{6}\right)^2 + 4 \cdot \left(\frac{5}{6}\right)^3 + \dots \right] \xrightarrow{\text{AGP}}$$

$$\lambda = \frac{5}{6}$$

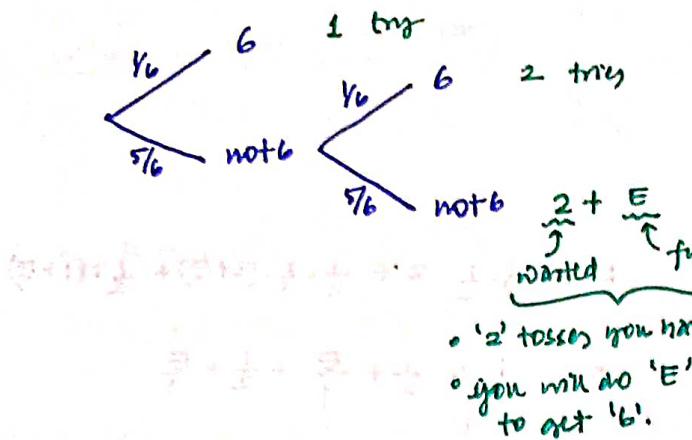
$$E = \frac{1}{6} [1 + 2\lambda + 3\lambda^2 + 4\lambda^3 + \dots]$$

$$S = 1 + 2\lambda + 3\lambda^2 + 4\lambda^3 + \dots \Rightarrow E = \frac{1}{6} \cdot S$$

$$E = \frac{1}{6} \left( \frac{1}{(1-\lambda)^2} \right)$$

$$\begin{aligned} \lambda S &= \lambda + 2\lambda^2 + 3\lambda^3 + \dots \\ S - \lambda S &= 1 + \lambda + \lambda^2 + \lambda^3 + \dots \\ S(1-\lambda) &= \frac{1}{1-\lambda} \Rightarrow S = \frac{1}{(1-\lambda)^2} \end{aligned} \quad E = \frac{1}{6} \cdot \frac{1}{(1-\lambda)^2} \Rightarrow E = \frac{1}{6} \times 6^2 \Rightarrow \boxed{E=6}$$

### Method 3:



$$E = \frac{1}{6} \cdot 1 + \frac{5}{6} \cdot \frac{1}{6} \cdot 2 + \frac{5}{6} \cdot \frac{5}{6} \cdot (2+E)$$

$$\Rightarrow E = \frac{1}{6} + \frac{10}{6^2} + \frac{25}{6^2} (2+E)$$

$$\Rightarrow 36E = 6 + 10 + 50 + 25E$$

$$11E = 66$$

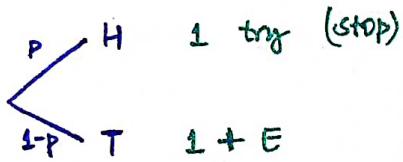
$$\boxed{E = 6}$$

### Ques GATE IT 2006

$$P(H) = P$$

N : no. of tosses till the 1st head appears

### Method 1:



$$E = p \cdot 1 + (1-p)(1+E)$$

$$E = p + 1-p + E - EP$$

$$EP = 1 \Rightarrow \boxed{E = \frac{1}{P}}$$

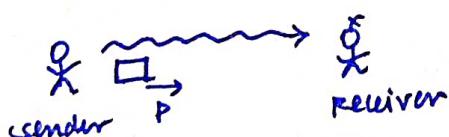
### NOTE → Template of Such question

Head or getting 6 or anything ⇒ success

$$P(\text{success}) = P$$

How many expected trials s.t. you get succeeded? =  $\frac{1}{P}$

### eg GATE PYQ in CN



Sender is sending data pkt to the receiver with probability = P

$$P(\text{receiving packet successfully}) = P$$

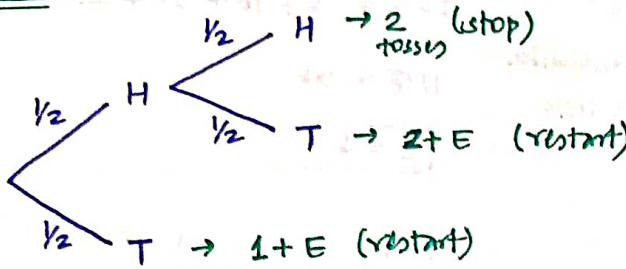
$$\text{Expected no. of tries (packets) to send just 1 packet successfully} = \frac{1}{P}$$

ques 10. TIFR 2015 A-6

$$P(H) = P(T) = \frac{1}{2}$$

TWO heads in a Row :

Method 1 :



Very Nice Question

$$E = \frac{1}{2} \cdot \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot \frac{1}{2} \cdot (2+E) + \frac{1}{2} \cdot (1+E)$$

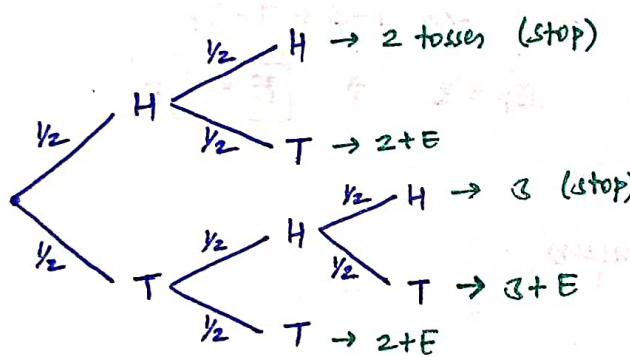
$$E = \frac{1}{2} + \frac{1}{2} + \frac{E}{4} + \frac{1}{2} + \frac{E}{2}$$

$$E = 1 + \frac{1}{2} + \frac{3E}{4}$$

$$\frac{E}{4} = \frac{3}{2} \Rightarrow E = 6$$

Method 2 : infinite series

Method 3 : (partially growing tree)



$$E = \frac{1}{2} \cdot \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot \frac{1}{2} (2+E) + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot 3 + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot (3+E) + \frac{1}{2} \cdot \frac{1}{2} \cdot (2+E)$$

$$E = \frac{1}{2} + \frac{1}{2} (2+E) + \frac{3}{8} + \frac{3}{8} + \frac{E}{8}$$

$$E = \frac{1}{2} + 1 + \frac{E}{2} + \frac{3}{4} + \frac{E}{8}$$

$$E - \frac{5E}{8} = \frac{9}{4} \Rightarrow \frac{3E}{8} = \frac{9}{4} \Rightarrow E = 6$$

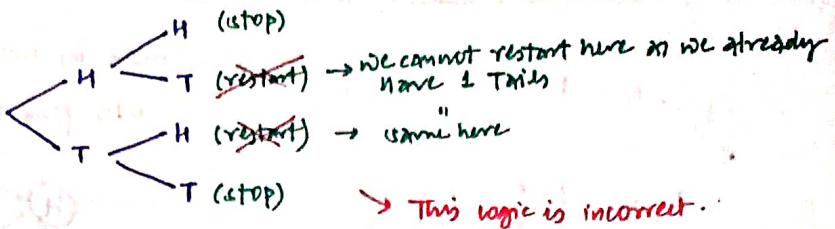
ques 11. GATE 2015

Same ques as 10.

Ans :  $E[X] = b$  tosses

either 2 heads or 2 tails

HH or TT



$$\frac{1}{2} H \rightarrow 1 + E_1$$

$$\frac{1}{2} T \rightarrow 1 + E_2 \Rightarrow E = \frac{1}{2}(1+E_1) + \frac{1}{2}(1+E_2)$$

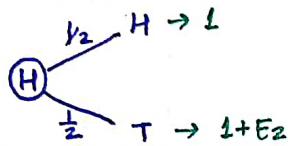
→ This logic is incorrect.

$E_1$ : tosses we need for HH or TT if I already have Head in my hand.

$E_2$ : tosses we need for HH or TT if I already have Tail in my hand.

Now, we'll calculate these  $E_1$  and  $E_2$ ,

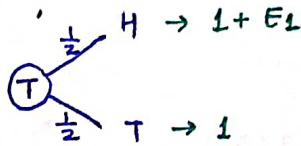
$E_1$ :



$$\Rightarrow E_1 = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (1+E_2)$$

$$E_1 = 1 + \frac{E_2}{2} \quad \text{--- (1)}$$

$E_2$ :



$$\Rightarrow E_2 = \frac{1}{2} \cdot (1+E_1) + \frac{1}{2} \cdot 1$$

$$E_2 = 1 + \frac{E_1}{2} \quad \text{--- (2)}$$

Using (1), (2) we get,

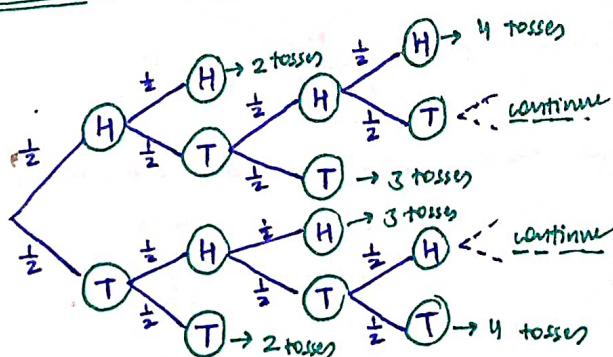
$$E_1 = 2, E_2 = 2$$

Now, put the values in:  $E = \frac{1}{2}(1+E_1) + \frac{1}{2}(1+E_2)$

$$= \frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 3 = 3$$

$$\Rightarrow E = 3$$

Method 2 : expand till infinity



$$E = 2 \left[ 2 \times \underbrace{\left( \frac{1}{2} \right)^2}_{S} + 3 \times \left( \frac{1}{2} \right)^3 + 4 \times \left( \frac{1}{2} \right)^4 + \dots \right]$$

$$E = 2 \times \frac{3}{2}$$

$$\boxed{E = 3}$$

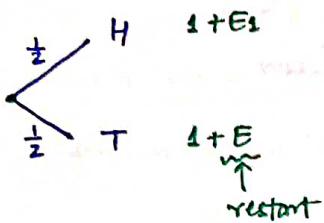
no. of tosses	Prob.
2	$2 \left( \frac{1}{2} \right)^2$
3	$2 \cdot \left( \frac{1}{2} \right)^3$
4	$2 \cdot \left( \frac{1}{2} \right)^4$
$\vdots$	$\vdots$

$$S = 2 \left( \frac{1}{2} \right)^2 + 3 \left( \frac{1}{2} \right)^3 + \dots$$

$$\begin{aligned} \frac{1}{2} S &= 2 \left( \frac{1}{2} \right)^3 + \dots \\ - &- - - - - \\ \frac{S}{2} &= 2 \left( \frac{1}{2} \right)^2 + \left( \frac{1}{2} \right)^3 + \left( \frac{1}{2} \right)^4 + \dots \end{aligned}$$

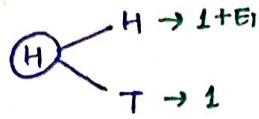
$$\Rightarrow \frac{S}{2} = \frac{1}{2} + \frac{\frac{1}{8}}{1 - \frac{1}{2}} \Rightarrow \frac{S}{2} = \frac{3}{4} \Rightarrow \boxed{S = \frac{3}{2}}$$

Ques 3. No. of tosses until you get HT



$E_1$ : we have head in hand, and looking for HT.

Let's find  $E_1$ ,



$$E_1 = \frac{1}{2}(1+E_1) + \frac{1}{2} \cdot 1$$

$$E_1 = 1 + \frac{E_1}{2}$$

$$E_1 = 2$$

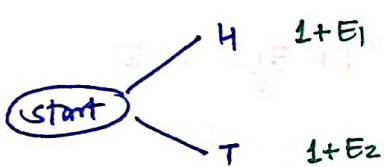
$$\therefore E = \frac{1}{2}(1+E_1) + \frac{1}{2}(1+E)$$

$$E = \frac{1}{2}(1+2) + \frac{1}{2} + \frac{E}{2}$$

$$E = \frac{3}{2} + \frac{1}{2} + \frac{E}{2}$$

$$E = 4$$

→ For HT or TH, we need how many tosses?

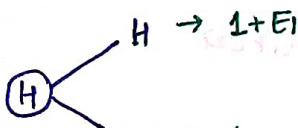


$$E = \frac{1}{2}(1+E_1) + \frac{1}{2}(1+E_2)$$

$$= \frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 3$$

$$E = 3$$

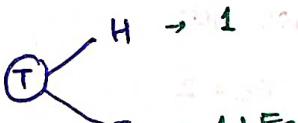
$E_1$ :



$$E_1 = \frac{1}{2}(1+E_1) + \frac{1}{2} \cdot 1$$

$$E_1 = 2$$

$E_2$ :



$$E_2 = 2$$

NOTE → no. of tosses for :

- TT  $\rightarrow 6$
- HH  $\rightarrow 6$
- HH or TT  $\rightarrow 3$
- HT  $\rightarrow 4$
- TH  $\rightarrow 4$
- HT or TH  $\rightarrow 3$

## CUMULATIVE DISTRIBUTION FUNCTION (CDF)

denoted as:  $F_x(x)/F(x)$

$$F(x) = P(X \leq x)$$

(ex)  $x = 1, 2, 3, 4$

$$P(X \leq 3) = P(X=1) + P(X=2) + P(X=3)$$

ques. Find the cumulative distribution func' of the total of heads obtained in 4 tosses of a balanced coin.

↳  $x = \text{no. of heads in 4 tosses}$

$$P(X=0) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$$

$$(TTTT) = \left(\frac{1}{2}\right)^4$$

$$P(X=1) = {}^4C_1 \cdot \left(\frac{1}{2}\right)^4$$

HTTT  
THTT  
TTHT  
TTTH

$$P(X=2) = {}^4C_2 \left(\frac{1}{2}\right)^4$$

$$P(X=3) = {}^4C_3 \left(\frac{1}{2}\right)^4$$

$$P(X=4) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^4$$

HHHH

choose 2 positions  
to put 'H'

$x$	$P(x)$	$F$
0	$\left(\frac{1}{2}\right)^4$	$F(0) = P(X \leq 0) = P(0) = \frac{1}{16}$
1	${}^4C_1 \left(\frac{1}{2}\right)^4$	$F(1) = P(X \leq 1) = P(0) + P(1) = \frac{1}{16} + \frac{4}{16} = \frac{5}{16}$
2	${}^4C_2 \left(\frac{1}{2}\right)^4$	$F(2) = P(X \leq 2) = P(0) + P(1) + P(2) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} = \frac{11}{16}$
3	${}^4C_3 \left(\frac{1}{2}\right)^4$	$F(3) = P(X \leq 3) = P(0) + P(1) + P(2) + P(3) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{15}{16}$
4	$\left(\frac{1}{2}\right)^4$	$F(4) = P(X \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4) = 1$

Hence, the cumulative distribution func' is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ \frac{1}{16} & \text{for } 0 \leq x < 1, \\ \frac{5}{16} & \text{for } 1 \leq x < 2, \\ \frac{11}{16} & \text{for } 2 \leq x < 3, \\ \frac{15}{16} & \text{for } 3 \leq x < 4, \\ 1 & \text{for } x \geq 4 \end{cases}$$

observe that this distribution is defined not only for the values taken on the given RV, but for all real numbers.

For instance, we can write  $F(1.7) = \frac{5}{16}$

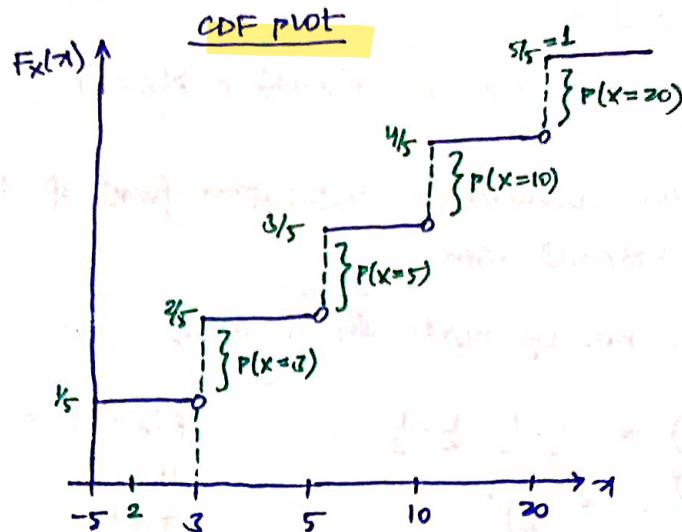
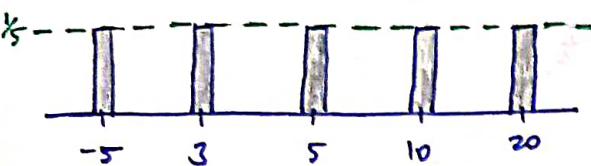
and  $F(100) = 1, F(0.5) = \frac{1}{16}$

although the prob of getting atmost 100 heads, atmost 1.7 heads or atmost 0.5 heads in 4 tosses of a balanced coin may not be of any real significance.

$$\bullet F(2.5) = P(X \leq 2.5) = P(X \leq 2) = F(2)$$

$$\bullet F(10) = P(X \leq 10) = P(X \leq 4) = F(4) = 1$$

### PMF plot



**NOTE** → as  $x$  varies from  $-\infty$  to  $\infty$  the graph of CDF i.e  $F_x(x)$  resembles a staircase with upward steps having height  $P(x=x_i)$  at each  $x=x_i$ .  
But note one thing that the graph of  $F_x(x)$  remains constant b/w 2 steps or event.

→ Can we find PMF from CDF (in DRV) ?

↳ Yes ⇔

② If  $X$  has the distribution func'  $F(1)=0.25$ ,  $F(2)=0.61$ ,  $F(3)=0.83$ ,  $F(4)=1$  for  $x=1, 2, 3, 4$ , find the prob. distribution of  $X$ .

↳ we have;

$$P(1) = F(1) = 0.25$$

$$P(x=2) = F_x(x=2) - F_x(x=1) = 0.61 - 0.25 = 0.36$$

$$P(x=3) = F_x(x=3) - F_x(x=2) = 0.83 - 0.61 = 0.22$$

$$P(4) = F(4) - F(3) = 1 - 0.83 = 0.17$$

$x$	$P(x=x)$
1	0.25
2	0.36
3	0.22
4	0.17

- ques. consider a RV  $X$  that takes values +1 and -1 with prob. 0.5 each. The values of the CDF  $F(x)$  at  $x=-1$  and +1 are
- (A) 0 and 0.5      (B) 0 and 1      (C) 0.5 and 1      (D) 0.25 and 0.75

Given  $P(X \leq 0) = P(X = -1) = 0.5$  and  $P(X = 1) = 0.5$

$$F_X(x = -1) = P(X = -1) = 0.5$$

$$F_X(x = 1) = P(X = 1) + P(X = -1) = 0.5 + 0.5 = 1$$

option (c) correct

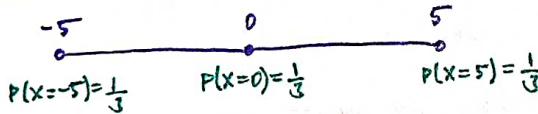
Dont make silly mistake here

## VARIENCE

- It measures how far a set of no. is spread out from their average
- Mean is one way to summarize data and variance is one other way. Together they serve as good summary of data.
- Mean/Expectation : one no. that summarize PMF. It gives us an idea about the avg of the data in the distribution.
- Variance : gives us data about how spread away the data is from the mean. It is average squared distance from mean. It is always non-negative.

Q)

### Dataset - 1 (PMF-1)



• Avg of these points =  $\frac{-5+0+5}{3} = 0$

Mean/Expectation

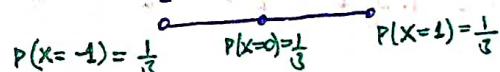
• Variance is average squared distance from mean.

• Variance =  $\frac{(Avg - (-5))^2 + (Avg - 0)^2 + (Avg - 5)^2}{3}$

$$= \frac{(0+5)^2 + 0^2 + (0-5)^2}{3}$$

$$= \frac{50}{3} = 16.67$$

### Dataset - 2 (PMF-2)



• Avg of these points =  $\frac{-1+0+1}{3} = 0$

• Variance =  $\frac{(Avg - (-1))^2 + (Avg - 0)^2 + (Avg - 1)^2}{3}$

$$= \frac{(0+1)^2 + 0^2 + (0-1)^2}{3}$$

$$= \frac{2}{3} = 0.67$$

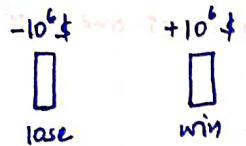
Variance will be low bcz points are closer to mean.

Expectation/mean of both the datasets is 0.

but it visible how different both datasets are,  
so using just expectation we cannot tell much about a dataset.  
Hence variance is also important

## GAMBLING EXAMPLE

### Game I



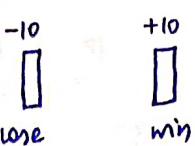
$$P(\text{lose } 10^6 \$) = \frac{1}{2} \quad P(\text{win } 10^6 \$) = \frac{1}{2}$$

Expected money won/loss = 0

$$\text{Variance} = \frac{(10^6)^2 + (10^6)^2}{2} = 10^{12}$$

here variance tells that the risk of loss is very high

### Game II



$$P(\text{lose } 10 \$) = \frac{1}{2} \quad P(\text{win } 10 \$) = \frac{1}{2}$$

Expected money won/loss = 0

$$\text{Variance} = \frac{(10)^2 + (10)^2}{2} = 100$$

the risk of loss is low.

## Sampling from a probability distribution.

$$\begin{aligned} P(H) &= \frac{1}{3} \\ P(T) &= \frac{2}{3} \end{aligned} \quad \left. \begin{array}{l} \text{underlying} \\ \text{prob.} \\ \text{distribution} \\ \downarrow \\ \text{nobody knows this} \end{array} \right\}$$

so we'll do the experiment to collect data. and let's suppose we'll create the perfect sample:

$$\text{HTTHTTHHTT} \rightsquigarrow P(H) = \frac{3}{9} = \frac{1}{3}$$

$$P(T) = \frac{2}{3}$$

sample prob.  
= real prob.

$$\Rightarrow P(X=i) = \frac{1}{6} \quad \text{on rolling a dice}$$

$\downarrow$   
1, 2, 3, 4, 5, 6

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

now, Expectation using Sample data :

Forget about prob. distribution

Sample data: 1, 2, 3, 4, 5, 6  $\rightarrow$  perfect sample

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

$$\text{Variance} = \frac{(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2}{6} = 2.916$$

**NOTE** →

Calculating using data

Calculate using  
prob distribution

Expectation

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\sum_k k P(X=k)$$

Variance

squared distance from mean

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

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

$$\{ E((X - \bar{x})^2) \mid \bar{x} = E[X] \}$$

**NOTE** → The most imp quality associated with a RV  $X$ , other than the mean, is its variance, which is denoted by  $V(X)$  or  $\sigma^2(X)$  and is defined as the expected value of the RV  $(X - E[X])^2$ . i.e,

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

$$\text{Now, } E[(X - \bar{X})^2] = E[X^2 - 2X\bar{X} + \bar{X}^2]$$

mean (average)  
 $\bar{X} \rightarrow \text{expectation}$   
 $\bar{X} = E[X]$

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

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

$$= E[X^2] - 2E[X]E[X] + [E[X]]^2 E[1]$$

buz  
 $E[aX] = aE[X]$

$$= E[X^2] - 2[E[X]]^2 + [E[X]]^2$$

$E[X]$  is a RV or a constant

$$\Rightarrow \boxed{\sigma^2(X) = E[X^2] - [E[X]]^2}$$

Simpler formula for Variance:

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

Both formulas are same

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

we'll always use this

ques. Consider the RV  $X$ , which has the PMF

$$P_X(x) = \begin{cases} \frac{1}{9}, & \text{if } x \text{ is an integer in the range } [-4, 4], \\ 0, & \text{otherwise.} \end{cases}$$

Find out the mean and variance

↳ mean = expectation,  $E[X] = 0$

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

$$= E[(X - 0)^2] = E[X^2]$$

$$\Rightarrow \sigma^2(X) = E[X^2] = \frac{60}{9} \quad \leftarrow \text{Variance}$$

$$E[X] = \sum_k x_k \cdot P(X=k)$$

$$E[X^2] = \sum_k x_k^2 \cdot P(X=k)$$

$$= 2\left((1)\frac{1}{9} + (2)^2 \cdot \frac{1}{9} + (3)^2 \cdot \frac{1}{9} + (4)^2 \cdot \frac{1}{9}\right) + 0 \cdot \frac{1}{9}$$

$$= \frac{60}{9}$$

**Method 2** → using the perfect sample:

forget prob dis, focus on data :  $-4, -3, -2, -1, 0, 1, 2, 3, 4$  ← perfect sample

$$\text{mean} = 0, \quad \text{variance} = \frac{2(16+9+4+1)}{9} = \frac{60}{9}$$

$$(-4+4)/2=0=\text{Mean}$$

### Method 3 $\rightarrow$

The mean  $E[X]$  is equal to 0. This can be seen from the symmetry of the PMF of  $X$  around 0, & can also be verified from the definition.

$$E[X] = \sum_{x=1}^4 x P_X(x) = \frac{1}{4} \sum_{x=1}^4 x = 0$$

Now,

$$\text{let } Z = (X - E[X])^2 = X^2.$$

$$P_Z(z) = \begin{cases} \frac{3}{4} & \text{if } z = 1, 4, 9, 16 \\ \frac{1}{4} & \text{if } z = 0 \\ 0 & \text{otherwise} \end{cases}$$

The variance of  $X$  is then obtained by

$$V(X) = E[Z] = \sum_z z P_Z(z) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{3}{4} + 4 \cdot \frac{2}{9} + 9 \cdot \frac{2}{9} + 16 \cdot \frac{2}{9} = \frac{60}{9}$$

ques For each RV;  $x, y, z$  and  $w$ . compute the mean and variance.

value $x$	1	2	3	4	5
pmf $P(x)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

$$E[X] = \text{mean} = \frac{1+2+3+4+5}{5} = 3$$

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

$$E[X^2] = 1^2 \cdot \frac{1}{5} + 2^2 \cdot \frac{1}{5} + 3^2 \cdot \frac{1}{5} + 4^2 \cdot \frac{1}{5} + 5^2 \cdot \frac{1}{5} = \frac{55}{5} = 11$$

$$= 11 - (3)^2$$

$$= 2$$

value $y$	1	2	3	4	5
pmf $P(y)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{1}{10}$

$$E[Y] = 1 \cdot \frac{1}{10} + 2 \cdot \frac{2}{10} + 3 \cdot \frac{3}{10} + 4 \cdot \frac{2}{10} + 5 \cdot \frac{1}{10} = 3$$

$$E[Y^2] = 1^2 \cdot \frac{1}{10} + 2^2 \cdot \frac{2}{10} + 3^2 \cdot \frac{3}{10} + 4^2 \cdot \frac{2}{10} + 5^2 \cdot \frac{1}{10} = 10$$

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

$$= 10 - (3)^2 = 1$$

(iii)

value z	1	2	3	4	5
pmf $p(z)$	$\frac{5}{10}$	0	0	0	$\frac{5}{10}$

$$E[z] = 1 \cdot \frac{5}{10} + \underbrace{2 \cdot 0 + 3 \cdot 0 + 4 \cdot 0}_{0} + 5 \cdot \frac{5}{10} = \frac{30}{10} = 3 \text{ (2)}$$

$$E[z^2] = 1^2 \cdot \frac{5}{10} + 5^2 \cdot \frac{5}{10} = 125 \cdot \frac{10}{10} = 125$$

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

$$= 125 - 3^2 = 4$$

(iv)

value w	1	2	3	4	5
pmf $p(w)$	0	0	1	0	0

$$E[w] = 0 \cdot 1 + 0 \cdot 2 + 1 \cdot 3 + 0 \cdot 4 + 0 \cdot 5 = 3$$

$$E[w^2] = 1 \cdot 3^2 = 9$$

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

$$= 9 - 3^2 = 0$$

Ques. A random integer N b/w 1 and 5 (inclusive) is chosen. Suppose that N has PMF  $P_N$  given by  $P_N(n) = k \cdot n^2$  for  $n = 1, 2, 3, 4, 5$

for some constant k.

(a) compute  $E[N]$ . Your final answer should not include the letter k.

$$\sum_{n=1}^5 P_N(n) = 1 \Rightarrow (1k + 4k + 9k + 16k + 25k) = 1$$

$$k = 1/55$$

$$\therefore P_N(n) = \frac{n^2}{55}$$

$$\text{now, } E[N] = 1 \cdot \frac{1^2}{55} + 2 \cdot \frac{2^2}{55} + 3 \cdot \frac{3^2}{55} + 4 \cdot \frac{4^2}{55} + 5 \cdot \frac{5^2}{55} = \frac{225}{55} = \frac{45}{11} = 4 \frac{1}{11}$$

(b) compute  $V(N)$ .

$$\text{now, } E[N^2] = 1^2 \cdot \frac{1^2}{55} + 2^2 \cdot \frac{2^2}{55} + 3^2 \cdot \frac{3^2}{55} + 4^2 \cdot \frac{4^2}{55} + 5^2 \cdot \frac{5^2}{55} = \frac{89}{5}$$

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

$$= \frac{89}{5} - \left(\frac{45}{11}\right)^2 = \frac{644}{605}$$

Ques. Consider a sequence of independent tosses of a coin that is biased so that it comes up heads with prob.  $3/4$  and tails with prob  $1/4$ . Let  $X_i$  be 1 if the  $i^{\text{th}}$  toss comes up heads and 0 otherwise.

(a) Compute  $E[X_1]$  and  $\text{Var}[X_1]$

$$P(H) = 3/4 \quad P(T) = 1/4$$

$$P(X_1 = 1) = 3/4$$

PMF of  $X_1$ :

$$P(X_1 = 1) = 3/4$$

$$P(X_1 = 0) = 1/4$$

$$\underline{E[X_1]} = 1 \cdot \frac{3}{4} + 0 \cdot \frac{1}{4} = \frac{3}{4}$$

$$\text{Now, } E[X_1^2] = 1^2 \cdot \frac{3}{4} + 0^2 \cdot \frac{1}{4} = \frac{3}{4}$$

$$\underline{\text{Var}[X_1]} = E[X_1^2] - (E[X_1])^2 = \frac{3}{4} - \left(\frac{3}{4}\right)^2 = \frac{3}{16}$$

(b) Compute  $\text{Var}[X_1 + 2X_2 + 3X_3 + 4X_4]$  it must be true only when  $X_1, X_2, X_3, X_4$  are independent

$$= V[X_1] + V[2X_2] + V[3X_3] + V[4X_4]$$

$$= V[X_1] + 2^2 V[X_2] + 3^2 V[X_3] + 4^2 V[X_4]$$

$$= \frac{3}{16} + 4 \cdot \frac{3}{16} + 9 \cdot \frac{3}{16} + 16 \cdot \frac{3}{16}$$

$$= \frac{3}{16} (1+4+9+16) = \frac{90}{16} = \frac{45}{8}$$

Ques. Suppose that  $P(X=0) = 1 - P(X=1)$ . If  $E[X] = 5\text{Var}(X)$ , find  $P(X=1)$ . Assume non-zero probabilities:

$$\hookrightarrow P(X=0) = 1-p$$

$$P(X=1) = p$$

$$E[X] = 1 \cdot p + 0 \cdot (1-p) = p$$

$$E[X^2] = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$

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

$$= p - p^2 = p(1-p)$$

$$\text{Now, } E[X] = 5V(X)$$

$$\Rightarrow p = 5p(1-p)$$

$$p - 5p(1-p) = 0$$

$$p(1 - 5(1-p)) = 0$$

$$p=0 \quad \text{or}$$

$$5(1-p)=1$$

$$\Rightarrow p = \frac{4}{5}$$

$$\Rightarrow P(X=1) = \frac{4}{5}$$

Ques. Let  $D_1$  and  $D_2$  be the outcomes (in  $\{1, 2, 3, 4, 5, 6\}$ ) of 2 independent fair die rolls. Let  $y_i$  be the RV which is equal to 1 if  $D_1 = i$  and 0 otherwise. Compute the  $\text{Var} \left[ \sum_{i=1}^6 y_i \right]$ .

↪  $D_1 \Rightarrow$  outcome of 1<sup>st</sup> die roll

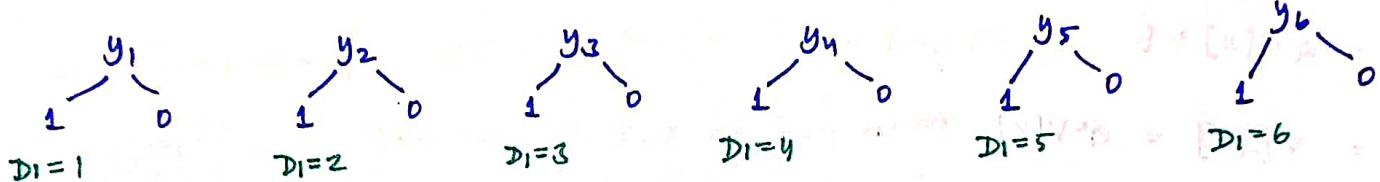
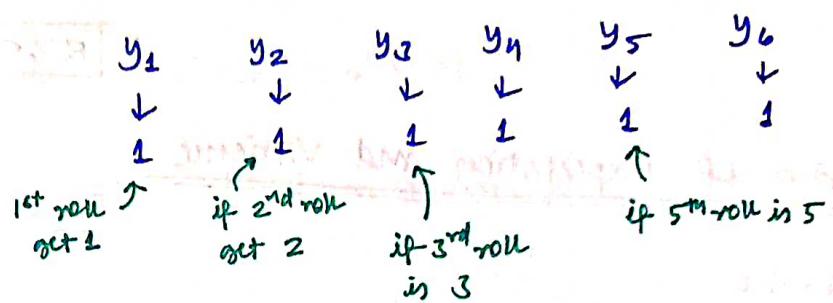
↪  $D_2 \Rightarrow$  outcome of 2<sup>nd</sup> die roll

$$\Rightarrow \{1, 2, 3, 4, 5, 6\}$$

Now,  $y_i$  be the RV which is equal to 1 if  $D_1 = i$  or 0 otherwise.

$$y_1 = 1 \text{ if } D_1 = 1$$

$$y_2 = 1 \text{ if } D_1 = 2$$



⇒ Value of  $y_i$  is totally based on first roll.

i.e. if  $D_1 = 5$ ,  $y_5 = 1$  and everything else is 0, no 2  $y$ 's will be 1 together.

⇒ At any time exactly one of the  $y = 1$

$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$
1	0	0	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0

$\leftarrow D_1 = 1 \rightarrow$  1<sup>st</sup> roll is 1  
 $\leftarrow D_1 = 2 \rightarrow$  1<sup>st</sup> roll is 2  
 $\leftarrow D_1 = 3 \rightarrow$  1<sup>st</sup> roll is 3

Now,  $\text{Var} \left[ \sum_{i=1}^6 y_i \right]$  This sum is constant.

$$= \text{Var} [y_1 + y_2 + y_3 + y_4 + y_5 + y_6]$$

$$= \text{Var}[1] = 0 \quad \text{Answer}$$

$$\left\{ \begin{array}{l} \text{Var}[1] = E[1^2] - (E[1])^2 \\ = E[1] - 1^2 \\ = 1 - 1 = 0 \end{array} \right\}$$

⇒  $\boxed{\text{Var}[\text{any constant}] = 0}$

Ques GATE CSE 2011

If the difference b/w the expectation of the square of a RV ( $E[X^2]$ ) and the square of the expectation of the RV ( $(E[X])^2$ ) is denoted by R, then

(A)  $R=0$

(B)  $R>0$

(C)  $R \geq 0$

(D)  $R<0$

$\hookrightarrow R = E[X^2] - (E[X])^2 = E[(X - E[X))^2] > 0$

Variance

$\therefore y \geq 0 \Rightarrow E[y] \geq E[0]$

$\Rightarrow E[y] \geq 0$

$\Rightarrow R > 0$

### Properties of Expectation and Variance

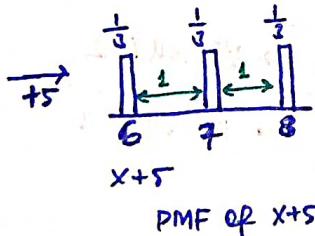
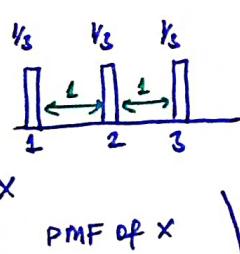
- $y = ax + b$

$$E[y] = a E[x] + b$$

$$V[y] = V[ax] = a^2 V[x]$$

- $E[x_1 + x_2] = E[x_1] + E[x_2]$  (Always True)

$$V[x_1 + x_2] = V[x_1] + V[x_2] \quad (\text{if } x_1 \text{ and } x_2 \text{ are Independent})$$

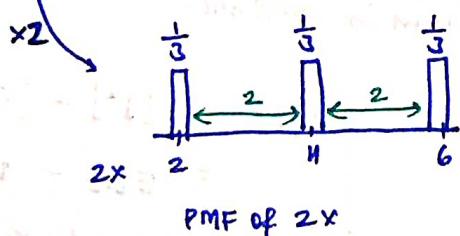


$\Rightarrow$  in this expectation will change

$$E[x+5] = E[x] + 5$$

but variance will be same, as spread is still the same

$$V[x+5] = V[x]$$



$\Rightarrow$  in this expectation will change

$$E[2x] = 2E[x]$$

Variance will also change as spread has changed

$$V[2x] = 2^2 V[x]$$

Ques. True/False

(i)  $V[x-y] = V[x] - V[y]$  ;  $x, y$  are indep.

**FALSE**

$$V[x+y] = V[x] + V[y] \quad (\text{formula})$$

$$V[x-y] = V[x] + V[-y]$$

$$= V[x] + (-1)^2 V[y]$$

$$V[x-y] = V[x] + V[y]$$

(ii)  $V[ax+by+c] = a^2 V[x] + b^2 V[y] + c$  ;  $x, y$  are indep

**FALSE**

$$V[ax+by+c] = a^2 V[x] + b^2 V[y]$$

$$\text{as } V[c] = 0 \\ \downarrow \text{constant}$$

(iii)  $V[ax+by+c] = a^2 V[x] + b^2 V[y] + V[c]$  ;  $x, y$  are indep

**TRUE**

(iv)  $V[ax+by+c] = a^2 V[x] + b^2 V[y]$  for any RVs  $x, y$

**FALSE**

True only if  $x, y$  are independent

(v)  $E[ax+by+c] = a E[x] + b E[y] + c$  for any RVs  $x, y$

**TRUE**

## STANDARD DEVIATION

$$S.D. = \sqrt{\text{Variance}}$$

RV  $X$  has a unit (marks, rupees etc)

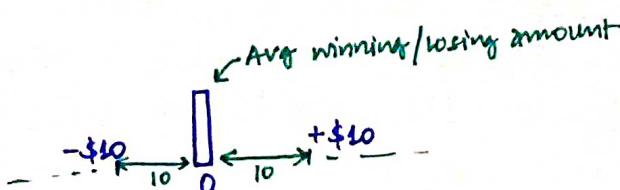
but the unit of variance is  $\text{unit}^2$ .

hence we need SD.

(ex)  $P(\text{winning \$10}) = \frac{1}{2}$  ← we did this example before

$$P(\text{losing \$10}) = \frac{1}{2}$$

$$E[\text{money win or lose}] = 0$$
  
$$\left\{ \frac{1}{2} \cdot 10 + \frac{1}{2} \cdot (-10) = 0 \right\}$$



$$\text{variance} = \frac{(-)^2 + ( )^2 + (+)^2}{3} \text{ unit}^2$$

$$\text{variance}( ) = 100$$

$$SD = 10$$

it means from the expected point we'll be deviating so here and there.

# Discrete Random Variables

Famous examples of DRVs

- Bernoulli  $\rightsquigarrow x=1 \text{ P } x=1 \text{ for some outcome w.p. } P$
- Binomial  $x=0 \quad 1-P \quad x=0, \text{ for some outcome w.p. } 1-P$

• Poisson

• Uniform

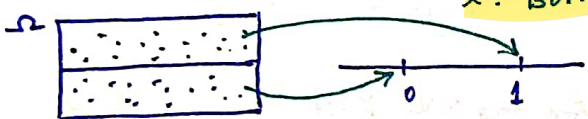
$$\text{DRV } Y = \begin{cases} 0 & \text{w.p. } P_1 \\ 1 & \text{w.p. } P_2 \\ 2 & \text{w.p. } P_3 \end{cases} \quad P_1 + P_2 + P_3 = 1$$

## Bernoulli RV

- For its simplicity, the Bernoulli RV is very important.
- It can take 2 values, 1 and 0. It takes on a 1 if an experiment with prob. P resulted in success and a 0 otherwise.

$$\text{i.e. } x = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$$

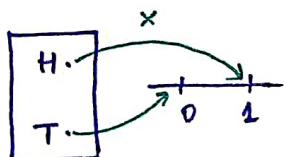
$x$ : Bernoulli RV



(ex) Toss a coin:

$$x(H) = 1$$

$$x(T) = 0$$



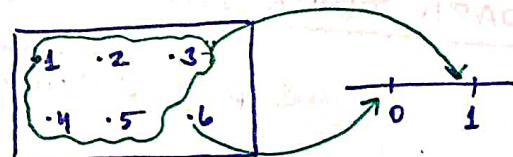
$$P(x=1) = P(H)$$

getting head is a success

$$P(\text{success}) = p = P(H)$$

(ex) Rolling a die!

$$\frac{1, 2, 3, 4, 5, 6}{\downarrow \uparrow} \quad \frac{1}{6}$$



$x$ : Bernoulli RV

**NOTE** → if we have some experiment where we can classify every outcome as success or failure then we say we can have Bernoulli RV.

Ques. consider a RV  $x$ , which takes values 1 or 0.

$$x = \begin{cases} 1 & \text{if a head} \\ 0 & \text{if a tail} \end{cases}$$

$$P(x=\pi) = \begin{cases} P & \text{if } \pi=1 \\ 1-P & \text{if } \pi=0 \end{cases}$$

Find the expectation and variance of  $x$ ?

$$E[x] = 1 \cdot p + 0 \cdot (1-p) = p$$

$$\begin{aligned} V(x) &= E[x^2] - (E[x])^2 \\ &= p - p^2 \\ &= p(1-p). \end{aligned}$$

$$E[x^2] = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$



Thanos: "with the snap of this finger, 50% of the population will vanish, chosen at random."

(Thanos snapped finger...)



Thanos: "Oops... didn't survive the Bernoulli trials!"

NOTE

Remember

BERNDULLI RV

$$1) \text{ PMF : } x = \begin{cases} 1 & p \cdot p \\ 0 & 1-p \end{cases}$$

$$2) \text{ Expectation : } E[x] = p$$

$$3) \text{ Variance : } V[x] = p(1-p)$$

## BINOMIAL RV

Repeated Independent trials of Bernoulli

Trials :  $n$

(e.g) tossing  $n$  times

no. of heads =  $k$

no. of tails =  $n-k$

performing experiment  $n$  times

no. of success =  $k$

no. of failures =  $n-k$

Ques. we do toss 5 times.

$x$  : no. of heads occurred.

$$P(x=3) = ?$$

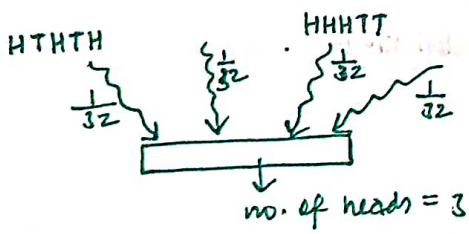
$$\hookrightarrow P(HTHTH) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{32}$$

$$\cancel{P(x=3) = \frac{1}{32}} \quad \text{Wrong}$$

in HTHTH is not the  
only sequence in which  
H is coming 3 times

we need to consider all the combinations in which 'H' is coming 3 times.

$\left\{ \begin{array}{l} \text{HHHHH} \\ \text{TTTTT} \\ \text{HTHTH} \\ \vdots \end{array} \right.$



$$\therefore P(x=3) = {}^5C_3 \cdot \frac{1}{32}$$

$$= 10 \cdot \frac{1}{32} = \frac{5}{16}$$

Ques. A fair coin is tossed  $n$  times.

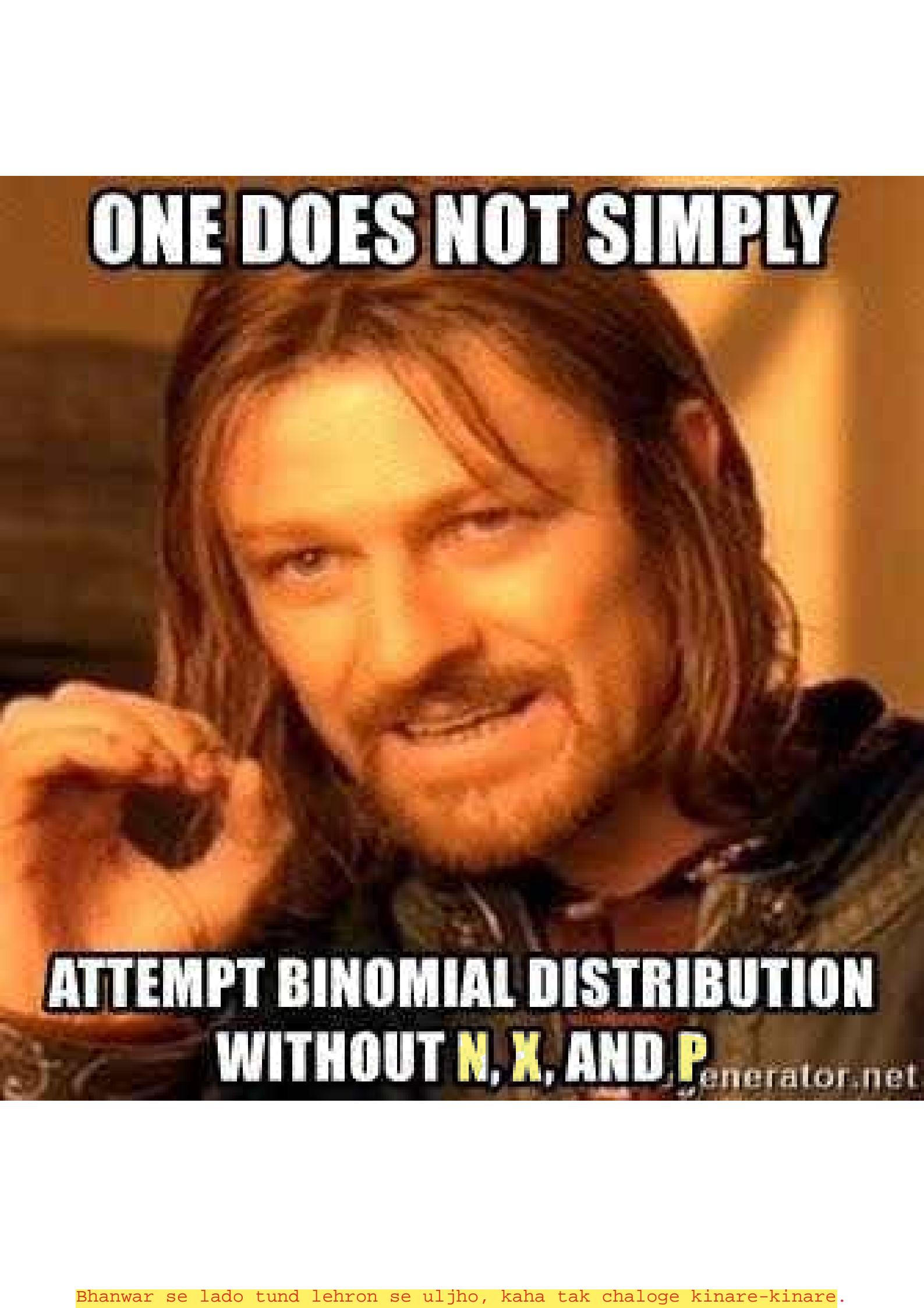
$$P(H) = p$$

$x$  = no. of H

what will be the  $P(x=k)$ ?

$$\hookrightarrow P(x=k) = {}^nC_k p^k (1-p)^{n-k}, \quad k=0,1,2,\dots,n$$

Binomial RV



**ONE DOES NOT SIMPLY**

**ATTEMPT BINOMIAL DISTRIBUTION  
WITHOUT N, X, AND P**

generator.net

### Bernoulli

$$x = 0, 1$$

$$\text{PMF: } P(X=1) = p$$

### Binomial

$$x = 0, 1, \dots, n$$

$$\text{PMF: } P(X=k) = {}^n C_k p^k (1-p)^{n-k}$$

→ Repeated independent trials of Bernoulli = Binomial

Ques. A fair coin is tossed  $n$  times.

$$P(H) = p$$

$$x = \#\text{heads}$$

What will be the  $P(X=k)$ ?

Find out the Expectation & Variance of Binomial RV.

$$\hookrightarrow P(X=k) = {}^n C_k \cdot p^k (1-p)^{n-k}$$

$${}^n C_r = \frac{n!}{r!(n-r)!}$$

x	$P(X=k)$
0	$(1-p)^n$
1	$n \cdot p \cdot (1-p)^{n-1}$
2	${}^n C_2 \cdot p^2 \cdot (1-p)^{n-2}$
⋮	⋮

[Method 1] →

$$E[X] = \sum_{k=0}^n k \cdot P(X=k)$$

after solving this

$$E[X] = \sum_{k=0}^n k \cdot {}^n C_k p^k (1-p)^{n-k} = np$$

(skipping derivation)

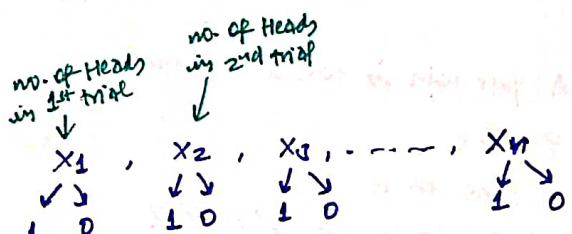
[Method 2] →  $n$  times tossing a coin

$X$ : No of heads →  $0, 1, 2, \dots, n$

$E[X]$  → expected no. of heads

We're doing  $n$  trials

Let's define a separate RV for each trial.



no. of heads

$$X = x_1 + x_2 + x_3 + \dots + x_n$$

Binomial

We have broken down our Binomial RV into  $n$  indep. Bernoulli RVs.

$$X = x_1 + x_2 + \dots + x_n$$

$$\left. \begin{array}{l} x_1 = 1 \\ x_2 = 0 \\ x_3 = 1 \\ x_4 = 1 \end{array} \right\} \text{no. of heads} = 3.$$

now,  $X = X_1 + X_2 + \dots + X_n$

Expectation

$$E[X] = E[X_1 + X_2 + \dots + X_n]$$

$$= \underbrace{E[X_1]}_P + \underbrace{E[X_2]}_P + \dots + \underbrace{E[X_n]}_P$$

$$X_1 = \begin{cases} 0 & \text{, no head in 1st trial} \\ 1 & \text{, head in 1st trial} \end{cases}$$

NICE One

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

Variance

$$V[X] = V[X_1 + X_2 + \dots + X_n]$$

$$= V[X_1] + V[X_2] + \dots + V[X_n]$$
  
$$\downarrow p(1-p) \quad \downarrow p(1-p) \quad \downarrow 1-p$$

$$= np(1-p) = npq \quad (q = 1-p)$$

$x_i$ 's are independent

$$\boxed{V[X] = npq}$$

We do not need the independent assumption for the expected value, since it is a linear func<sup>n</sup> of RVs, but we need it for variance

NOTE → for Binomial RV

$X$  : no. of success in  $n$  trials

Range of  $X$  :  $0, 1, 2, \dots, n$

PMF of  $X$  :  $P(X=k) = {}^n C_k p^k (1-p)^{n-k}$

$E[X] = np$  → no. of trials → prob. of success

$V[X] = npq$   $(q = 1-p)$

Ques. 30 people are invited to a party. Each person accepts the invitation, independently of all others, with probability  $\frac{1}{3}$ .

Let  $X$  be the no. of accepted invitation.

Compute the following:

$$(a) E[X]$$

$$(b) V[X]$$

$$(c) E[X^2]$$

$$(d) E[X^2 - 4X + 5]$$

$\hookrightarrow$   $X$  is binomial

$$X = X_1 + X_2 + \dots + X_n \quad \text{for each person} = X_i$$

$\begin{matrix} \downarrow & \downarrow \\ 1 & 0 \\ \text{Accept} & \text{Reject} \end{matrix}$

$$E[X] = 30 \times \frac{1}{3} = 10$$

$$V[X] = npq$$

$$= np(1-p) = 30 \times \frac{1}{3} \times \frac{2}{3} = \frac{20}{3}$$

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

$$\frac{20}{3} \neq E[X^2] - 100$$

$$E[X^2] = \frac{20}{3} + 100 = \frac{320}{3}$$

$$E[X^2 - 4X + 5] = E[X^2] - 4E[X] + 5$$

$$= \frac{320}{3} - 4 \times 10 + 5 = \frac{215}{3}$$

Ques. Suppose that a fair die is rolled 72000 times. Each roll turns up a uniformly random member of the set  $\{1, 2, 3, 4, 5, 6\}$  and the rolls are independent of each other. For each  $j \in \{1, 2, 3, 4, 5, 6\}$  let  $X_j$  be the no. of times that the die comes up  $j$ .

(a) Compute  $E[X_3]$  and  $V[X_3]$ .

You are rolling a dice 72000 times

and  $X_3$ : no. of 3's

$\hookrightarrow$  success

$X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_6$   
 $\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow$   
no. of 1's    no. of 2's    no. of 3's    no. of 4's    no. of 5's    no. of 6's

in general template  
 $X$ : No. of (success)

$$E[X_3] = np = 72000 \times \frac{1}{6}$$

prob. of success

$$= 12000$$

$$V[X_3] = npq = 72000 \times \frac{1}{6} \times \frac{5}{6} \quad (q = 1 - p = \frac{5}{6})$$

$$= 10000$$

(b) Compute  $V[X_1 + X_2]$ .

$X_1$ : no of 1's.

$X_2$ : no of 2's.

$y = X_1 + X_2$  = no. of 1's and 2's.

success: either 1 comes or 2 comes in any trial

$$P(\text{success}) = 2/6 = 1/3$$

$$E[y] = n \cdot p = 72000 \times \frac{1}{3} = 24000$$

$$V[y] = n \cdot p \cdot q = 72000 \times \frac{1}{3} \times \frac{2}{3} = 16000$$

(c)  $E[X_1 + X_2 + X_3] = ?$

success: either 1, 2 or 3 can come

$$P(\text{success}) = 1/2 (= 3/6)$$

$$E[X_1 + X_2 + X_3] = np = 72000 \times \frac{1}{2} = 36000$$

(OR)

{we don't need any assumption here}  
about dep. or indep.

$$E[X_1 + X_2 + X_3] = E[X_1 + X_2] + E[X_3]$$

$$= 24000 + 12000 = 36000$$

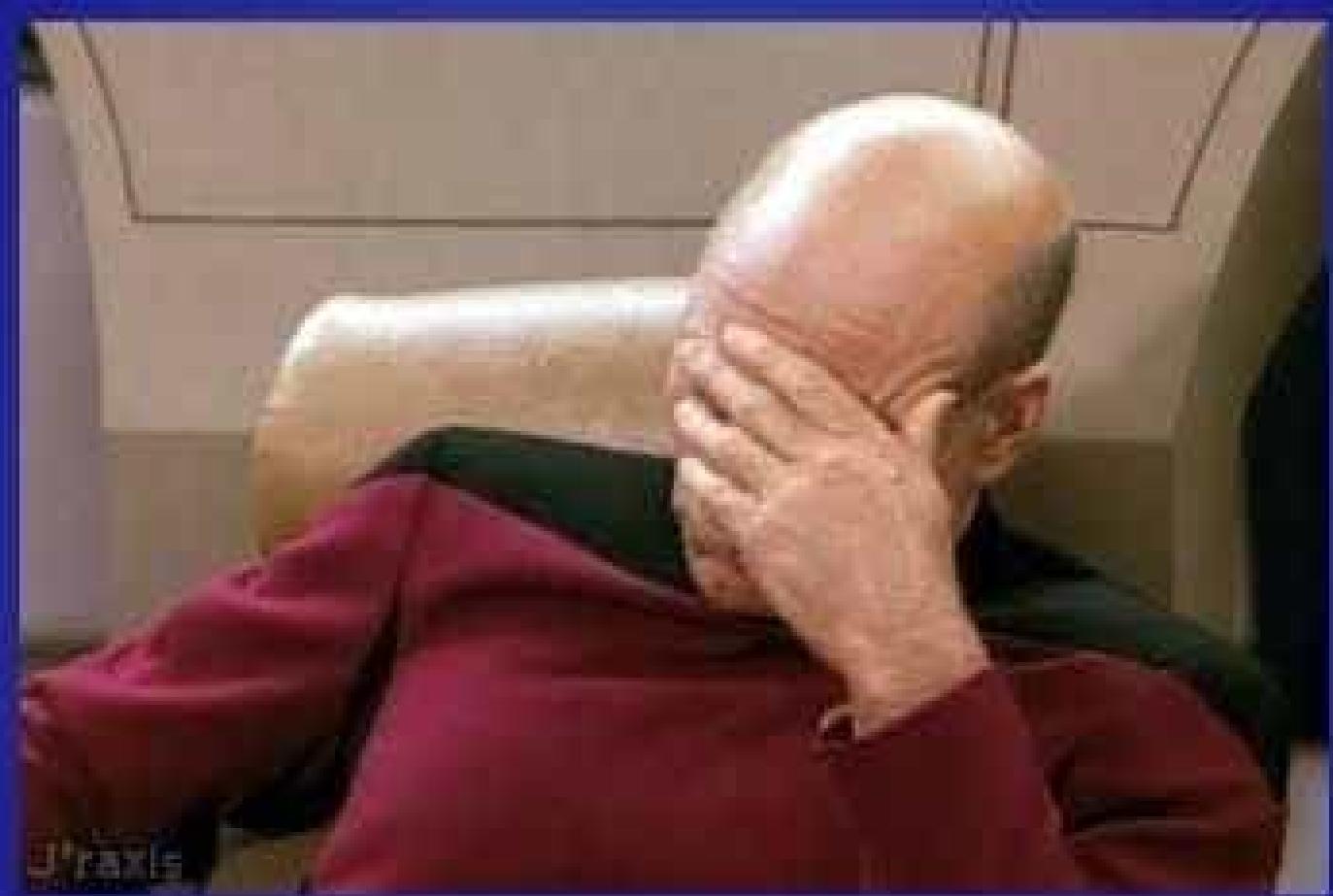
$$\begin{matrix} \downarrow \\ \text{we already did} \end{matrix} \quad \begin{matrix} \downarrow \\ \text{we already did} \end{matrix}$$

$[X_1 + X_2 + X_3 = X_1 + X_2 + X_3]$

for next slide  
lets go to the relationship between variance and covariance

Properties of Covariance & Correlation

# **SMALL SAMPLE SIZE**



# **POISSON DISTRIBUTION**

[monologoforner.net](http://monologoforner.net)

## Poisson Distribution

A R.V  $X$  has the Poisson distribution with parameter  $\lambda$ , where  $\lambda > 0$ , if the PMF of  $X$  is

$$P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k=0, 1, 2, \dots$$

$e = \text{Euler's constant}$   
 $\approx 2.71828$

$X$  can be seen as representing the no of events occurring in a fixed interval where events occur randomly throughout the interval.

Bernoulli :  $X = 0, 1$

Binomial :  $X = 0, 1, 2, \dots, n$

Poisson :  $X = 0, 1, 2, \dots, \infty$

Binomial and Poisson are same, its just that we put  $n=\infty$  in terms of Poisson.  
Binomial with  $n=\infty$  is Poisson.

$$\frac{n C_k p^k (1-p)^{n-k}}{\downarrow n=\infty} \quad \left\{ \begin{array}{l} n \uparrow, p \downarrow \\ \text{and } p \downarrow \rightarrow p \text{ is very small} \end{array} \right. \quad = \frac{e^{-\lambda} \lambda^k}{k!}$$

### Mean and Variance of Poisson RV

$$P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad k=0, 1, 2, \dots$$

$$E[X] = \sum_{k=0}^{\infty} k \cdot P(X=k) = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda} \lambda^k}{k!} = \lambda$$

$$E[X] = V[X] = \lambda$$

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

Ques. On an avg there are 2 accidents per day, then what is prob for 4 accidents per day?

Ans. On an avg there are 2 accidents per day, then what is prob for 4 accidents per day?

$$\hookrightarrow E[X] = \lambda = 2 \leftarrow 2 \text{ accidents per day}$$

$$P(X=4) = \frac{e^{-2} 2^4}{4!} = \frac{e^{-2} 2^4}{24} \approx 0.09$$

**NOTE** More precisely, the Poisson PMF with parameter  $\lambda$  is a good approximation for a binomial PMF with parameters  $n$  and  $p$ , provided  $\lambda=np$ ,  $n$  is very large, and  $p$  is very small. i.e.,

$$\frac{e^{-\lambda} \lambda^k}{k!} \approx {}^n C_k p^k (1-p)^{n-k}, \quad k=0, 1, \dots, n.$$

In this case, using the Poisson PMF may result in simpler models & calculations. For ex, let  $n=100$  and  $p=0.01$ . The prob of 5 successes in 100 trials is calculated using the binomial PMF as  ${}^{100} C_5 (0.01)^5 (1-0.01)^{95} = 0.00290$ .

Using the Poisson PMF with  $\lambda=np=100 \cdot 0.01 = 1$ , this prob is approximated by  $e^{-1} \cdot \frac{1}{5!} = 0.00306$ .

# POISSON DISTRIBUTION BE LIKE



**Me: Mean and variance are different**

**Poisson distribution:**



## How do I identify a Poisson RV?

some examples of RV that generally obey the Poisson prob. law:

- 1) The no. of misprints on a page (or group of pages) of a book.
- 2) The no. of people in a community who survive to age 100.
- 3) The no. of wrong telephone nos. that are dialed in a day.
- 4) The no. of packages of dog biscuits sold in a particular store each day.
- 5) The no. of customers entering a post office on a given day.
- 6) The no. of vacancies occurring during a year in a federal judicial system.
- 7) The no. of  $\alpha$ -particles discharged in a fixed period of time from some radioactive material.

Ques. Suppose that the no. of typographical errors on a single page of this book has a poisson distribution with parameter  $\lambda = \frac{1}{2}$ . Calculate the prob. that there is at least one error on this page.

↳ on a single page avg errors ( $\lambda$ ) =  $\frac{1}{2}$

$X \rightarrow$  no. of errors on this page

$$\text{now, } P(X \geq 1) = 1 - P(X < 1)$$

$$= 1 - P(X = 0)$$

$$= 1 - e^{-\lambda} \approx \boxed{0.393}$$

$$P(X=0) = \frac{e^{-\lambda} \lambda^k}{k!} = \frac{e^{-1/2} (1/2)^0}{0!}$$

$$= e^{-1/2}$$

Ques. On an avg there are 2 accidents per day, what is the prob after 4 accidents in 2 days?

↳  $\lambda = 2 \rightarrow \text{per day}$        $P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!}$

↓ now  $\lambda = 4 \rightarrow \text{per 2 days}$

we have  
to modify the  $\lambda$   
in per question

$$= \frac{e^{-4} (4)^4}{4!} \approx 0.195$$

**IMPORTANT TWIST**

ques. Patients arrive at a hospital accident and emergency department at random at a rate of 6 per hour.

(a) Find the prob that, during any 90 min period, the no. of patients arriving at the hospital accident & emergency department is

$$(i) \text{ exactly } 7 \quad P(X=7)$$

$$(ii) \text{ at least } 3 \quad P(X \geq 3)$$

given  $\lambda = 6$

$$60 \text{ min} = 6 \text{ patients}$$

$$90 = \frac{6}{60} \times 90 = 9 \text{ patients}$$

$$\boxed{\lambda = 9}$$

$$P(X=7) = \frac{e^{-\lambda} \lambda^k}{k!} = \boxed{\frac{e^{-9} 9^7}{7!}}$$

$$P(X \geq 3) = 1 - P(X \leq 2)$$

$$= 1 - [P(X=0) + P(X=1) + P(X=2)]$$

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

$$\frac{e^{-\lambda} \lambda^k}{k!}$$

(b) A patient arrives at 11.30 am. Find the prob that the next patient arrives before 11.45 am.

exactly one patient in 15 min.

$$60 = 6$$

$$15 = \frac{6}{60} \times 15 = 3/2$$

$$\boxed{\lambda = 3/2}$$

$\uparrow$  updated  $\lambda$

$$P(X=1) = \frac{e^{-\lambda} \lambda^k}{k!}$$

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

$$= \frac{3}{2} e^{-3/2}$$

Ques. GATE CSE 2013

Suppose  $p$  is the no. of cars per minute passing through a certain road junction b/w 5 PM and 6 PM, and  $p$  has poisson distribution with mean 3. What is the prob. of observing fewer than 3 cars during any given minute in this interval?

- (A)  $\frac{8}{(2e^3)}$       (B)  $\frac{9}{(2e^3)}$       (C)  $\frac{17}{(2e^3)}$       (D)  $\frac{26}{(2e^3)}$

↪  $p$ : no. of cars per minute

$\lambda = 3$  cars per minute

$$P(p < 3) = P(p=0) + P(p=1) + P(p=2)$$

$$= e^{-\lambda} + e^{-\lambda}\lambda + \frac{e^{-\lambda}\cdot\lambda^2}{2!}$$

$$= e^{-3} \left( 1 + 3 + \frac{3^2}{2!} \right) = \frac{17}{2e^3}$$

Ques. GATE CSE 2017

If a RV  $X$  has a poisson distribution with mean 5, then the expectation  $E[(x+2)^2] = ?$

$$\begin{aligned} \hookrightarrow E[(x+2)^2] &= E[x^2 + 4x + 4] \\ &= E[x^2] + 4E[x] + 4 \\ &= 30 + 4(5) + 4 \\ &= 54 \end{aligned}$$

$$\text{in poisson, } E[x] = V[x] = \lambda = 5$$

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

$$5 = E[x^2] - 25$$

$$E[x^2] = 30$$

Ques. GATE CSE 2007

In a multimer operating system on an avg, 20 requests are made to use a particular resource per hour. The arrival of requests follows a poisson distribution. The prob. that either 1, 3 or 5 requests are made in 45 min is given by:

- (A)  $6.9 \times 10^6 \times e^{-20}$       (B)  $1.02 \times 10^6 \times e^{-20}$       (C)  $6.9 \times 10^3 \times e^{-20}$       (D)  $1.02 \times 10^3 \times e^{-20}$

↪ requests 60 min  $\rightarrow$  20 requests

45 min  $\rightarrow \frac{20}{60} \times 45 = 15$  requests

$$\lambda = 15$$

modified  $\lambda$

$$\begin{aligned} \text{now, } P(x=1) + P(x=3) + P(x=5) &= \frac{e^{-\lambda}\cdot\lambda^1}{1!} + \frac{e^{-\lambda}\cdot\lambda^3}{3!} + \frac{e^{-\lambda}\cdot\lambda^5}{5!} \\ &= e^{-15} \left( 15 + \frac{(15)^3}{3!} + \frac{(15)^5}{5!} \right) \\ &= 6905.625 \times e^{-15} \\ &= 6.9 \times 10^3 \times e^{-20} \times e^5 = 6.9 \times 151.63 \times 10^3 \times e^{-20} \\ &= \boxed{1.02 \times 10^6 \times e^{-20}} \end{aligned}$$

## DISCRETE UNIFORM RV

A RV  $X$  has a discrete uniform distribution if each of the 'n' values in its range, say  $x_1, x_2, \dots, x_n$ , have equal probability. Then,

$$P(x_i) = \frac{1}{n}$$

where  $P(x)$  represents the PMF.

Q)  $\{A, B, C, D\}$

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

Q)  $\{1, 2, 3, 4\}$

$$P(1) = P(2) = P(3) = P(4) = \frac{1}{4}$$

Q)  $\{2, 50, 103, 110\}$

$$P(2) = P(50) = P(103) = P(110) = \frac{1}{4}$$

Q)  $\{1, 2, 3, 4, \dots, n\}$

$$P(1) = P(2) = \dots = P(n) = \frac{1}{n}$$

Q)  $\{2, 4, 6, 8, \dots, 2n\}$

$$P(2) = P(4) = \dots = P(2n) = \frac{1}{n}$$

Q) Roll a dice and let  $X$  be the upward face showing. Find mean & variance

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

PMF:  $P(X=i) = \frac{1}{6}$

$$E[X] = \sum_x x \cdot P(x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

$$E[X^2] = E[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{81}{6}$$

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

$$= \frac{81}{6} - 3.5$$

Ques. If  $x$  is uniformly distributed on the set  $\{1, 2, 3, \dots, N\}$ , then

(i)  $P(x) = ?$

$$P(x) = \frac{1}{N} \quad x \in \{1, 2, 3, \dots, N\}$$

(ii)  $E[x] = \frac{1+2+3+\dots+N}{N} = \frac{\frac{N(N+1)}{2}}{N} = \frac{(N+1)}{2}$

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

$$E[x^2] = \frac{1^2 + 2^2 + \dots + N^2}{N} = \frac{\frac{N(N+1)(2N+1)}{6}}{N} = \frac{(N+1)(2N+1)}{6}$$

$$V[x] = \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4}$$

$$= \frac{(N+1)}{2} \cdot \frac{(N-1)}{6} = \frac{N^2-1}{12}$$

Ques. If  $x$  is a uniformly distributed RV on the set  $\{a, a+1, \dots, b-1, b\}$   
then find out

(i)  $P(x=x) = \frac{1}{(b-a+1)}$   $\rightarrow$  total no. of elements in set

(ii)  $E[x] = \frac{(b-a+1)}{2} \cdot \frac{(a+b)}{(b-a+1)} = \frac{(a+b)}{2}$

(iii)  $V[x] = ?$

$$\begin{aligned} E[x^2] &= \frac{a^2 + (a+1)^2 + \dots + (b-1)^2 + b^2}{(b-a+1)} = \frac{\frac{b(b+1)(2b+1)}{6} - \frac{(a+1)a(2a-1)}{6}}{(b-a+1)} \\ &= \frac{2b^3 + 2b^2 + b - 2a^3 + 3a^2 - a}{6(b-a+1)} \end{aligned}$$

$$\text{now, } V[x] = E[x^2] - (E[x])^2 = \frac{(b-a+1)^2 - 1}{12}$$

## Discrete RVs

1. Bernoulli
2. Binomial
3. Poisson
4. Uniform

## Continuous RVs

1. Uniform
2. Exponential
3. Normal

## CONTINUOUS RV

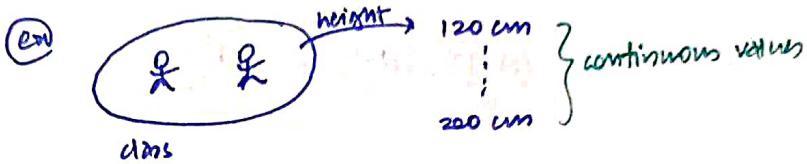
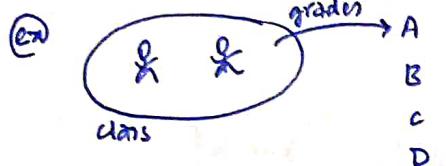
### Discrete RV

v/s

### continuous RV

$X$  takes countable finite values

$X$  will take uncountable values.



- In this we have PMF

- we have PDF here

Probability Density Function (PDF) (denoted by  $f(x)$ )

A RV  $X$  has a PDF  $f(x)$  if;  $P(a \leq X \leq b) = \int_a^b f(x) dx \quad \forall a, b.$

→ if  $a=b \Rightarrow \int_a^a f(x) dx = 0$

→ For valid pdf -

$$P(X \leq \infty) = \int_{-\infty}^{\infty} f(x) dx = 1$$

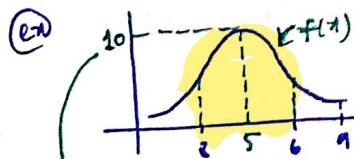
$$\rightarrow P(X \leq b) = \int_{-\infty}^b f(x) dx$$

What is  $f(x)$ , intuitively?

→ prob. density function

→ it tells likelihood.

$f(x)$  cannot tell you the prob directly but comparatively it can tell which range is more likely



$P(2 \leq X \leq 6)$  is more than  $P(6 \leq X \leq 9)$

prob cannot be more than 1.  
so  $f(x)$  doesn't tell us prob.

$|f(5)=10|$  is not prob.

→ Point probability is zero.

mathematically,

$$P(X=a) \Rightarrow P(a \leq X \leq a) = 0$$

$$\int_a^a f(x) dx = 0$$

$$\Rightarrow P(X=a) = 0$$

Prob. of any point is zero.

$$\rightarrow P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

intuitively,

in DRV,

$$P(X=b) = \frac{30}{100}$$

no. of times  
 $X=b$

total

in CRV,

$$P(X=3.0012786310) = \frac{1}{100} = 0$$

finite

Ques. Let  $X$  be a continuous RV with the following PDF.

$$f(x) = \begin{cases} ce^{-x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

where  $c$  is a +ve constant

(a) Find  $c$ .

(b) Find  $P(1 < X < 3)$ .

$$\int_{-\infty}^{\infty} f(x) dx = 1 = \left( \int_{-\infty}^0 f(x) dx \right) + \left( \int_0^{\infty} f(x) dx \right) = 1$$

$$\Rightarrow \int_0^{\infty} ce^{-x} dx = 1 \Rightarrow -c [e^{-x}]_0^{\infty} = -c \left[ \frac{1}{e^{\infty}} - \frac{1}{e^0} \right] = 1$$

$$\Rightarrow c = 1$$

$$\text{Now, } P(1 < X < 3) = \int_1^3 f(x) dx$$

$$= \int_1^3 e^{-x} dx = [-e^{-x}]_1^3 = [e^{-1} - e^{-3}]$$

Ques. GATE CSE 2016

A PDF on the interval  $[a, 1]$  is given by  $\frac{1}{\pi x^2}$ . and outside this interval the value of the funcn is 0. The value of  $a$  is —.

$$f(x) = \begin{cases} \frac{1}{\pi x^2}, & a \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\int_a^1 f(x) dx = 1 \Rightarrow \int_a^1 \frac{1}{\pi x^2} dx = 1 \Rightarrow \left[ \frac{-1}{\pi x} \right]_a^1 = 1 \Rightarrow -\frac{1}{\pi} + \frac{1}{\pi a} = 1$$

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

## Discrete RV

- $X$  can take only discrete values in a set.

$$\text{Ex: } X = \{0, 1, 2, \dots, n\}$$

- PMF  $P(X=x)$

- CDF

$$F(k) = P(X \leq k) = \sum_{x \leq k} P(X=x)$$

- Mean

$$E[X] = \sum x P(x)$$

- Variance

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

## continuous RV

- $X$  can take all possible values in an interval of real line

$$\text{Ex: } X \in [0, 1]$$

- PDF  $f(x)$

- CDF

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

- Mean

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

- Variance

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

## UNIFORM DISTRIBUTION

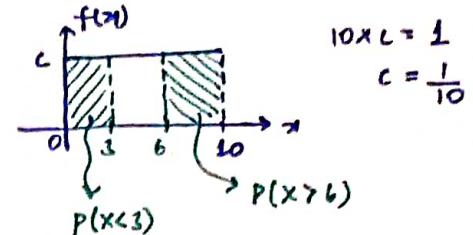
ques. If  $X$  is uniformly distributed over  $(0, 10)$ , calculate the prob that

$$(a) X < 3$$

$$(b) X > 6$$

$$(c) 3 < X < 8$$

$$\Rightarrow f(x) = \begin{cases} \frac{1}{10}, & x \in [0, 10] \\ 0, & \text{otherwise} \end{cases}$$



$$P(X < 3) = \frac{1}{10} \times 3 = \frac{3}{10}$$

$$P(X < 3) = \int_0^3 \frac{1}{10} dx = \frac{3}{10}$$

$$P(X > 6) = \frac{1}{10} \times (10 - 6) = \frac{4}{10}$$

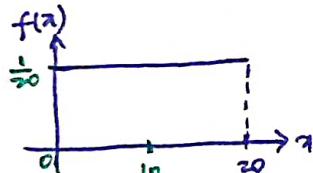
$$P(X > 6) = \int_6^{10} \frac{1}{10} dx = \frac{4}{10}$$

$$P(3 < X < 8) = \frac{1}{10} \times (8 - 3) = \frac{5}{10}$$

$$P(3 < X < 8) = \int_3^8 \frac{1}{10} dx = \frac{1}{2}$$

ques. For the uniform PDF described ... with  $a=0$  and  $b=20$ ,  $f(x) = \frac{1}{20} = 0.05$  for  $x \in [0, 20]$ .  
 $E[X] = ?$

$$P(1 \leq X \leq 15) = ?$$

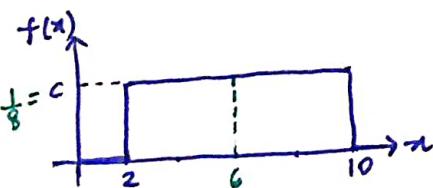


$$P(1 \leq X \leq 15) = \frac{1}{20} \times (15 - 1) = \frac{14}{20}$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{20} \int_0^{20} x dx = \frac{1}{20} \left[ \frac{x^2}{2} \right]_0^{20} = \frac{1}{20} \times \frac{20 \times 20}{2}$$

$$E[X] = 10$$

ques. Bus is uniformly late b/w 2 and 10 minutes. How long can you expect to wait?  
 with what SD? If its > 7 mins late, you'll be late for work.  
 What's the prob of you being late?



$$\circ E[X] = \frac{10 + 2}{2} = 6 \text{ mins}$$

$$\circ V[X] = \frac{(b-a)^2}{12} = \frac{8^2}{12}$$

$$c \times (10 - 2) = 1$$

$$c = \frac{1}{8}$$

$$\sigma = \sqrt{\frac{8^2}{12}} = \frac{8}{\sqrt{12}} = 2.31 \text{ mins}$$

$$\circ P(7 < X < 10) = \frac{1}{8} \times (10 - 7) = \frac{3}{8}$$

## GEOMETRIC DISTRIBUTION

[Position are fixed]  $\rightarrow$  no combination of  $p \& q$

$$\begin{aligned} x = 1 &\rightarrow p \\ = 2 &\rightarrow qp \\ = 3 &\rightarrow qqp \\ &\vdots \end{aligned}$$

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

Mean :

$$\begin{aligned} E[x] &= \sum x P(x) \\ &= 1 \cdot p + 2 \cdot pq + 3 \cdot pq^2 + \dots \\ &= p [1 + 2q + 3q^2 + \dots] \\ &= p [1 - q]^{-2} \\ &= p \cdot p^{-2} = \frac{1}{p} \end{aligned}$$

$$p+q=1$$

$$p = 1-q$$

$$E[x] = \frac{1}{p}$$

Variance :

$$V[x] = \frac{q}{p^2}$$

## UNIFORM DISTRIBUTION

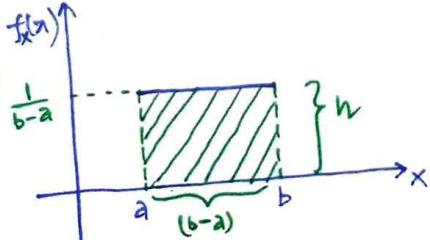
When a RV 'x' is equally likely to be observed in a finite range & is likely to have zero value outside this finite range, then the RV is said to have a uniform distribution.

$\rightarrow x \sim \text{uniform}(a, b)$  means that  $x \in [a, b]$ ,  $a \rightarrow \min^m \text{ value}$ ,  $b \rightarrow \max^m \text{ value}$

$\rightarrow x$  is a continuous RV which is in between  $a$  &  $b$ .

$\rightarrow x = \text{continuous RV}$

$$\rightarrow \int_{-\infty}^{\infty} f_x(x) = 1$$



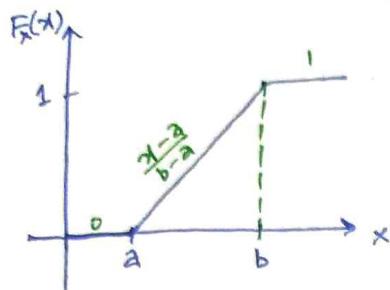
$$(b-a) \cdot h = 1 \Rightarrow h = \frac{1}{(b-a)}$$

(pdf)  $f_x(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{otherwise} \end{cases}$

CDF

$$F(x < a) = 0$$

$$F(x > b) = 1$$



$$\begin{aligned}
 F_x(x) &= P(x \leq x) = \int_{-\infty}^x f_x(x) dx \\
 &= \int_{-\infty}^a 0 dx + \int_a^x \frac{1}{b-a} dx ; a < x < b \\
 &= \frac{1}{b-a} [x]_a^x = \frac{x-a}{b-a} ; a < x < b
 \end{aligned}$$

$$F_x(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a}, & x \in [a, b] \\ 1 & , x \geq b \end{cases}$$

Mean:

$$\begin{aligned}
 \mu &= E[x] \\
 &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^a x \cdot 0 dx + \int_a^b x \cdot \frac{1}{b-a} dx + \int_b^{\infty} x \cdot 0 dx \\
 &= \left[ \frac{x^2}{2(b-a)} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}
 \end{aligned}$$

$$\boxed{\mu = E[x] = \frac{a+b}{2}}$$

Variance:

$$\begin{aligned}
 V[x] &= E[x^2] - [E(x)]^2 \\
 &= \int_a^b x^2 \frac{1}{b-a} dx + \left( \frac{a+b}{2} \right)^2 = \\
 &= \left[ \frac{x^3}{3} \right]_a^b \frac{1}{b-a} + \frac{(a+b)^2}{4} = \frac{b^3 - a^3}{(b-a) 3} + \frac{(a+b)^2}{4} \\
 &= \frac{1}{12} (b-a)^2
 \end{aligned}$$

$$\boxed{V[x] = \frac{1}{12} (b-a)^2}$$

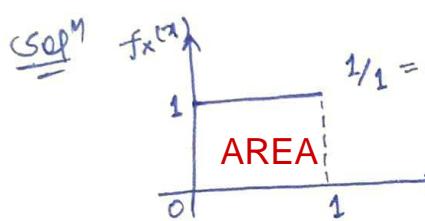
ques. A R.V is uniformly distributed over the interval 2 to 10.  
Its variance will be -

- (A)  $\frac{16}{3}$       (B) 6      (C)  $\frac{256}{9}$       (D) 36

Soln  $V(x) = \frac{(b-a)^2}{12} = \frac{(10-2)^2}{12} = \frac{64}{12} = \boxed{\frac{16}{3}}$

ques. X is uniformly distributed random variable that takes values between 0 & 1. The value of  $E[x^3]$  will be,

- (A) 0      (B)  $\frac{1}{8}$       (C)  $\frac{1}{4}$       (D)  $\frac{1}{2}$



$$f(x) = 1, 0 < x < 1$$

Now, CRV ↗

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

$$= \int_{x=0}^1 x^3 \cdot 1 dx = \left[ \frac{x^4}{4} \right]_0^1 = \boxed{\frac{1}{4}}$$

### Some real life examples of Poisson's distribution

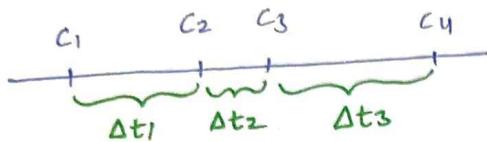
- 1) No. of calls received at a call centre per hour  
↳ Requirements & Assumptions:  
a) calls happen at a constant rate ( $\lambda$ )  
b) occurrence of one call does not affect another.  
c) Two or more events/calls cannot occur simultaneously.  
No. of patients arriving at the emergency b/w 10 PM - 11 PM.
- 2) No. of customers at the counter per hr.
- 3) No. of insurance claims in a year.
- 4) No. of goals in a sports event in b/w 2 teams.
- 5) No. of visitors on your website per minute.

## EXponential Distribution

- Continuous Distribution → continuous R.V
- Let's take an example of a Poisson process for example the calls received at a call centre.

↳ generates  
Poisson distributed RV

uneven  
distribution



We don't know the distribution b/w  $\Delta t_i$ 's  
so we need exponential distribution.

If  $c_1, c_2, c_3, c_4 \dots$  etc follows the Poisson distribution/process then the time interval in between the two Poisson events follows exponential distribution, in the above figure  $\Delta t_1, \Delta t_2, \Delta t_3$  etc follow exponential distribution.

$P(10 \text{ calls in } 1 \text{ hr}) \rightarrow$  can be answered using Poisson's distribution

but,  $P(\text{next event/call occurs in } 3 \text{ min}) \rightarrow$  can't be answered with Poisson's distribution  
But Exponential distribution can.

similar to  
↓ Assume  $c_2$  has happened  
↳ next event is  $c_3$

$$P(\Delta t_2 < 3 \text{ min})$$

Poisson distribution measures the no. of events per unit time.

★ Poisson distribution measures the no. of events per unit time.  
↳ parameter is ' $\lambda$ ' (rate)  $\Rightarrow X \sim \text{Poisson}(\lambda)$

Exponential distribution is the distribution of time intervals b/w events in a Poisson process.  
↳ parameter is ' $\lambda$ ' (rate)  $\Rightarrow X \sim \text{Exp}(\lambda)$

So, if the  $\Delta t_i$ 's (time intervals b/w events), if we think of them as RV ~~X~~ 'X', it follows Exponential distribution.  $X \sim \text{Exp}(\lambda)$ .

### ★ PDF

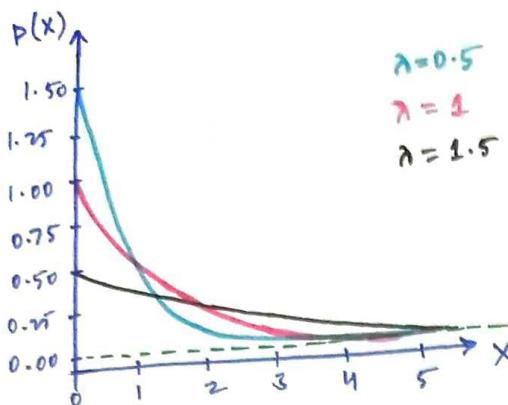
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

### CDF

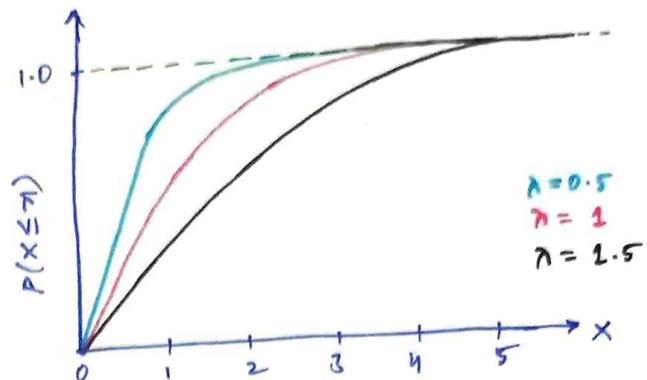
$$\begin{aligned} F_X(x) &= P(X \leq x) = \int_{-\infty}^x f(x) dx = \int_{-\infty}^0 0 dx + \int_0^x \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^x e^{-\lambda x} dx = \lambda \left( \frac{e^{-\lambda x}}{-\lambda} \right)_0^x = 1 - e^{-\lambda x}, \quad x \geq 0 \end{aligned}$$

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

## PDF for Exponential distribution



## CDF for Exponential distribution



• MEAN :  $\mu = E(x)$

$$\begin{aligned} &= \int_0^\infty x (\lambda e^{-\lambda x}) dx \\ &= \left[ -x e^{-\lambda x} \right]_0^\infty + \int_0^\infty e^{-\lambda x} dx = 0 + \left[ \frac{e^{-\lambda x}}{-\lambda} \right]_0^\infty = \frac{1}{\lambda} \end{aligned}$$

using Integration by parts

$$\boxed{\mu = E(x) = \frac{1}{\lambda}}$$

• VARIANCE :  $V(x) = E(x^2) - [E(x)]^2$

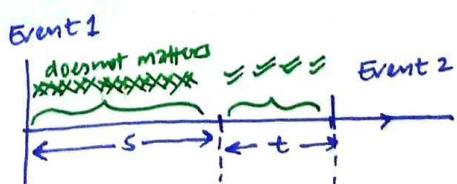
$$\begin{aligned} &= E(x^2) - \mu^2 = \int_0^\infty x^2 \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} \\ &= \frac{1}{\lambda^2} \end{aligned}$$

on doing integration by parts.

$$\boxed{V(x) = \frac{1}{\lambda^2}}$$

## ★⇒ Memoryless Property of Exponential Distribution

$$P(x > s+t | x > s) = P(x > t) \quad \forall s, t \geq 0$$



$$= \frac{P(x > s+t \wedge x > s)}{P(x > s)} = \frac{P(x > s+t)}{P(x > s)}$$

$$= \frac{1 - P(x \leq s+t)}{1 - P(x \leq s)} \Rightarrow \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(x > t)$$

→ Prob that time interval b/w 2 events is ' $s+t$ ' given that time ' $s$ ' has already passed

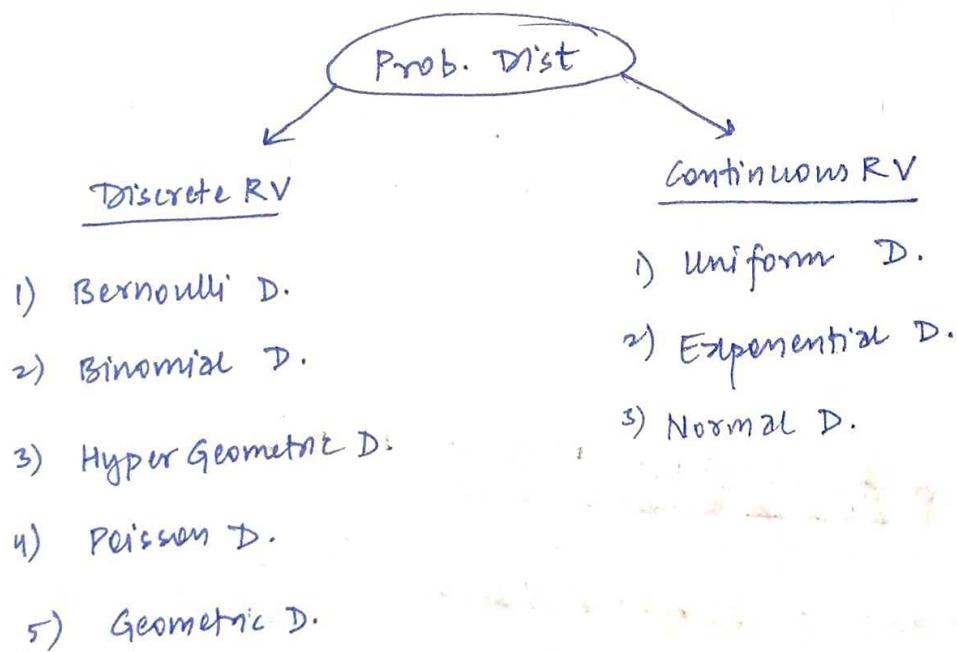
Ques. Assume that the duration in minutes of a telephone conversation follows the exponential distribution.

$f(x) = \frac{1}{5} e^{-x/5}$ ,  $x \geq 0$ . The prob that the conversation will exceed 5 minutes is :

- (A)  $1/e$       (B)  $1 - 1/e$       (C)  $1/e^2$       (D)  $1 - 1/e^2$

Soln prob that the conv. will exceed 5 min

$$\begin{aligned} \hookrightarrow P(X > 5) &= \int_{x=5}^{\infty} f(x) dx \\ &= \int_5^{\infty} \frac{1}{5} e^{-x/5} dx = \frac{1}{5} \left[ \frac{e^{-x/5}}{-1/5} \right]_5^{\infty} \\ &= -1 \left[ e^{-x/5} \right]_5^{\infty} = -1 [e^{-\infty} - e^{-5/5}] = e^{-1} = \boxed{1/e} \end{aligned}$$



# PROBABILITY DISTRIBUTION

## NORMAL / GAUSSIAN DISTRIBUTION

- used for continuous RV. → continuous distribution.
- most widely used & popular distribution.
- ex: height, length of leaves, weights of people in a set of population.  
If not normally distributed they are approximately Normal or Gaussian distributed.
- It is found in so many natural phenomena that a lot of mathematicians/statisticians when prove something like a theoretical model, & have to make an assumption on the distribution of a R.V 'x', if they don't know exactly whether it is a Binomial or Poisson or any other. The most typical assumption they make that 'x' follows Normal distribution. → so very useful

→ A R.V 'x' is normally distributed, represented as.

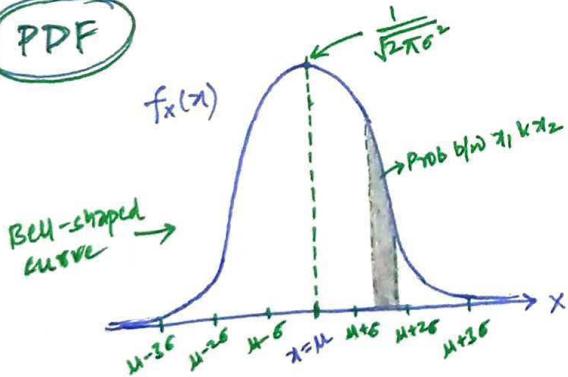
$$x \sim N(\mu, \sigma^2)$$

mean                      Variance

$\mu$  &  $\sigma^2$  should be finite.

$$\Rightarrow \sigma = \text{std. dev} = \sqrt{\text{variance}}$$

**PDF**



$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \sigma \neq 0$$

①  $x \rightarrow$  continuous R.V

②  $\mu = \text{mean}$   
 $\sigma^2 = \text{Variance} \Rightarrow \sqrt{\sigma^2} = \sqrt{V} = \text{SD}$

$\sigma = \text{Standard Deviation}$

③ Symmetric curve (about,  $x=\mu$ )

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

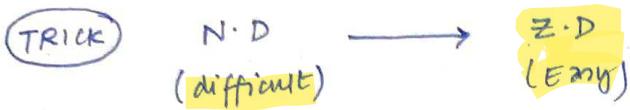
$$\Rightarrow \int_{-\infty}^{\mu} f_x(x) dx = 0.5 ; \quad \int_{-\infty}^{\mu} f_x(x) dx = 0.5$$

$$\int_{-\infty}^{\mu} f_x(x) dx = 0.5$$

$$\text{Now, } P[x_1 < x < x_2] = \int_{x_1}^{x_2} f_x(x) dx$$

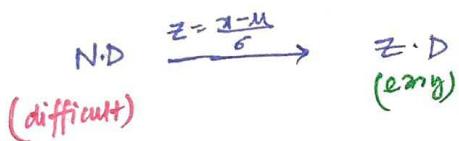
$$= \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

→ it is difficult to solve  
so we use a trick off  
"standard Normal distribution"

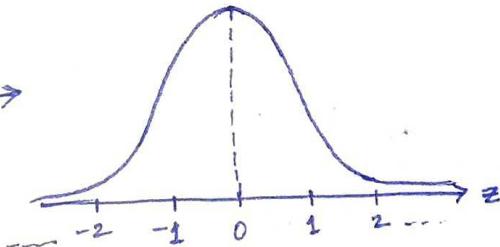


**★ Standard Normal Distribution** : If  $X \sim N(\mu, \sigma^2)$  then we can convert it to another normal variable 'z' with  $\mu=0$  &  $\sigma^2=1$  by the following transformation :

$$z = \frac{x - \mu}{\sigma} \sim N(0,1)$$



$$Z = \frac{x - \mu}{\sigma}$$



[Z DIST.] → standard Normal dist.

$$z = \frac{x - \mu}{\sigma} \rightarrow \sigma z = x - \mu$$

$$\therefore x = \mu + \sigma z \Rightarrow dx = 0 + \sigma dz$$

$$\begin{aligned}
 \text{Now, } P[x_1 < x < x_2] &= P\left[\frac{x_1 - \mu}{\sigma} < \frac{x - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma}\right] = P[z_1 < z < z_2] \\
 &= \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\phi} e^{-\frac{z^2}{2}} \cdot \phi dz \\
 &= \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} dz \quad \checkmark
 \end{aligned}$$

# **BINOMIAL DISTRIBUTION**

## **POISSON DISTRIBUTION**

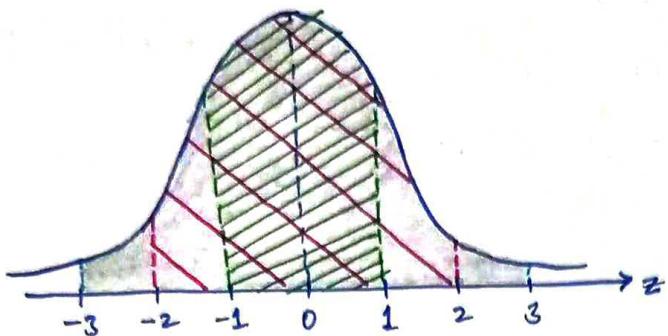
## **NORMAL DISTRIBUTION**

## **STANDARD NORMAL DISTRIBUTION**



Z distribution :

99% chances are that our prob will be in the region,  $-3 < z < 3$ .



$$P(-1 < z < 1) = 0.6827$$

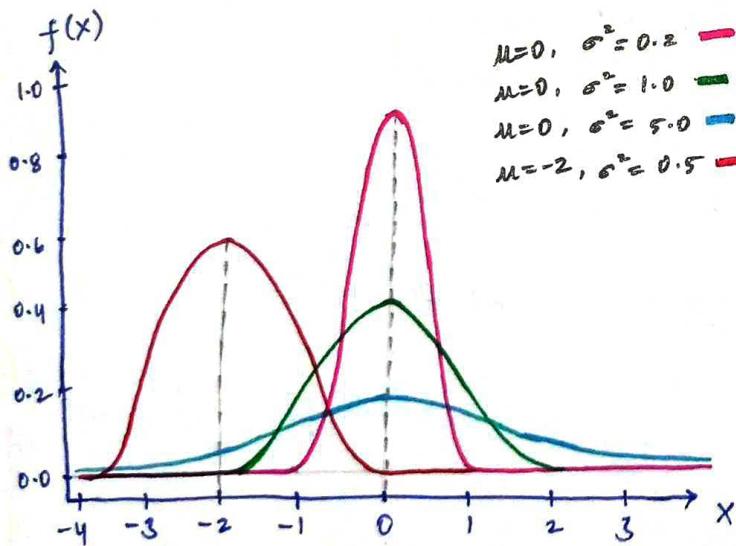
$$P(-3 < z < 3) = 0.9973$$

$$P(-2 < z < 2) = 0.9545$$

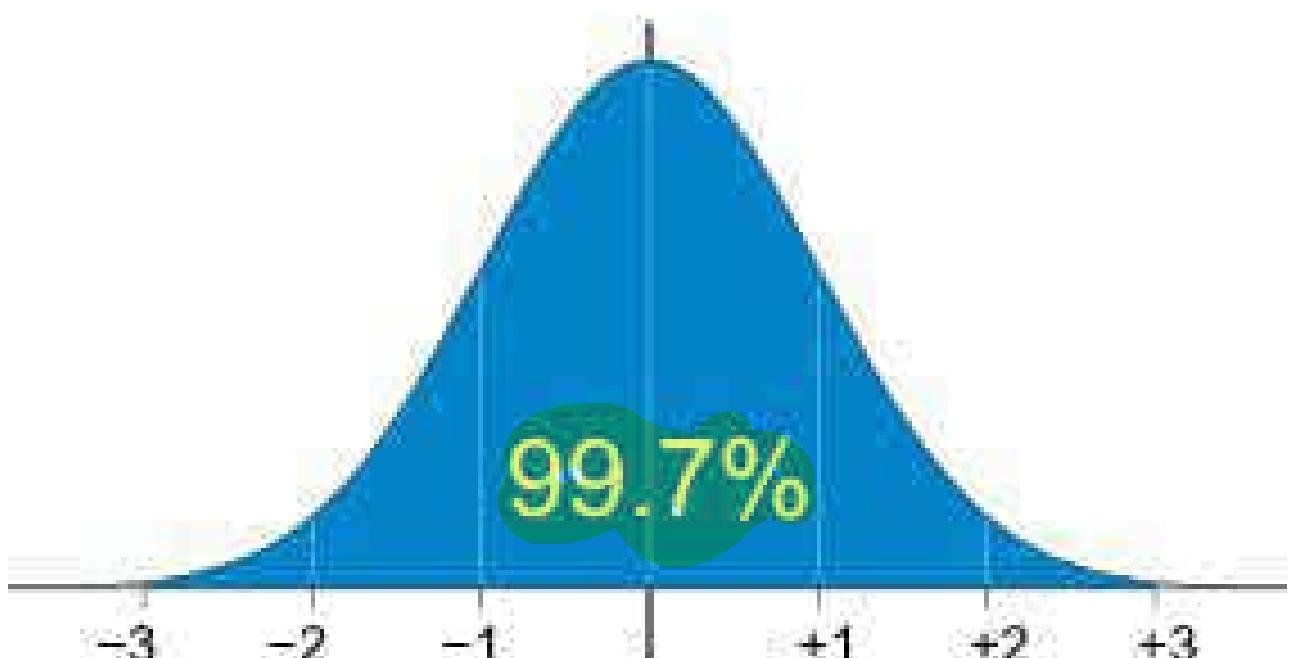
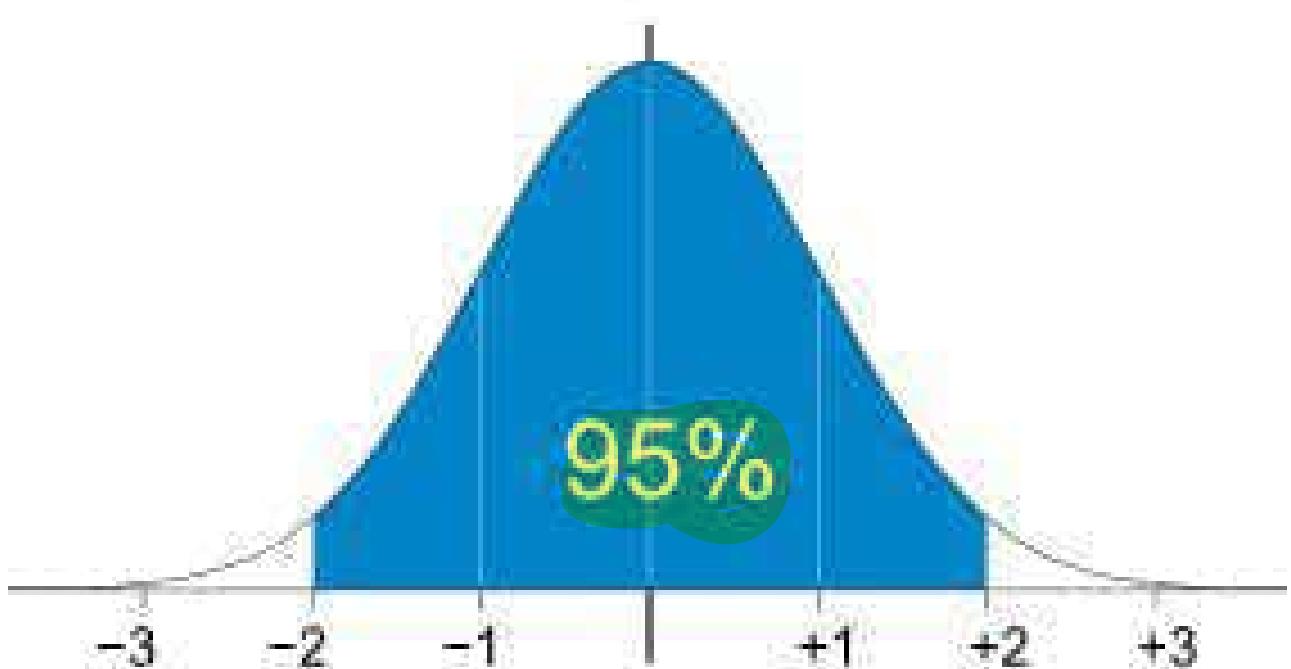
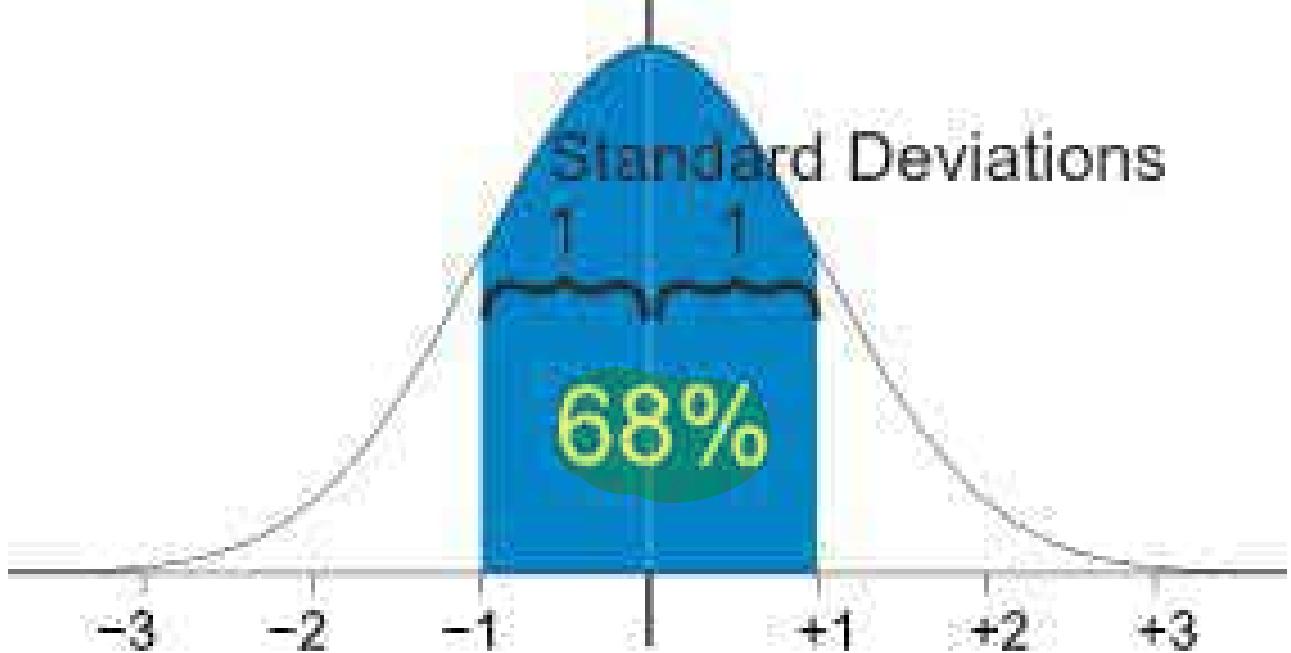
$$\rightarrow P(0 < z < 1) = ??$$

$$\hookrightarrow \text{plot is symmetric} \Rightarrow P(0 < z < 1) = \frac{P(-1 < z < 1)}{2} = \frac{0.6827}{2} = 0.34135 \quad \begin{cases} \text{both symmetric graphs} \\ \text{& overall area = 1.} \end{cases}$$

★ PDF for Normal distribution - Variations



As Variance increases,  
the peakedness of the  
curves decreases in the PDF  
 $\rightarrow$  Variance  $\uparrow$  peak  $\downarrow$



Ques. A nationalized bank has found that the daily balance available in its savings accounts follows a normal distribution with a mean of Rs 500. & a standard deviation of Rs 50. The percentage of savings account holders, who maintain an average daily balance more than Rs 500 is —.

Soln ① Normal distribution  $\curvearrowright$

$$\mu = 500$$

$$\sigma = 50$$

$$P(X > 500) = P(X > \mu) = \int_{\mu}^{\infty} f(x) dx = 0.5$$

$$\% = 0.5 \times 100 = \boxed{50\%}$$

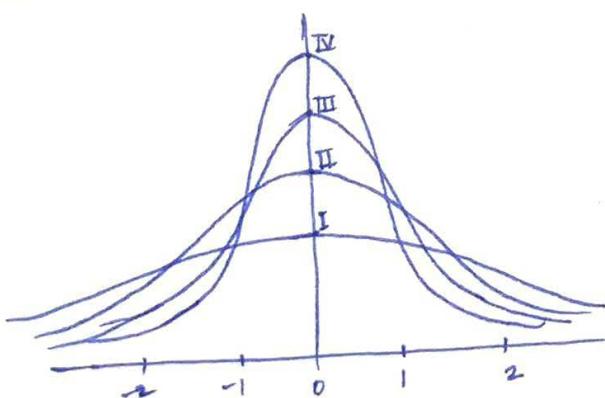
Method 2 :  $P[X > 500]$

convert to  $Z$  distribution

$$= P\left[\frac{X-\mu}{\sigma} > \frac{500-\mu}{\sigma}\right] = P\left[Z > \frac{500-500}{50}\right] = P[Z > 0] = 0.5$$

$$\% = 0.5 \times 100 = \boxed{50\%}$$

Ques. Which one has the lowest variance



(A) I

(B) II

(C) III

(D) IV

Soln

Variance highest  $\rightarrow$  Peak lowest

Variance lowest  $\rightarrow$  Peak highest

$\Rightarrow$  IV  $\curvearrowleft \rightarrow$  lowest variance

Ques. The annual precipitation data of a city is normally distributed with mean & standard deviation as 1000mm & 200mm, respectively. The probability that the annual precipitation will be more than 1200 mm is.

(A) 25%.

(B) 50%.

(C) 75%.

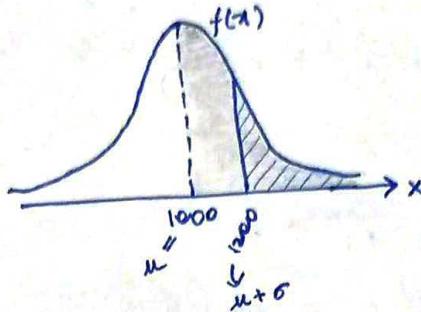
(D) 100%.

Soln ① Normal dist  $\Rightarrow$

$$\mu = 1000$$

$$\sigma = 200$$

$$P(X > 1200) = ??$$



$$P(X > \mu) = 50\%$$

$P(X > 1200)$  would obviously  $< 50\%$ .

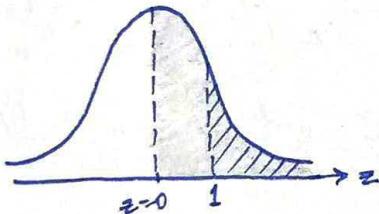
So (A) is correct

Method 2  $Z$ -distribution

$$P(X > 1200) = P\left[\frac{X-\mu}{\sigma} > \frac{1200-\mu}{\sigma}\right]$$

$$Z = \frac{X-\mu}{\sigma}$$

$$= P\left[Z > \frac{1200-1000}{200}\right] = P(Z > 1)$$



$$P(Z > 1) = P(Z > 0) - P[0 < Z < 1]$$

$$= 0.5 - \frac{0.6827}{2} \approx 0.16$$

Ques.  $f_x(x) = \frac{1}{\sqrt{8\pi}} e^{-\frac{(x-1)^2}{8}}$ ,  $-\infty < x < \infty$  then  $\int_{-\infty}^{\infty} f_x(x) dx =$

(A) 0

(B) 1/2

(C)  $1 - \frac{1}{e}$

(D) 1

$$\text{Soln} \quad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Rightarrow 2\pi\sigma^2 = 8\pi \Rightarrow \sigma^2 = 4$$

$$\sigma = 2 \quad (\sigma^2 = 4 \text{ coz S.D can't be -ve})$$

$$\Rightarrow \mu = 1$$

$$\text{Now, } \int_1^{\infty} f_x(x) dx = \int_M^{\infty} f_x(x) dx = 0.5$$

Ques. Consider the following 2 Normal distributions.

$$f_1(x) = e^{-\pi x^2} \quad f_2(x) = \frac{1}{2\pi} e^{\left\{-\frac{1}{4\pi}(x^2+2x+1)\right\}}$$

If  $\mu$  &  $\sigma$  denote the mean & standard deviation, respectively, then

(A)  $\mu_1 < \mu_2$  &  $\sigma_1^2 < \sigma_2^2$

(B)  $\mu_1 < \mu_2$  &  $\sigma_1^2 > \sigma_2^2$

(C)  $\mu_1 > \mu_2$  &  $\sigma_1^2 < \sigma_2^2$

(D)  $\mu_1 > \mu_2$  &  $\sigma_1^2 > \sigma_2^2$

Soln NP:  $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

(i)  $f_1(x) = e^{-\pi x^2}$

$$\Rightarrow -\pi x^2 = -\frac{1}{2} \frac{(x-\mu_1)^2}{\sigma_1^2}$$

$$\Rightarrow \pi(x-0)^2 = \frac{1}{2\sigma_1^2} (x-\mu_1)^2$$

$$\Rightarrow \pi = \frac{1}{2\sigma_1^2}, \quad \mu_1 = 0$$

$$\therefore \sigma_1^2 = \frac{1}{2\pi}$$

$$\Rightarrow \boxed{\mu_1 > \mu_2, \sigma_1^2 < \sigma_2^2}$$

(ii)  $f_2(x) = \frac{1}{2\pi} e^{\left\{-\frac{1}{4\pi}(x^2+2x+1)\right\}}$

$$\Rightarrow -\frac{1}{4\pi}(x^2+2x+1) = -\frac{x^2+2x+1}{2\pi}$$

$$\Rightarrow -2x \frac{[x+1]}{2\pi} = \frac{(x+\mu_2)^2}{\sigma_2^2}$$

$$\Rightarrow \mu_2 = -1, \quad \sigma_2^2 = 2\pi$$

Ques.  $I = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-x^2/8} dx$

$\int_0^\infty e^{-t} t^{n-1} dt = \Gamma(n)$

Soln Method 1 : Using Euler's Gamma function :

$$\Rightarrow \text{Let, } \frac{x^2}{8} = t \Rightarrow x^2 = 8t \Rightarrow 2x dx = 8dt \quad dx = \frac{4}{x} dt \Rightarrow dx = \frac{4}{\sqrt{8t}} dt$$

$$\begin{array}{c|cc} x & 0 & \infty \\ \hline t & 0 & \infty \end{array}$$

After changing variable & substituting :

$$\Rightarrow I = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-t} \frac{4}{\sqrt{8t}} dt \Rightarrow I = \frac{4}{\sqrt{8\sqrt{2\pi}}} \int_0^\infty e^{-t} t^{-1/2} dt$$

$$\Rightarrow I = \frac{1}{\sqrt{\pi}} \int_0^\infty e^{-t} t^{-1/2} dt \Rightarrow I = \frac{1}{\sqrt{\pi}} \times \sqrt{\frac{1}{2}} \Rightarrow I = \frac{1}{\sqrt{\pi}} \times \sqrt{\frac{1}{2}} \Rightarrow \boxed{I=1}$$

$$\begin{aligned} -\frac{1}{2} &= n-1 \\ \Rightarrow n &= 1 - 1/2 = 1/2 \end{aligned}$$

$$\text{Method 2 : } I = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-x^2/8} dx \Rightarrow I = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/8} dx$$

In Normal dis :  $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Comparing powers of exponent :  $\Rightarrow$

$$\Rightarrow -\frac{x^2}{8} = -\frac{(x-\mu)^2}{2\sigma^2}$$

$$\Rightarrow \frac{(x-0)^2}{4} = \frac{(x-\mu)^2}{\sigma^2} \rightarrow \mu=0, \sigma^2=4 \Rightarrow \sigma=2$$

Now,  $I = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/8} dx$   $\rightarrow$  '6' term is missing here, so we'll add it.

$$I = \sqrt{2^2} \int_0^\infty \frac{1}{\sqrt{2\pi}\sqrt{2^2}} e^{-x^2/8} dx$$

$$I = 2 \int_0^\infty \frac{1}{\sqrt{2\pi(2)^2}} e^{-\frac{(x-0)^2}{2(2)^2}} dx$$

$$I = 2 \int_m^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 2 \times \frac{1}{2} = 1$$

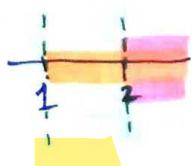
Ques. An exponential random variable is given such that its CDF is given by  $F_Z(z) = \begin{cases} 1-e^{-z} & ; z \geq 0 \\ 0 & ; \text{otherwise} \end{cases}$

Find : probability  $\Rightarrow P[Z > 2 | Z > 1]$ .

$$F_Z(z) = \begin{cases} 1-e^{-z} & ; z \geq 0 \\ 0 & ; \text{otherwise} \end{cases}$$

$$\text{Sop} \quad P[A|B] = \frac{P[A \cap B]}{P[B]}$$

$$\Rightarrow P[Z > 2 | Z > 1] = \frac{P[Z > 1, Z > 2]}{P[Z > 1]}$$



$$= \frac{P[Z > 2]}{P[Z > 1]} = \frac{\int_{z=2}^\infty f(z) dz}{\int_{z=1}^\infty f(z) dz} = \frac{\int_2^\infty e^{-z} dz}{\int_1^\infty e^{-z} dz} = e^{-1} = 0.37$$

P.M.F :  $f(z) = \frac{d}{dz} [CDF]$   
 $\Rightarrow f(z) = \frac{d}{dz} [1-e^{-z}] = e^{-z}$

$$\text{Method 2} \quad \frac{P[Z > 2]}{P[Z > 1]} = \frac{1 - P[Z \leq 2]}{1 - P[Z \leq 1]} = \frac{1 - F_Z(z=2)}{1 - F_Z(z=1)} = \frac{1 - [1 - e^{-2}]}{1 - [1 - e^{-1}]} = \frac{e^{-2}}{e^{-1}} = e^{-1} = 0.37$$

making use of  
recat directy

Ques. If calls arrive at a telephone exchange such that the time of arrival of any call is independent of the time ~~interval~~ of arrival of earlier or future calls, the probability distribution function of the total no. of calls in a fixed time interval will be.

- (A) Poisson      (B) Gaussian      (C) Exponential      (D) Gamma

Soln' Total no. of calls are specific  $\Rightarrow$  discrete RV

$\Rightarrow$  Poisson ✓      b/w Gaussian, Exponential  $\rightarrow$  continuous RV  
Gamma is not present

Ques. The second moment of a Poisson-distributed RV is 2.  
The mean of the RV is \_\_\_\_.

Soln'  $E[X^n]$ ,  $n=1$  (first moment)  
 $n=2$  (second moment)

$$\Rightarrow E[X^2] = 2$$

Now, In Poisson's :  $E[X] = V[X] = \lambda$  (let)

$$\Leftrightarrow V[X] = E[X^2] - [E[X]]^2$$

$$\Rightarrow V[X] = 2 - (\lambda)^2$$

$$\Rightarrow \lambda = 2 - \lambda^2 \Rightarrow \lambda^2 + \lambda - 2 = 0$$

$$\Rightarrow \lambda = 1, \cancel{\lambda = -2} \quad \text{variance can't be -ve}$$

$$\Rightarrow \text{variance} = \boxed{\text{Mean} = 1}$$

Ques. The prob of a resistor being defective is 0.02. There are 50 such resistors in a circuit. The prob of 2 or more defective resistors in the circuit is.

Soln'

$P(D) = 0.02$	<u>Method 1 [Binomial]</u>
$n = 50$	$P(X) = {}^n C_x p^x q^{n-x}$
$P(X \geq 2) = ?$	$n = 50$ $p = 0.02$ $q = 1 - 0.02 = 0.98$
$P(X \geq 2) = 1 - P(X < 2)$	$x = 0, 1$
$= 1 - P(X=0) + P(X=1)$	$\Rightarrow 1 - P(X=0) + P(X=1)$
$= 1 - {}^{50} C_0 p^0 q^{50} + {}^{50} C_1 p^1 q^{49}$	$\Rightarrow 1 - {}^{50} C_0 (0.02)^0 (0.98)^{50} + {}^{50} C_1 (0.02)^1 (0.98)^{49}$
$= 1 - (0.98)^{50} + 50 \times 0.02 (0.98)^{49}$	$\Rightarrow 1 - 0.73577$
$\Rightarrow P(X \geq 2) = 0.264$	$= \boxed{0.264}$

Method 2 [Poisson]  
 $\Rightarrow \lambda = np$   
 $\lambda = 50 \times 0.02 = 1$

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) \\ &= 1 - (P(X=0) + P(X=1)) \\ &= 1 - e^{-1} \left[ \frac{1^0}{0!} + \frac{1^1}{1!} \right] \\ &= 1 - e^{-1}(2) = \boxed{0.26} \end{aligned}$$

ques. Two random variables  $X$  &  $Y$  are distributed according to

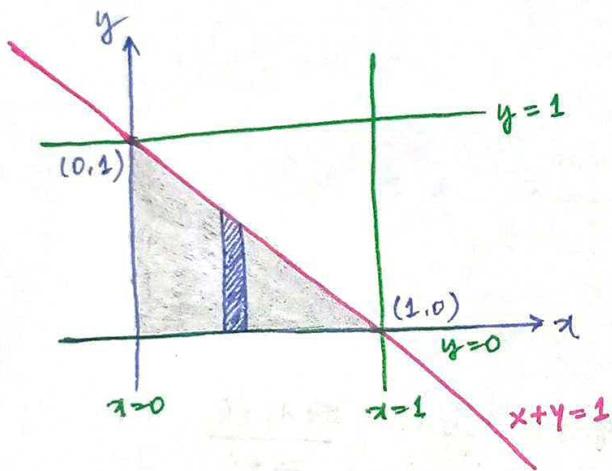
$$f_{X,Y}(x,y) = \begin{cases} (x+y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

The prob  $P(X+Y \leq 1)$  is —.

Soln

$$= \iint_{\substack{x=0 \\ y=0}}^{y=1} f(x,y) dx dy$$

→ here we are having difficulty in finding  $f(x,y)$  & the limits of integration so we have to use the concept of calculus.



using MARIO strip concept

$$\begin{aligned} x &= 0 \text{ to } 1 \\ y &= 0 \text{ to } 1-x \end{aligned}$$

Now,

$$\int_{x=0}^1 \int_{y=0}^{1-x} (x+y) dx dy = 0.33$$

ques. If  $X$  &  $Y$  are RV. such that  $E[2X+Y] = 0$  &  $E[X+2Y] = 33$ ,  
then  $E[X] + E[Y] = ?$

Soln  $E[2X+Y] = 0 \rightarrow E[2X] + E[Y] = 0 \rightarrow 2E[X] + E[Y] = 0$

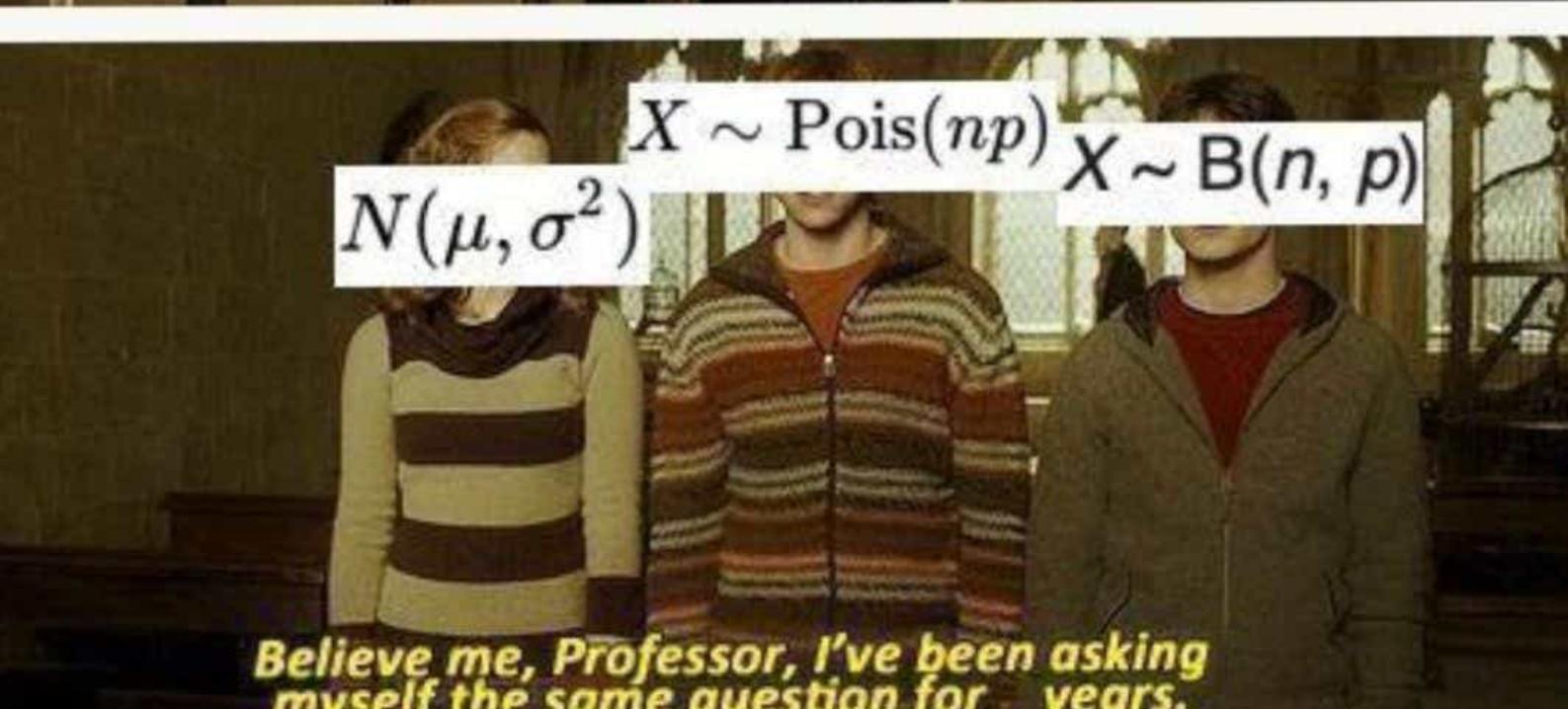
$$E[X+2Y] = 33 \rightarrow E[X] + E[2Y] = 33 \rightarrow \frac{E[X] + 2E[Y] = 33}{3E[X] + 3E[Y] = 33}$$

$$\Rightarrow E[X] + E[Y] = 11$$



*Why is it when something happens, it is always you three?*

$$X \sim \text{Pois}(np) \quad X \sim \text{B}(n, p)$$
$$N(\mu, \sigma^2)$$



*Believe me, Professor, I've been asking myself the same question for years.*

# STATISTICS & PYQs

1) **MEAN** = Average

↳ [Ungrouped Data]

Ex:  $\{1, 8, 3, 2, 6\}$

$$\text{Avg} = \frac{\sum x_i}{n} = \frac{1+8+3+2+6}{5} = \frac{20}{5} = \boxed{4}$$

↳ [Grouped Data]

Ex

$x_i$	$f_i$
8	1
6	1
5	3
7	2
4	2
9	1

frequency repetition  
can be written in ungrouped data

$\text{Avg} = \frac{\sum x_i}{n} = \frac{\sum x_i f_i}{\sum f_i}$

$$= \frac{8+6+15+14+8+9}{10} = \frac{60}{10} = \boxed{6}$$

Ans.

Flow rate (litre/sec)	7.5 to 7.7	7.7 to 7.9	7.9 to 8.1	8.1 to 8.3	8.3 to 8.5	8.5 to 8.7
Frequency	1	5	35	17	12	10

Mean flow rate of the liquid is

(A) 8.00 l/s      (B) 8.06 l/s      (C) 8.16 l/s      (D) 8.26 l/s

Soln: take avg of flow rate as  $\frac{(\text{upper value} + \text{lower value})}{2}$  ex:  $x_i = \frac{7.5+7.7}{2} = 7.6$

Now,  $\text{Avg} = \frac{\sum x_i f_i}{\sum f_i} = \frac{(1 \times 7.6) + (5 \times 7.8) + (35 \times 8) + (17 \times 8.2) + (12 \times 8.4) + (10 \times 8.6)}{80}$

$$= \frac{652.8}{80}$$

$$= \boxed{8.16 \text{ l/s}}$$

Modified table will be using  $x_i = \frac{(x_i)^{\text{upper}} + (x_i)^{\text{lower}}}{2}$

$x_i$	7.6	7.8	8	8.2	8.4	8.6
$f_i$	1	5	35	17	12	10

## 2) MEDIAN

- Step 1 : Sort all the values (either Ascending / Descending).
- Step 2 : Select the middle value.

(ex)  $\{2, 1, 3\}$

$$\hookrightarrow 1, \underset{\uparrow}{2}, 3$$

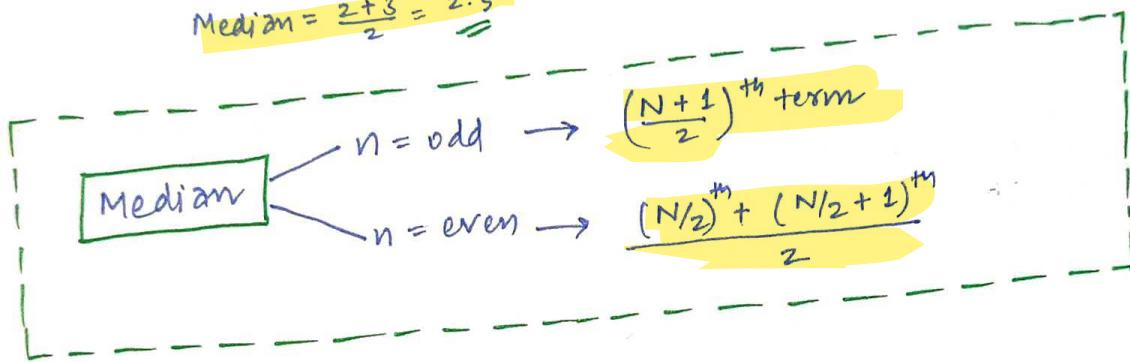
Median = 2

(ex)  $\{1, 4, 2, 3\}$

$$\hookrightarrow 1, \boxed{2, 3}, 4$$

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

- Step 3 : If we have even no. of values, we need to take avg of middle 2 elements.



(ex)  $\{1, 2, 7, 6, 8\}$

$$\hookrightarrow 1, 2, \underset{\uparrow}{6}, 7, 8 \Rightarrow \text{Median} = \left(\frac{5+1}{2}\right)^{\text{th}} = 3^{\text{rd}} \text{ term}$$

Median  $\hookrightarrow \boxed{6}$

$$\text{Mean} = \frac{1+2+6+7+8}{5} = \boxed{4.8}$$

Ques. The spot speeds (expressed in km/h) observed at a road section are 66, 62, 45, 79, 32, 51, 56, 60, 53 & 49. The median speed is ??

Soln In Ascending order : 32, 45, 49, 51, 53, 56, 60, 62, 66, 79

$$n = 10 \Rightarrow \text{Median} \rightarrow \frac{\left(\frac{10}{2}\right)^{\text{th}} + \left(\frac{10}{2} + 1\right)^{\text{th}}}{2} = \frac{5^{\text{th}} + 6^{\text{th}}}{2} = \frac{53 + 56}{2} = \boxed{54.5}$$

Ques. The marks obtained by a set of students are:

38, 84, 45, 70, 75, 60, 48

The mean & median marks, respectively are

(A) 45 & 75

(B) 55 & 48

(C) 60 & 60

(D) 60 & 70

Soln

$$\text{Mean} = \frac{38 + 84 + 45 + 70 + 75 + 60 + 48}{7} = \frac{420}{7} = \underline{\underline{60}}$$

In Ascending order: 38, 45, 48, 60, 70, 75, 84

$$\text{Median} = \left(\frac{N+1}{2}\right)^{\text{th}} \text{ term} = \left(\frac{7+1}{2}\right)^{\text{th}} = 4^{\text{th}} \text{ term} \rightarrow \underline{\underline{60}}$$

### 3) MODE

The value that is observed most frequently i.e., <sup>has</sup> highest freq in  $x_i$ :

<u>Ex</u> )	$x_i$	$f_i$
1	1	1
2	2	2
3	3	3
7	7	4 → 7 is mode of the dataset
10	10	2

<u>Ex</u> )	Height	frequency
150	10	
155	20	
160	70	→ 160 is mode of the dataset.
170	15	
180	4	

Ex)  $\{1, 2, 2, 3, 3, 3\} \rightarrow \boxed{\text{Mode} = 3} \rightarrow \text{UNIMODEL} \rightarrow \text{only one value of mode}$

Ex)  $\{1, 2, 2, 3, 3\} \rightarrow \boxed{\text{Mode} = 2 \& 3} \rightarrow \text{BIMODEL} \rightarrow 2 \text{ values of mode}$

Ex)  $\{1, 2, 2, 3, 3, 4, 4\} \rightarrow \boxed{\text{Mode} = 2, 3 \& 4} \rightarrow \text{TRIMODEL}$

Ex)  $\{1, 2, 2, 3, 3, 4, 4, 5, 5\} \rightarrow \boxed{\text{Mode} = 2, 3, 4 \& 5} \rightarrow \text{MULTIMODEL}$

Ex)  $\{1, 2, 3, 4, 5\} \rightarrow \text{All have same frequency}$   
 $\hookrightarrow \boxed{\text{No Mode}}$

Ques. Marks obtained by 100 students in an examination are given in the table

S.N.O.	Marks obtained	No. of students
1	25	20
2	30	20
3	35	40
4	40	20

What would be the mean, median and mode of the marks obtained by the students?

- (A) Mean : 33, Median : 15, Mode : 40
- (B) Mean : 35, Median : 32.5, Mode : 40
- (C) Mean : 33, Median : 35, Mode : 35
- (D) Mean : 35, Median : 32.5, Mode : 35

Copy No of students  $\rightarrow f_i$

Marks obtained  $\rightarrow x_i$

$\Rightarrow \underline{\text{Mode}} = 35$  as its frequency is max.  
 {35 came 40 times which is highest}

don't do silly mistakes

$$\Rightarrow \underline{\text{Mean}} : \frac{\sum x_i f_i}{\sum f_i} = \frac{25(20) + 30(20) + 35(40) + 40(20)}{100} = \frac{3300}{100} = 33 \approx$$

$\Rightarrow \underline{\text{Median}}$ :	1	2	3	-----	20	$\rightarrow 25 \text{ marks}$
	21	-----	40	-----	40	$\rightarrow 30 \text{ marks}$
	41	-----	60	-----	60	$\} 35 \text{ marks}$
	61	-----	80	-----	80	$\rightarrow 40 \text{ marks}$

$$n = 100$$

$$\begin{aligned} &\Rightarrow \frac{\left(\frac{100}{2}\right)^{\text{th}} + \left(\frac{100}{2} + 1\right)^{\text{th}}}{2} \\ &= \frac{50^{\text{th}} + 51^{\text{th}}}{2} = \frac{35 + 35}{2} \end{aligned}$$

$$\Rightarrow \underline{\text{Median}} = 35 \approx$$

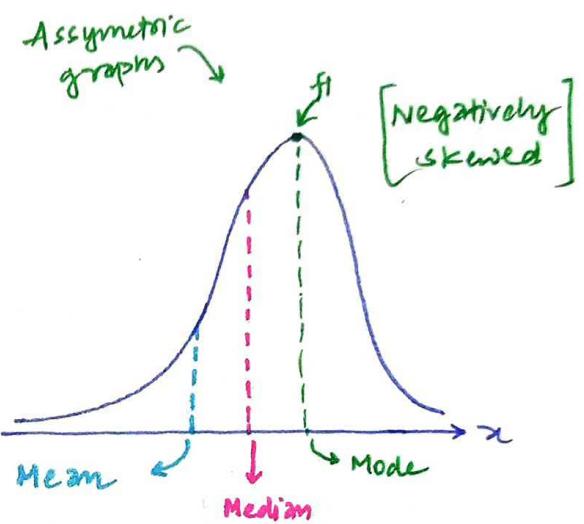
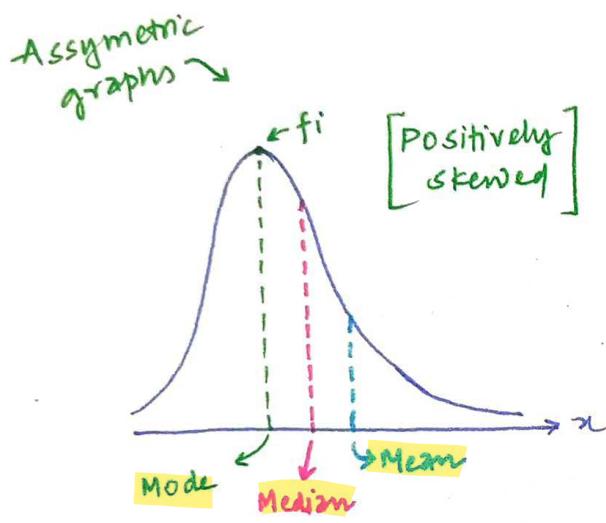
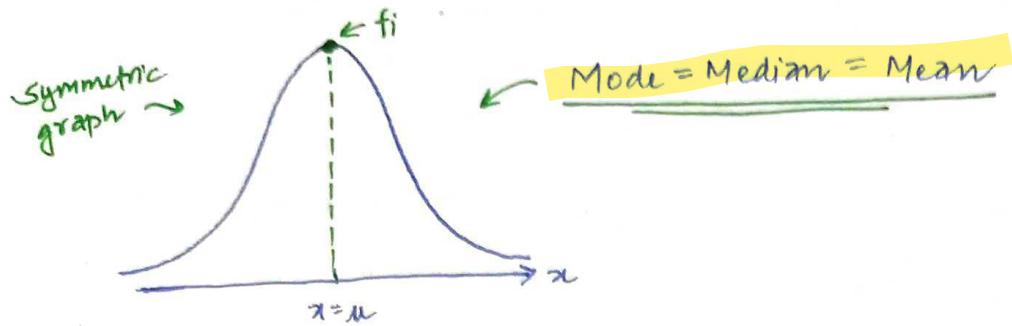
(M2) for Median wrong cumulative frequency method

S.NO	$x_i$	$f_i$	$Cf$
1	25	20	20
2	30	20	40
3	35	40	80
4	40	20	100

$$\text{Median} \rightarrow \frac{50^{\text{th}} + 51^{\text{th}}}{2}$$

$$= \frac{35 + 35}{2} = 35 \approx$$

## Normal distribution



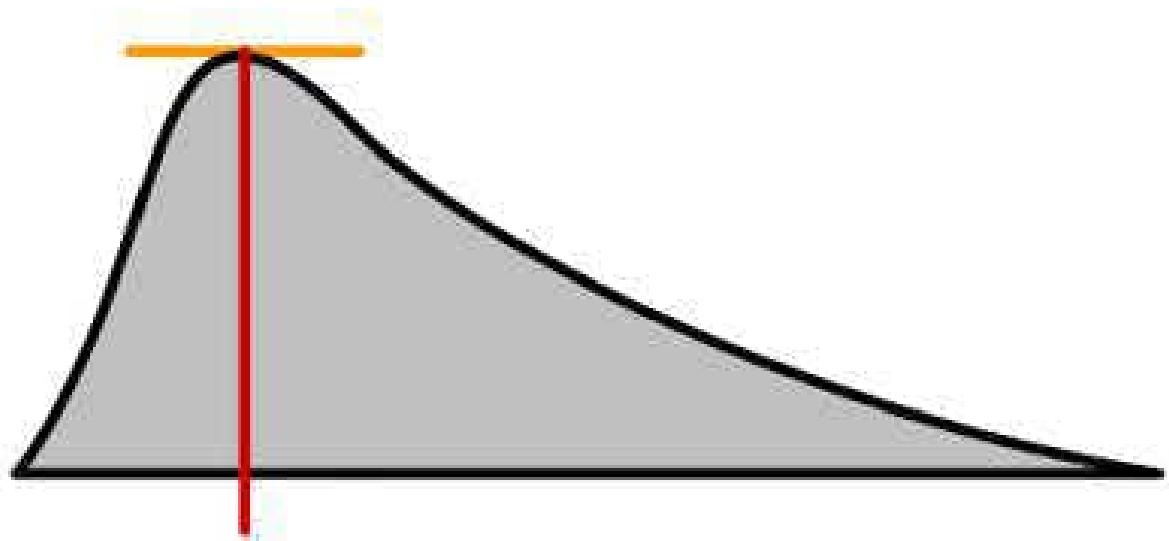
$$\text{Mode} \leq \text{Median} \leq \text{Mean}$$

$$\text{Mode} \geq \text{Median} \geq \text{Mean}$$

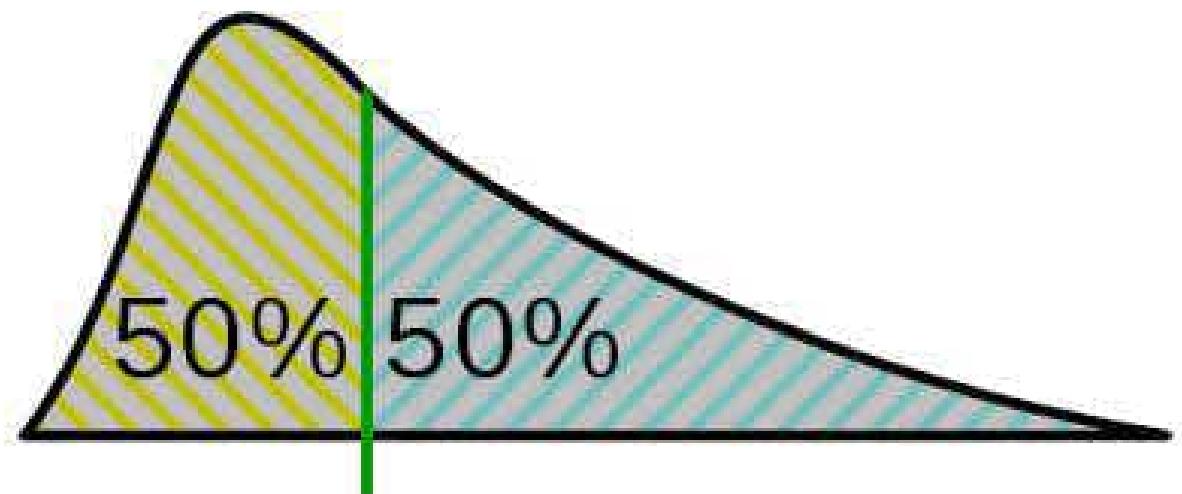
$$\begin{aligned} \text{Mean} &= \text{Median} \\ 1 \text{ Mode} + 2 \text{ Mean} &= 3 \text{ Median} \end{aligned}$$

Negatively  
skewed.

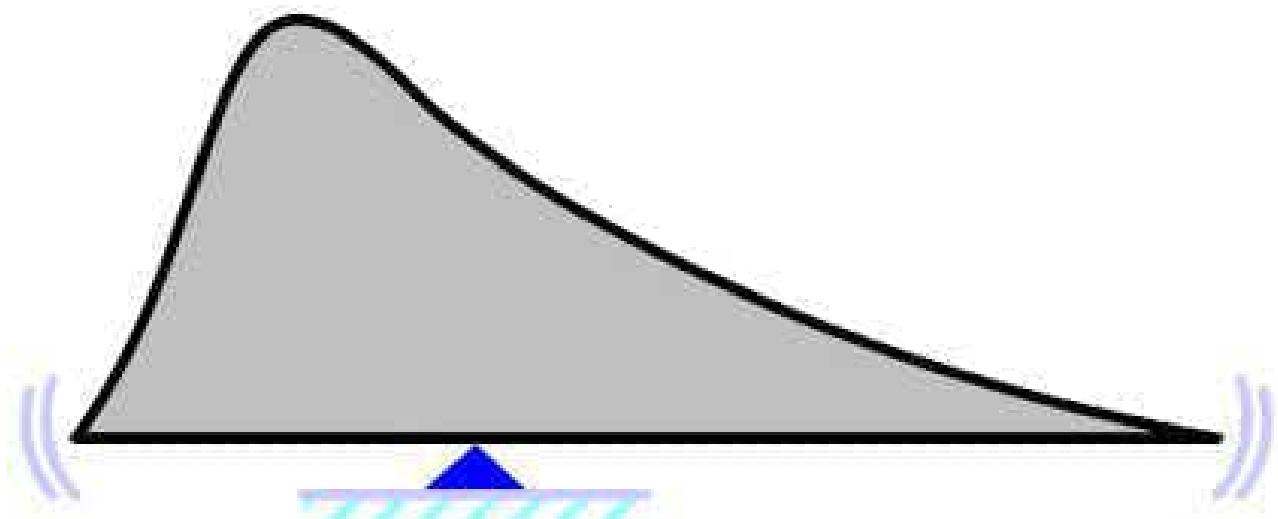
Positively  
skewed.



mode



median



mean

## Variance & Std Deviation of a dataset

$$x = \{2, 3, 4, 7, 9\}$$

Method 1:

x	$\mu$	$x-\mu$	$(x-\mu)^2$
2	5	-3	9
3	5	-2	4
4	5	-1	1
7	5	2	4
9	5	4	16

$$\boxed{\text{Variance}(\sigma^2) = \frac{\sum(x-\mu)^2}{n}}$$

$$\boxed{\text{Std Dev. } \sigma = \sqrt{\text{Variance}}}$$

$$= \sqrt{\frac{\sum(x-\mu)^2}{n}}$$

$$\hookrightarrow \mu = \frac{2+3+4+7+9}{5} = 5$$

$$\hookrightarrow \sum(x-\mu)^2 = 9+4+1+4+16 = 34$$

$$\Rightarrow \text{Variance}, \sigma^2 = \frac{\sum(x-\mu)^2}{n} = \frac{34}{5} = 6.8$$

$$\Rightarrow \text{std dev. } \sigma = \sqrt{6.8}$$

Method 2:

$$\text{Variance} = E[x^2] - (\bar{x})^2 \quad \text{Mean, } \mu = \frac{\sum x_i}{n} = \frac{\sum x_i f_i}{\sum f_i}$$

$$\boxed{\text{Variance} = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

$$\boxed{= \frac{\sum x_i^2 f_i}{\sum f_i} - \left(\frac{\sum x_i f_i}{\sum f_i}\right)^2}$$

x	$x^2$
2	4
3	9
4	16
7	49
9	81

$$\Rightarrow V = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2$$

$$V = \frac{159}{5} - \left(\frac{25}{5}\right)^2$$

$$\sigma^2 \rightarrow V = 6.8$$

$$\sum x_i = 25 \quad \sum x_i^2 = 159$$

$$n = 5$$

$$\text{std dev, } \sigma = \sqrt{V} = \sqrt{6.8}$$

$$\rightarrow = 2.607$$

Ques which data set has more VARIATION among these two,

(1) [1, 2, 3]

(2) [101, 102, 103]

Set 1

⇒ Coefficient of Variation =  $\frac{\sigma}{\mu}$

data Set 1

$x_i$	$x_i^2$
1	1
2	4
3	9

$$\sum x_i = 6 \quad \sum x_i^2 = 14 \quad n = 3$$

$$\text{Variance, } \sigma_1^2 = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2$$

$$= \frac{14}{3} - \left(\frac{6}{3}\right)^2 = \frac{14}{3} - 4 = \frac{2}{3}$$

$$\sigma_1 = \sqrt{\frac{2}{3}}$$

$$\mu_1 = \frac{\sum x_i}{n} = \frac{6}{3} = 2$$

$$\mu_1 = 2$$

Now, coefficient of variations,

$$\frac{\sigma_1}{\mu_1} = \frac{\sqrt{2/3}}{2}$$

$$\left(\frac{\sigma_1}{\mu_1}\right) > \left(\frac{\sigma_2}{\mu_2}\right)$$

coefficient of variation of data set 1 > data set 2

$$\frac{\sigma_1}{\mu_1} = \frac{\sqrt{2/3}}{102}$$

data Set 2

$x_i$	$\mu$	$(x_i - \mu)^2$
101	102	1
102	102	0
103	102	1

$n = 3$

$$\text{Variance, } \sigma_2^2 = \frac{\sum (x_i - \mu)^2}{n} = \frac{2}{3}$$

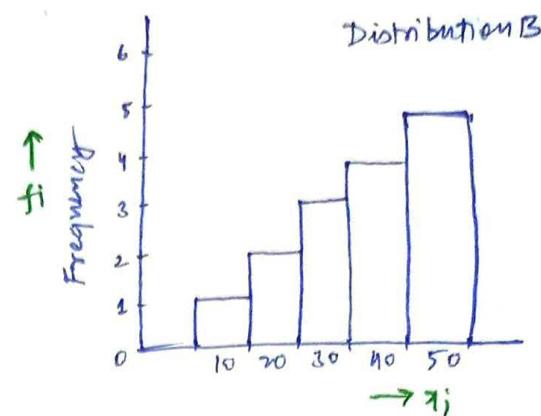
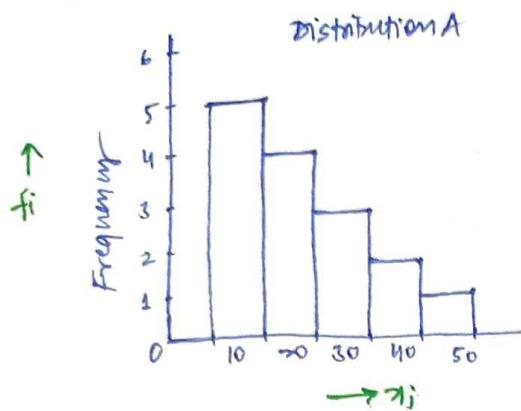
$$\sigma_2 = \sqrt{\frac{2}{3}}$$

$$\mu_2 = \frac{\sum x_i}{n} = \frac{306}{3} = 102$$

$$\mu_2 = 102$$

⇒ Data set 1 has more Variation

Ques. For the distribution given below:



Which of the following is correct for the above distributions?

- (a) Std dev of A is significantly lower than std dev of B.
- (b) Std dev of A is slightly lower than std dev of B.
- (c) Std dev of A is same as std dev of B.
- (d) Std dev of A is significantly higher than std dev of B.

Soln'

Distribution A

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
10	5	50	500
20	4	80	1600
30	3	90	2700
40	2	80	3200
50	1	50	2500
$\Sigma \Rightarrow$		15	10500
		350	

Distribution B

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
10	1	10	100
20	2	40	800
30	3	90	2700
40	4	160	6400
50	5	250	12500
$\Sigma \Rightarrow$		15	22500
		550	

$$\sigma_1 = \sqrt{\frac{\sum x_i^2 f_i}{n} - \left(\frac{\sum x_i f_i}{n}\right)^2}$$

$$n = \sum f_i = 15$$

$$\sigma_2 = \sqrt{\frac{\sum x_i^2 f_i}{n} - \left(\frac{\sum x_i f_i}{n}\right)^2}$$

$$n = \sum f_i = 15$$

$$\sigma_1 = \sqrt{\frac{10500}{15} - \left(\frac{350}{15}\right)^2} = 12.472$$

$$\Rightarrow \boxed{\sigma_1 = \sigma_2}$$

option (c) ✓

$$\sigma_2 = \sqrt{\frac{22500}{15} - \left(\frac{550}{15}\right)^2} = 12.472$$

Ques. Let  $X$  be a normal RV with mean 1 & variance 4. The prob  $P[X \leq 0]$  is

(A) 0.5

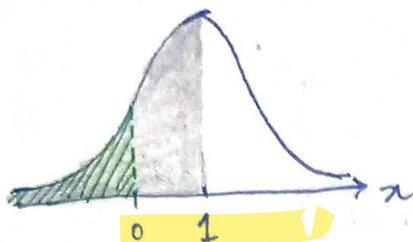
(B) greater than zero  
& less than 0.5

(C) greater than 0.5

& less than 1.0

(D) 1.0

Sol<sup>n</sup>



$$\text{If } \mu = 1 \Rightarrow P[X \leq 1] = 0.5$$

$$\text{We want } P[X \leq 0] \Rightarrow 0 < P[X \leq 0] < P[X \leq 1]$$

$$0 < P[X \leq 0] < 0.5$$

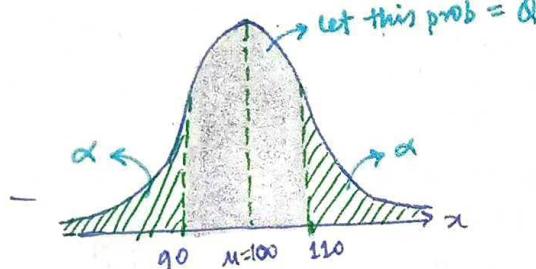
both prob is  
always +ve

Ques. For a R.V.  $X$  ( $-\infty < x < \infty$ ), following normal distribution, the mean is  $\mu = 100$ .

If the prob is  $P=\alpha$  for  $x \geq 110$ , then the prob of  $X$  lying between 90 & 110, i.e.,  $P(90 \leq x \leq 110)$  will be equal to.

- (A)  $1-2\alpha$       (B)  $1-\alpha$       (C)  $1-\frac{\alpha}{2}$       (D)  $2\alpha$

Sol<sup>n</sup>



$$1 = \alpha + \alpha + \alpha$$

$$1 = 2\alpha + \alpha$$

$$\alpha = 1-2\alpha$$

Ques. Let  $U$  &  $V$  be 2 independent zero mean Gaussian RV of variances  $\frac{1}{4}$  &  $\frac{1}{9}$  respectively. The prob  $P[3V \geq 2U]$  is

(A)  $4/9$

(B)  $1/2$

(C)  $2/3$       (D)  $5/9$

Sol<sup>n</sup>  $P[3V \geq 2U] = P[3V - 2U \geq 0]$

Let a normal RV,  $Y = 3V - 2U$

$$\Rightarrow P[Y \geq 0]$$

Convert from NOR. dis. to Z-dis.

$$= P\left[\frac{Y-\mu}{\sigma} \geq \frac{0-\mu}{\sigma}\right] \quad \begin{matrix} \text{this is} \\ \text{mean of 'y'} \\ \text{not 'V' or 'U'.} \end{matrix}$$

$$= P[Z \geq \frac{-\mu}{\sigma}]$$

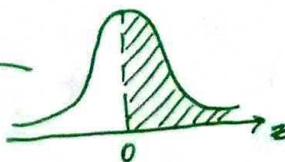
$$= P[Z \geq \frac{0}{\sigma}] = P[Z \geq 0] = \boxed{1/2}$$

$$Y = 3V - 2U$$

$$\begin{aligned} \mu &= E[Y] = E[3V - 2U] \\ &= E[3V] - E[2U] \\ &= 3E[V] - 2E[U] \end{aligned}$$

$$\mu = E[Y] = 0 \quad \therefore$$

-----



Given  
 $V, U$  are  
zero mean  
RV



Ab hum wo **school/college wale darpok**  
nahi rahe Ab hmko chahiye full **marks**

# END OF THE END

