PROBABILITY

12.1 WHY STUDY PROBABILITY

- Used extensively in multiple areas by CSE, EE, Mech, Civil engineers, also has extensive applications in Physics and other sciences.

- Software :- Quicksort → $O(n^2)$ worst case time complexity.
  On randomly choosing the pivot we can get $O(n \log n)$ time complexity.

- There are probabilistic Data Structures and algorithms, we use these Data Structures and Algorithms to use to improve the time complexity of the algorithm.

- Machine Learning, Data Science, Artificial Intelligence :-
  A Lot of algorithms based on probability → Bayes Theorem → Naive Bayes
  ↓
  Bayesian Inference.

- Build / Simulate Especially at defence organizations such as
  DRDO / ISR / CERN / BARR.
  - We try to simulate real world systems like Internet traffic, Road traffic etc.

12.2 INTRODUCTION

— Two ways to introduce Probability

1. Set Theoretic (Axiomatic). — intuitive, Venn diagrams, Combinatorics.
2. Measure Theoretic · More Mathematical).

Sample Space (S)

Experiment : Flipping 2 coins (distinct).

Outcomes = {(H,H), (H,T), (T,H), (T,T)} = S.

Sample Space :- Set of all possible outcomes of the experiment

Experiment :- Conduct a 7-way horse race.

Outcomes : ordering of 7 horses.

2,1,3,6,5,4,7.

S = {7! possible orderings}.

Experiment :- 2 distinct dice are rolled,

Outcomes S = {(1,1) (1,2) (1,3), (1,4), (1,5) (1,6)
(2,1) (2,2) - - - - (2,6)
(3,1) (3,2) - - - - (3,6)
⋮
(6,1) - - - (6,6)}.

36 total outcomes.

Experiment :- light Bulb, Measuring the # hrs the light bulb works

Outcomes $S = \{ x \;\; 0 \le x < \infty \}$

$x \in \mathbb{R}$

here the sample space is of infinite size.

Event (E) : Any subset of S is an event.

(eg) $E = \{ x : 0 \le x \le 6 \}$ ⇒ light bulb works for 0 to 6 hours.

$E \subseteq S$

(eg) If two die are rolled. Event that sum = 6.

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

(eg) E : first horse wins out of the 7 horses.

1 — — — — — —

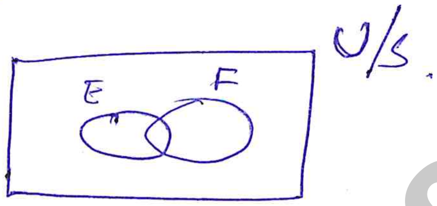
E = { 6! possible orderings where 1 is the first }

(eg) E = 1st coin is always H.

$E = \{ (H,T), (H,H) \}$.

12.3 Axioms Of Probability, Properties And Examples

- For a given experiment
  - S: The sample space is the set of all possible outcomes similar to universal set in the set theory.
  - E: Event :- Set of outcomes



eg :- Throwing / Tossing 2 coins (distinct)

$S = \{(H,H),(H,T),(T,H),(T,T)\}$

E = All outcomes where first coin is H.

$E = \{(H,H),(H,T)\}$

$F = \{(H,H),(T,T)\}$ all outcomes with both coins of the same value.

$E \cap F = \{(H,H)\}$

$E \cup F = \{(H,H),(T,T),(H,T)\}$

$E^c = \{(T,T),(T,H)\}$

Frequentist Vs Bayesian way of thinking about probability.

- The Bayesian approach is more related to the Bayes theorem
- The frequentist approach.

1. Conduct the experiment n times
2. n(E) = no of times the outcome of the experiment ∈ E

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

eg fair coin $\to$ H, $\to$ T

$S = \{H, T\}$ exp: Toss of a fair coin.

$E = \{H\}$

$P(H) = ?$

10 times $\to$ 6 H, $\to$ 4 T $\quad P(H) = \frac{6}{10}$

1000 $\to$ 506 H, $\to$ 494 T $\quad P(H) = \frac{506}{1000}$

⋮ $\quad$ ⋮

∞ $\quad$ 0.5

Axiomatic Approach (self evident)

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

2. $P(S) = 1$

3. if $E_1, E_2, E_3 \cdot E_n$ are mutually exclusive events $E_i \cap E_j = \emptyset$

$$P(E_1 \cup E_2 \cup \ldots E_n) = \sum_{\forall i} P(E_i)$$

example. A fair die is rolled.

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

$E_1 = \{2\}$ $\quad P(E_1) = 1/6$

$E_2 = \{4\}$ $\quad P(E_2) = 1/6$

$E_3 = \{6\}$ $\quad P(E_3) = 1/6$.

$P(E_1 \cup E_2 \cup E_3) = 1/6 + 1/6 + 1/6$

$= \underline{1/2}$

4. $P(E^c) = 1 - P(E)$

5. If $E \subseteq F$ (if E is a subset of F) then $P(E) \le P(F)$

6. $P(E \cup F) = P(E) + P(F) - P(E \cap F)$ (can be proved using principle of inclusion & exclusion).

7. $P(E_1 \cup E_2 \cup E_3 \ldots \cup E_n) = P(E_1) + P(E_2) \ldots + P(E_n)$

$- (P(E_1 \cap E_2) + P(E_2 \cap E_3) + \ldots P(E_n \cap E_n))$ 2 way

$+ (P(E_1 \cap E_2 \cap E_3) + P(E_2 E_3 E_4) + \ldots)$ 3 way

$\vdots$

$\ldots$ n way.

- If we have a sample space with equally likely outcomes then.

$$P(E) = \frac{|E|}{|S|}$$

for example if we have a fair die $S = \{1,2,3,4,5,6\}$.

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}).$$

$$P(S) = 1 = P(\{1,2,3,4,5,6\}) = \sum_i P(\{i\}) = 1$$

$$\therefore P(\{i\}) = 1/6.$$

→ example: If 2 dice are thrown.

E = sum of dice $= 7$.

$|S| = 36$ $(1,1)\ (1,2) \ldots (6,6)$

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

$$P(E) = \frac{6}{36} = \frac{1}{6}.$$

→ example if we have 6 white and 5 Black balls, if we randomly pick 3 balls what is the probability that we get 1 white and 2 black balls

No of ways we can choose 1 white and 5 black balls $= {}^6C_1 \times {}^5C_2$

Total no of ways we can choose 3 from 11 balls $= {}^{11}C_3$

$$\therefore P(E) = \frac{{}^6C_1 \times {}^5C_2}{{}^{11}C_3}$$

(eg) If we have n balls and one ball is special among them

experiment :- K balls are drawn one at a time randomly.

Event : Special ball is picked.

$$|S| = {}^{n}C_{K}.$$

$$|E| = {}^{1}C_{1} \cdot {}^{n-1}C_{K-1}$$

Matching Problem

eg → If we have N men who throw/keep their hats on the table and then pick the hat from the table randomly. the probability that all get a hat other than their hat.

E_i = Event that the ith person has picked the correct hat.

$P(E_1 \cup E_2 \cup E_3 \cdots E_n)$ = Probability that atleast one person picks his hat

$1 - P(E_1 \cup E_2 \cdots \cup E_n)$ = Prob. that no person pick his hat.

$$= \underbrace{P(E_1)}_{1/n} + \underbrace{P(E_2)}_{1/n} + \cdots P(E_n) = \frac{n}{n} = 1$$

$$- \big(P(E_1 \cap E_2) + P(E_2 \cap E_3) + \cdots\big) - 2 \text{ way} \quad \left(\tfrac{1}{n} \cdot \tfrac{1}{n-1}\right) {}^{n}C_2 = \frac{1}{2!}$$

$$+ \big(P(E_1 \cap E_2 \cap E_3) + \cdots\big) - 3 \text{ way}. \quad -\frac{1}{3!}$$

$-$

$\vdots$

n way

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} \cdots \frac{1}{n!}$$

$$1 - P\left(\bigcup_{i=1}^{n} E_i\right) = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots (-1)^n \frac{1}{n}.$$

If we have infinitely many people the expression convert.

$$= 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots \frac{1}{\infty}.$$

$$= e^{-n} = 0.36788.$$

12.4 Conditional Probability And Examples

Example:- Two dice are rolled.

$S = \{(1,1)\ (1,2) \ldots (6,6)\}$

F = { outcome of the first die 3 = $\{(3,1), (3,2), (3,3), (3,4)$
$(3,5); (3,6)\}$

E = Sum of two dice = 7 = $\{(1,6), (2,5), (3,4), (4,3), (5,2),$
$(6,1)\}$

Probability that event E has occurred given that event F has already occurred.

$P(E/F) = ?$

$$\boxed{P(E|F) = \frac{P(E\cap F)}{P(F)}} = \frac{1/36}{1/6} = 1/6$$

if $P(F) \neq 0$.

(Q) Student taking one hour exam, P.(Student finishes the exam under x hrs) = $\frac{x}{2}$

Given that the student is working at 0.75 hours what is the probability that the student the full 1 hours.

$0 < x \leq 1$

F = Student uses the full one hr

F^c = Student finishes exam under 1 hour.

$P(F^c) = 1/2$ $\quad F \cup F^c = S$

$$P(F) = 1 - 1/2 = 1/2$$

L_x = Student finishes in x hours.

L_x^c = Student is still working at x hours.

$L^c_{0.75}$.

$$P(F|L^c_{0.75}) = \frac{P(F\cap L^c_{0.75})}{P(L^c_{0.75})} = \frac{P(F)}{1-P(L_{0.75})}$$

$$= \frac{1/2}{1-\frac{0.75}{2}} = \underline{\underline{0.8}}.$$

12.5 MULTIPLICATION THEOREM

- example on conditional probability.
- Select n balls sequentially and randomly without replacement the urn contains r red balls, b blue balls $n \le r+b$, Given that k out of n balls are blue, what's the probability that the 1st ball is blue.

Lets take Event B = Event that 1st picked ball is blue.

B_k = Event that k out of n balls picked are blue.

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

$$P(B_k) = \frac{{}^{b}C_{k} \times {}^{r}C_{n-k}}{{}^{b+r}C_{n}} = \frac{n(B_k)}{n(S)}$$

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

$$P(B) = \frac{b}{r+b}.$$

$$P(B_k|B) = \frac{{}^{b-1}C_{k-1} \cdot {}^{r}C_{n-k}}{{}^{r+b-1}C_{n-1}}$$

$$\therefore P(B|B_k) = \frac{\left(\dfrac{{}^{b-1}C_{k-1} \cdot {}^{r}C_{n-k}}{{}^{r+b-1}C_{n-1}}\right)\dfrac{b}{r+b}.}{\left(\dfrac{{}^{b}C_{k} \cdot {}^{r}C_{n-k}}{{}^{b+r}C_{n}}\right)}$$

$$= \left(\frac{k}{n}\right).$$

Multiplication Rule.

We have $P(E_1|E_2) = \dfrac{P(E_1 \cap E_2)}{P(E_2)}$ if $P(E_2) \neq 0$

$$\Rightarrow P(E_1 \cap E_2) = P(E_1|E_2) \cdot P(E_2)$$

can also be written as

$$P(E_1E_2) = P(E_1|E_2)P(E_2)$$

Generalization

$$P(E_1E_2E_3 \cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_2E_1)P(E_4|E_3E_2E_1) \cdots P(E_n|E_{n-1}E_{n-2} \cdots E_1)$$

(example) Matching problem. n people throw their hats and randomly pick back their hats

$$P(\text{no one picks their hat back}) = \sum_{i=0}^{n} (-1)^i / i!$$

Q. P(Exactly k persons have picked correctly) = ?

Let $A = \{$k persons would pick correctly$\}$ - can be done in ^nC_k ways.

E = Everyone in A has picked correctly.

G = Everyone other than people in set A have picked.

We need to calculate $P(G \cap E) = P(G \cap E) = P(G|E)P(E)$

$P(G|E) =$ - Probability that $n-k$ people have not picked the correct hat.

$$= \sum_{i=0}^{(n-k)} \frac{(-1)^i}{i!}$$

$P(E) =$ F_1 = event that 1st person in A has picked the hat correctly.

F_2 = event that 2nd person in set A has picked the hat correctly.

F_3 - ,, 3rd ,, .

$$P(E) = P(F_1 F_2 \cdots F_k) = P(F_1)P(F_2|F_1)P(F_3|F_2F_1)P(F_4|F_3F_2F_1) \cdots P(F_k|F_{k-1} \cdots F_1)$$

$$= \frac{1}{(n)} \cdot \frac{1}{(n-1)} \cdot \frac{1}{(n-2)} \cdots \frac{1}{(n-k+1)}$$

$$\frac{(n-k)!}{n!}$$

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

$$= \left[\sum_{i=0}^{(n-k)} \left(\frac{(-1)^i}{i!}\right)\right]\left[\frac{(n-k)!}{n!}\right]$$ × no of ways we can form set A $({}^nC_k)$

12.6 Independent Events

- Example

Experiment: Toss a coin & throw a dice.

E = the coin is H

F : The dice is 3.

$P(E \cap F) = P(E/F)P(F)$. – definition of conditional probability.

↓

The outcome of H given that the outcome of the dice is 3. the outcome of the coin is not dependent on the outcomes of the dice therefore we can replace $P(E|F)$ by $P(E)$ itself

$\Rightarrow P(E \cap F) = P(E)P(F)$.

Such events where outcome of one does not impact the outcome of other are known as Independent Events.

Note :- Independent Events and Mutually Exclusive events are not the same, independence of events means that the outcome of one of the event does not have any impact on the outcome of the other event, whereas if two or more events are mutually exclusive then the occurance of one prevents the occurance of the other event or events.

Independent Events $P(E \cap F) = P(E) \cdot P(F)$.

Mutually Exclusive Events $P(E \cap F) = 0$ as $E \cap F = \emptyset$

example. A Deck of 52 cards is available.

experiment :- A card is picked randomly and replaced.

The second card is also picked randomly.

$P(\underbrace{\text{the first card is jack}}_{E_1} \text{ and the } \underbrace{\text{second is 8}}_{E_2})$

→ E_1 and E_2 are independent events if we get first card as jack we do not have any impact on E_2 because it is replaced, hence these events are independent events.

→ If after the 1st card is picked it is not replaced then the prob of the second card being 8 is no longer same it differs and depends on what the outcome of the 1st event was and also the number of cards in the deck have decreased.

(example). 2 distinct coins are tossed

E : 1st coin is H

F : 2nd coin is T

$P(EF) = P(E)P(F) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

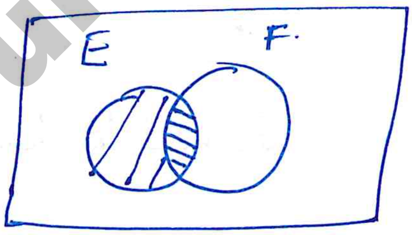
Generalization for Independent Events

$$P(E_1 E_2 E_3 \cdots E_n) = P(E_1) \cdot P(E_2) P(E_3) \cdots P(E_n)$$

$$= \prod_{i=1}^{n} P(E_i)$$

Note: If E and F are independent events then E & F^c are also independent Events.

Proof $P(E) = P(E \cap F) + P(E \cap F^c)$



$= P(E)P(F) + P(E \cap F^c)$

$P(E \cap F^c) = P(E)[1 - P(F)]$

$= P(E) \cdot P(F^c)$

Which means E and F^c are independent events.

(example) An infinite sequence of trails are performed.

success probability p.

Trail : Throwing a dice

success - outcome = 1

$p = 1/6$.

(a) Prob. of atleast one success in n independent trails.

$$= 1 - P(\text{no succes in } n \text{ trails}).$$

$$= 1 - P(F_1 F_2 \ldots F_n)$$

$$= 1 - P(F_1)P(F_2)\ldots P(F_n)$$

$$= 1 - (1-p)^n$$

(b) Prob. of exactly k successes in n trails.

- We need to have k sucesses and (n-k) failures.

$$= {}^nC_k\, p^k (1-p)^{n-k}.$$

Note: $S_1, S_2 \ldots S_n$ are all independent events

S_i = Success in the ith trail.

(c) Prob of all n trails are successful.

$$= \underline{\underline{p^n}}$$

12.7 Law Of Total Probability

(Example). Manufacturing

- We have two factories X & Y for manufacturing bulbs.
- Bulbs made in X work for 5000 hours in 99% cases.
- Bulbs made in Y work for 5000 hours in 95% cases.

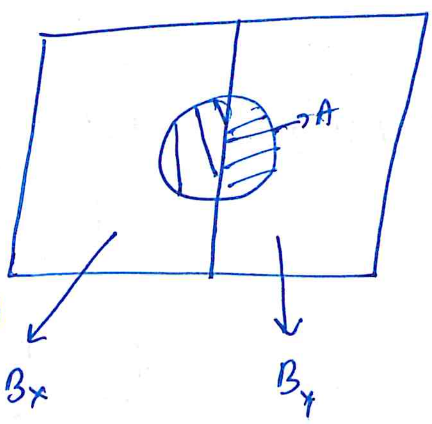
- We create a package which have 60% bulbs from X and 40% of the bulbs from Y, now what is the probability that the bulb will work for 5000 hours?

Let A: be the event that the bulb works for 5000 hours.

B_X: Event that a bulb is made at X

B_Y: Event that a bulb is made at Y.

$$P(A) = P(A \cap B_X) + P(A \cap B_Y)$$



$$= P(B_X) \cdot P(A|B_X) + P(B_Y) P(A|B_Y)$$

$$= (0.6 \times 0.99) + (0.4 \times 0.95)$$

$$= \underline{0.974}$$

$S = B_X \cup B_Y$

$B_X \cap B_Y = \phi$

A Generalization of above example is as follows

If $B_1\ B_2\ B_3\ ---\ B_n$ are events such that.

$B_i \cap B_j = \phi$ Mutually Exclusive.

$B_1 \cup B_2 \ldots \cup B_n = S$ — Mutually Exhaustive Events.

Then $P(A) = P(A \cap B_1) + P(A \cap B_2) \ldots\ldots + P(A \cap B_n)$.

$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots\ldots P(A|B_n)P(B_n)$

$$\boxed{P(A) = \sum_{i=1}^{n} P(A|B_i)\,P(B_i)}$$

12.8 Bayes Theorem

From defination of conditional probability we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) \neq 0$$

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

$$\boxed{P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \quad \text{if } P(B) \neq 0}$$

$P(B)$ and $P(A)$ are known as marginal probability.

$P(A|B)$ and $P(B|A)$ are the conditional probabilities.

Alternate way of writing Bayes theorem :

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

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

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

Applications of Bayes theorem

- Medical Diagonosis
  - Detection of Breast Cancer.

Approx 1% of women in 40-50 have breast cancer.

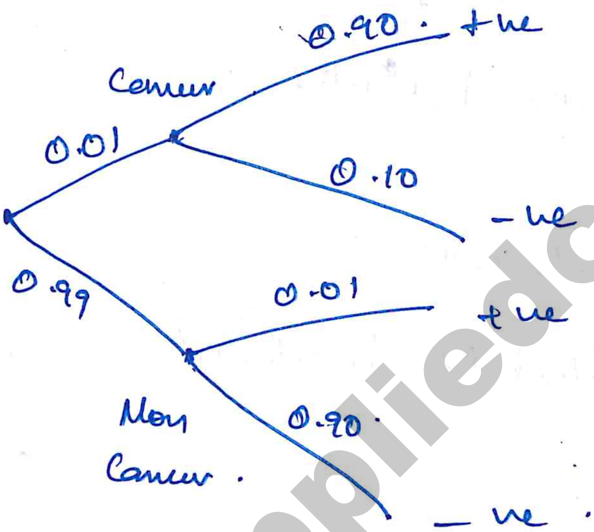
Mammogram (X-ray) → cheapest (not perfect)

- If a woman has breast cancer, the test will result in true value 90% of the time.
- If a woman does not have breast cancer, then the test will result true → 10% of the times.

If the doctor wants to predict, what is the prob that the women has cancer given the result is +ve

$$P(\text{cancer} \mid +ve) = \frac{P(+ve \mid \text{cancer})\, P(\text{cancer})}{P(+ve)}$$

$$= \frac{0.9 \times 0.01}{P(+ve)}$$



$$P(+ve) = P(\text{Cancer} \cap +ve) + P(\text{Noncancer} \cap +ve)$$

$$= P(\text{cancer})\, P(+ve \mid \text{cancer}) + P(\text{Noncancer})\, P(+ve \mid \text{Noncancer})$$

$$= (0.9)(0.01) + (0.1)(0.99)$$

$$= 0.108$$

$$P(\text{cancer} \mid +ve) = \frac{0.9 \times 0.01}{0.108} = \frac{9}{108} = \underline{\underline{8.3\%}}.$$

Actuall probability = 8.3% people learn age and get tensed but it is not fair.

In such a case the doctor will take a more advanced test.

— <u>Odds</u> :- Mostly used in betting.

$$\text{Odds of } A \text{ happening} = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

The odds in favor of A are 2 to 1 $\rightarrow \dfrac{P(A)}{P(A^c)} = \dfrac{2}{1}$

12.9 Solved Problems.

A random bit string of length n is constructed by tossing a fair coin n times and setting a bit to 0 or 1 depending on outcomes head and tail, respectively. The probability that two such randomly generated strings are not identical is.

A. $\dfrac{1}{2^n}$

B. $1 - \dfrac{1}{n}$

C. $\dfrac{1}{n!}$

D. $1 - \dfrac{1}{2^n}$

_ _ _ _ . . . _ S_1

1 2 3 4 n

_ _ _ _ - - - _ S_2.

Prob that the two strings are different = 1 − Prob that both are same.

$= 1 - \left(\frac{1}{2}\cdot\frac{1}{2}\cdot \ldots n \text{ times}\right)$

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

Q → For each element in a set of size 2n, an unbiased coin is tossed. The 2n coin tosses

A. $\frac{^{2n}C_n}{4^n}$

B. $\frac{^{2n}C_n}{2^n}$

C. $\frac{1}{^{2n}C_n}$

D. $\frac{1}{2}$.

Solution.

Given that we have a set of the 2n elements $\{a_1, a_2, \dots a_{2n}\}$

We can choose n out of 2n coins in ${}^{2n}C_n$ ways.

Getting heads on n coins we need to have $(1/2) \times (1/2) \dots$ n times
$= (1/2)^n$ probability.

and

getting tails on n coins we need to have $= (1/2) \times (1/2) \dots$ n times
$= (1/2)^n$ probability.

$\therefore$ Total probability $= {}^{2n}C_n \times \left(\frac{1}{2}\right)^n \times \left(\frac{1}{2}\right)^n$

$= {}^{2n}C_n \times \frac{1}{4^n}$ (option A).

Q) Let A and B be two arbitrary events, then, which of the following is true?

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

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

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

D. $P(A \cup B) \leq P(A) + P(B)$.

Solution

A. $P(A \cap B) = P(A)P(B)$ – it is true only when the 2 events are independent.

×

B. $P(A \cup B) = P(A) + P(B)$ it is only true when $P(A \cap B) = 0$

×

C. $P(A|B) = P(A \cap B) \cdot P(B)$ – From conditional probability we know that

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

×

D. $P(A \cup B) \le P(A) + P(B)$

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

$\therefore$ D is true.

Q → Let $P(E)$ denote the probability of the event E. Given $P(A) = 1$, $P(B) = 1/2$, the values of $P(A|B)$ and $P(B|A)$ respectively are?

A. $\left(\frac{1}{4}\right), \left(\frac{1}{2}\right)$.

B. $\left(\frac{1}{2}\right), \left(\frac{1}{4}\right)$.

C. $\left(\frac{1}{2}\right), 1$.

D $1, \left(\frac{1}{2}\right)$.

Soln

$P(A|B)$ = Prob of A given B occures.

= $P(A)=1$ $P(B)=1/2$

Will occour always therefore irrespective of occourance of B

A will occour. ∴ $P(A|B)=1$

Also by using the defn of conditional probability

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

As $P(A)=1$ it is like the complete sample space, therefore.

$A \cap B = B$

$$P(B|A)= \frac{P(A\cap B)}{P(A)} = \frac{P(B)}{P(A)} = \frac{1/2}{1} = 1/2.$$

Option D is correct.

Q → Box P has 2 red balls and 3 blue balls and box Q has 3 red balls and 1 blue balls. A ball is selected as follows (i) Select a box (ii) Choose a ball from the selected box such that each ball in the box is equally likely to be choosen. The probabilities of selecting boxes P and Q are 1/3 and 2/3 respectively. Given that a ball selected in the above process is a red ball, the probability that it came from the box P is :-

A. $4/19$ B. $5/19$ C. $2/9$ D. $19/13$.

Solution

P: 2r, 3b

Q: 3r, 1b

$P(P) = 1/3$ $P(Q) = 2/3$

$$P(P \mid red) = \frac{P(P \cap red)}{P(red)} = \frac{\frac{1}{3} \cdot \frac{2}{5}}{P(red)} \Rightarrow \frac{2/15}{19/30} \Rightarrow \frac{4}{19}$$

$$P(red) = \frac{1}{3} \cdot \frac{2}{5} + \frac{2}{3} \cdot \frac{3}{4}$$

12.10 Random Variable : An Introduction

- A random variable is a mapping from an event to $\mathbb{R}$

example :- Toss 3 distinct coins || $8 = 2^3$ outcomes $= |S|$.

Random variable X - denotes the number of heads

$P(X=0) = P(\text{All tails}) = \frac{1}{2^3}$

$$P(x=2) = {}^3C_2 \frac{1}{2^2} \cdot \frac{1}{2}$$

In the above example the random variable x can take only values $= \{0,1,2,3\}$, example of one form of random variables known as discrete random variables.

(2) X = amount of rainfall on a given day.

$P(x \geq 2\text{cm}) = 0.95$

$P(x \leq 1\text{cm}) = 0.99$

This information is very useful for many people like farmers, people residing in that area.

hence $X \in [0, \infty)$

(3) X = height of students.

$P(x \geq 180\text{ cm}) = 1\%$

Useful for designing costumes / T-shirt for a public event.

designing benches / chairs.

(4) X = time spent on a website

let $P(x \geq 10\text{ min}) = 80\%$

A very good sign

let $P(x \leq 1\text{ min}) = 90\%$

We can guess that there is something wrong on our website.

(5) $X = \#$ visitors to a website on a given date.

$$P(x \geq 1000) = 0.1\%$$

$$\frac{1}{1000} = \frac{1}{3\,\text{years}}$$ once in 3 years.

(6) $X = \#$ children in a family

$P(x \leq 2) = 95\%$ India (population control)

$P(x = 0) = 90\%$ Japan (New policies can be decided by governments).

- These example here describes a discrete random variable.

Discrete & Continous Random Variables

- If # of possible values the random variable can take is countable then the R.V is discrete

- If the possible values the random variable can take are uncountable then the R.V. is continous

12.11 PMF, CDF AND PDF OF RANDOM VARIABLES

Probability Mass Function

- Discrete random variable

X = # children in a family.

$X = \{0, 1, 2, 3, 4 \dots 6\}$

- If a survey is done and we have data for large amount of families.

H_1 — 0

H_2 — 2

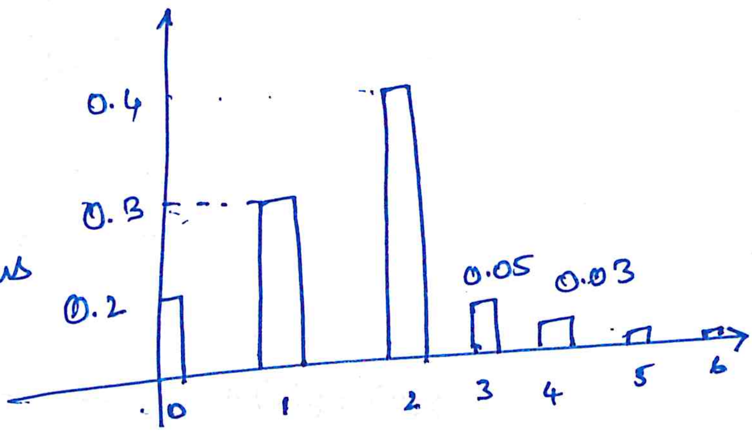
H_3 — 1

. . .

. . .

. . .

A graph can be plotted using this data, known as Histogram.



Many of the important questions can be answered by looking at this graph.

For ex $P(x=4) = ?$ $= 0.03$

$P(x <= 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3) = 95\%$

- The probability mass function is a function which mapps from a value of a random variable to its probability.

en $P(X=a)$ to $\rightarrow p(a)$

example.

X is a discrete random variable.

$X \in \{0, 1, 2, 3 \dots\}$

PMF $P(X=i) = p(i) = c\frac{\lambda^0}{i!}$ for some λ: +ve value

(a) $P(X=0) = p(0) = \frac{c\lambda^0}{0!} = c$.

(b) $P(X \leq 2) = P(X=0) + P(X=1) + p(X=2)$

$$= \frac{c\lambda^0}{0!} + \frac{c\lambda^1}{1!} + \frac{c\lambda^2}{2!}$$

$$= c + \lambda c + \frac{\lambda^2}{2}c$$

(c) $p(0) + p(1) + \dots = 1$

$$\sum_{i=0}^{\infty} c\frac{\lambda^i}{i!} = c\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = ce^{\lambda} = 1.$$

$$c = \frac{1}{e^{\lambda}} = e^{-\lambda}$$

Cumulative distribution function (CDF)

If x is a discrete R.V

$$x \in \{1, 2, 3 - - - \}$$

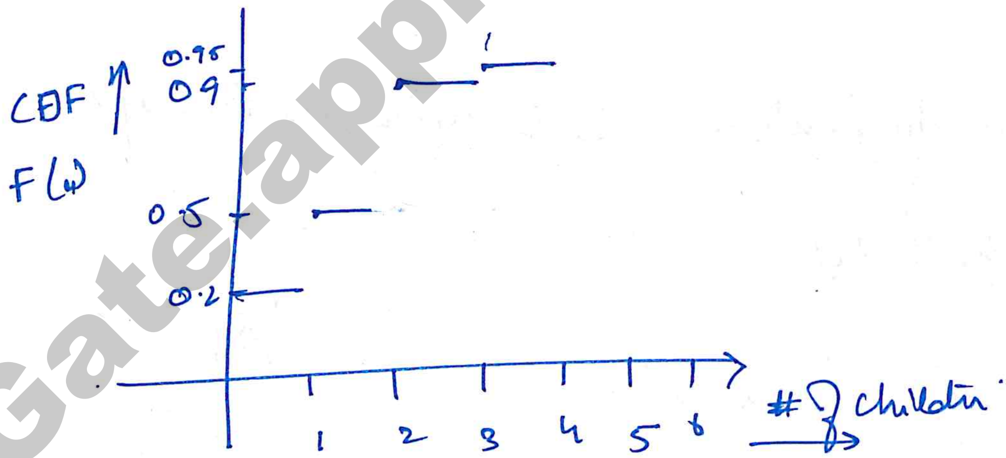
$$F_x(a) = F(a) = P(x=a) + P(x=a-1) + P(x=a-2) + \cdots P(x=0)$$

$$= P(x \le a)$$

$$= \sum_{x \le a} P(x = x) = \sum_{x \le a} p(x)$$

↑ r.v ↑ value

We can also plot the CDF

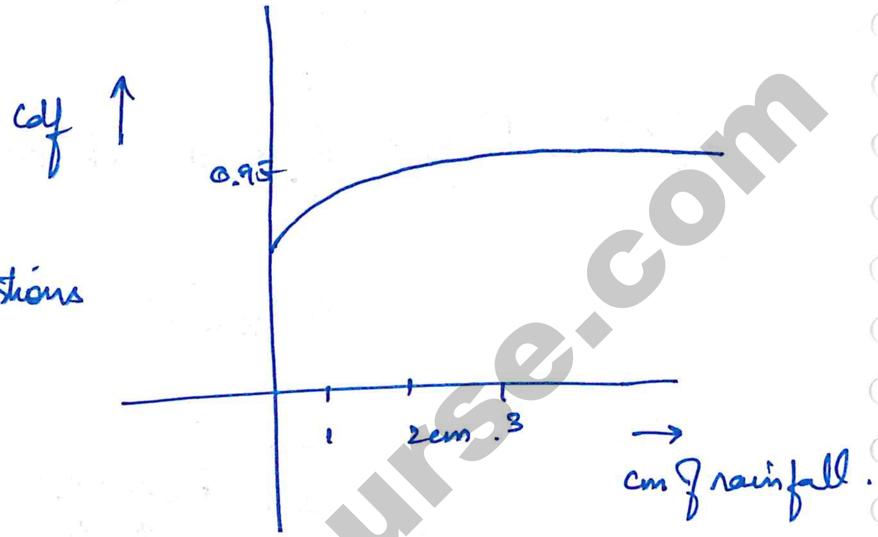


CDF for a continous random variable

X = rainfall on a particular day.

$F_X(a) = F(a) = P(X \le a)$

$X \in [0, \infty)$



We can answer various questions about the rainfall like

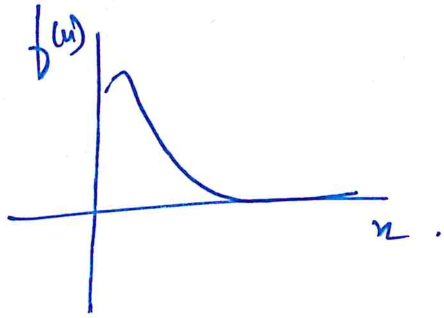
(a) $P(X \le 2\,cm) = 0.95$

(b) $P(1 < X \le 3\,cm) = P(X \le 3) - P(X \le 1)$

$= 0.96 - 0.90 = 6\%$.

Probability density function :- It is defined as the slope of the curve represented as the CDF.

$$f_X(a) = \left[\frac{d\, F_X(x)}{dx}\right]_{x=a}.$$



It is the slope of the curve at $x=a$.

→ In case of a continuous random variable it does not make sense to have PDF defined for a given point, because of the point actually depends on the accuracy by which we are measuring the point. for ex. $x = 1$ cm could also be 1.0000001 or 0.999999999 cm its not possible to measure it, therefore it is defined by differentiating the cdf.

12.12 Expectation

If x is a discrete random variable then Expectation $E[x]$ is defined as

$$E[x] = \sum_{x} x \, \underbrace{p(x)}_{P(x=x)}$$

Pmf

$P(x=x)$.

# of children in a family example.

H_1	2
H_2	0
H_3	0
H_4	1
H_5	1
H_6	2
H_7	2
H_8	0
H_9	3
H_{10}	4

Avg # of children in a family $= \dfrac{2+0+0+\dots 3+4}{10}$.

The above can be rewritten as- $\sum x\,p(x)$

$$\Rightarrow \quad 2 \times \frac{3}{10} + 0\left(\frac{3}{10}\right) + 4 \times \left(\frac{1}{10}\right)$$

$$\Rightarrow \sum_{i=0}^{4} i \times p(i).$$

- Expectation is the mean or average of the random variable.

$$\mu = \mu_x = E[X] = \sum_x x\,p(x) \quad \text{in case of discrete r.v.}$$

$$\mu_x = E[X] = \int_x x f(x)\,dx \quad \text{in case of a continuous random variable.}$$

Variance

→ # of children per household problem.

City 1

H_1 — 2

H_2 — 2

H_3 — 2

⋮

H_{10} — 2

$\mu = 2$

City 2

$H_1 \to 0$

$H_2 \to 0$

$H_3 \to 0$

⋮

H_9 — 4

$H_{10} \to 4$

$\mu = 2$

- Mean gives us an idea about the average of the data in the distribution.
- Variance gives us idea about how spread away the data is from the mean.
- In the above example both the city$_1$ and city$_2$ have the same mean but in case of city$_1$ the data is all $x_i = 2$ whereas in city$_2$ it is spread

Variance is defined as $Var(X) = \sum_i \frac{(x_i - \mu)^2}{n}$

In the city, Variance = 0

city Variance $= \frac{40}{10} = 4$.

Variance is also defined as $E\left((X-\mu)^2\right)$

- In case of discrete random variable

$$E[g(x)] = \sum_x g(x)p(x)$$

- In case of continuous random variable

$g(x) = y,$ $\int_x g(x) f(x)\, dx$

→ $\text{Variance}(X) = \sum \frac{(x_i - \mu)^2}{n}$

$$= E[\overbrace{(x-\mu)^2}^{g(x)}]$$

$$= E[x^2 - 2\mu x + \mu^2] \quad \text{--- (1)}$$

Properties of Expectation

1. $E[ax+b] = aE[x] + b.$

$$\int_x (ax+b) f(x)\,dx = a\int x f(x)\,dx + b\int f(x)\,dx.$$

These properties can be applied on eq ①

$$E(c) = \int_x c f(x)\,dx = c\int_x f(x)\,dx = c \cdot 1$$

$$= E[x^2] - 2\mu E[x] + \mu^2$$

$$= E[x^2] - 2\mu^2 + \mu^2 = E[x^2] - \mu^2$$

PMF OF MULTIPLE VARIABLES

example :- x - rainfall on a particular day y : temperature on that day.

X - # of children, Y - # of females.

We can define CDF, PDF defined on two variables.

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

2cm   40°c.

- If we have two random variables X and Y then.

1. $E[X+Y]$ = avg(temperature + rainfall)

$$= E[X] + E[Y].$$

2. If X and Y are independent then $P(X=x \cap Y=y) = P(X=x).P(Y=y)$

$\Downarrow$

$\underline{\underline{P(X=x) \times P(Y=y)}}$

$$E[XY] = \sum_{x,y} xy\, P(x,y)$$

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

$$= \sum_{x} x\, P(x) . \sum_{y} y \cdot P(y)$$

$$= E(X) E(Y)$$

12.13 Probability Distributions: Bernoulli and Binomial

- Mathematicians and statisticians have observed natural phenomenon and modelled their probability in the form of distributions.

eg 1. Coin Toss — T – 0, H – 1

X :- discrete random variable $X \in \{0,1\}$

$$P(X=1) = 1/2 = p$$

$$P(X=0) = 1/2 = q \quad (1 - 1/2)$$

2. Will it rain tomorrow? — T (1), F (0) $\quad X \in \{0,1\}$

$$P(X=1) = 0.15 = p$$

$$P(X=0) = 0.85 = q$$

3. Will a customer purchase a product?
   - Y (1)
   - N (0)

$$P(X=1) = 0.05 = p$$

$$P(X=0) = 0.95 = q$$

4. What will be the gender of the new born baby — M (0), F (1)

$$P(X=1) = 0.5 = p$$

$$P(X=0) = 0.5 = q$$

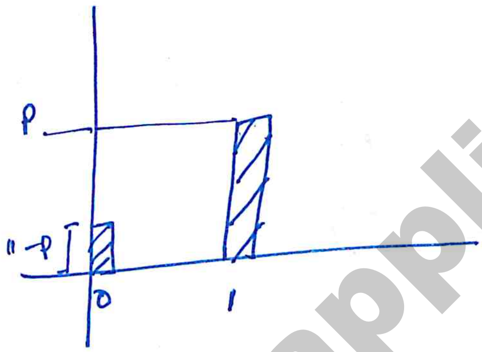
In general $x \in \{0,1\}$

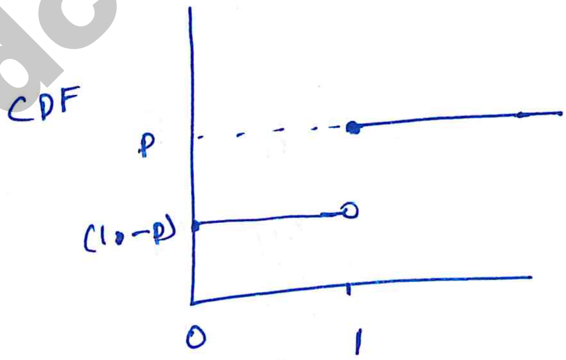
$P(x=1) = p$ success probability.

$P(x=0) = 1-p = q$ = failure probability.

- If x is a random variable which follows Bernoulli's distribution, then ($X \sim$ Bernoulli(p)), X has two outcomes $X \in \{0,1\}$ and its parameter p represents its probability of success $\underline{p = P(x=1)}$.

- p is the parameter of the distribution. Given p we can plot the the P.M.F of the distribution & the CMF of the distribution also.





$\rightarrow$ Mean $= E(x) = \mu$

$= \sum_i x\,p(x)$

$= 0 \cdot q + 1 \cdot p$

$= p.$

– Variance$(X) = E[X^2] - \mu^2$

$$= \sum_i x^2 p(x) - p^2$$

$$= 0 \cdot q + 1p - p^2$$

$$= p(1-p) = \underline{\underline{pq}}.$$

Bionomial Distribution

example 1:: If we want to test a new drug/medicine, we would like to test for how many patients will benefit from this drug.

$$P_1 \quad P_2 \quad \cdots \quad P_n$$

$$0/1 \quad 0/1 \quad \cdots \quad 0/1$$

Prob(K patients will benefit out of n patients) = ?

→ Out of n patients how many will benefit will be answered by a Bionomial random variable

2. Toss n coins at a time and prob of getting a single head is p. Probability of getting K heads in n tosses is given by bionomial distribution.

i) All the n tosses are independent

ii) $P(X=1) = 0.5 = p$, for each of the trails the probability of success remains the same.

iii) Each of the trail is a Bernoulli(p).

$$\boxed{P(k \text{ heads in } n \text{ trails}) = {}^nC_k\, p^k\, q^{(n-k)}}$$

$X \in \{0, 1, 2 \ldots n\}$

$P(X=k) = P(k) = {}^nC_k\, p^k (1-p)^{n-k}$ — PMF. (Prob. Mass function)

$P(X \leq k) = \sum_{i=0}^{k} {}^nC_i\, p^i (1-p)^{n-i}$ — CDF. (Cumulative density function).

→ If X is a Binomial random variable it is represented as $X \sim$ Bernoulli (n, p)

$X = X_1 + X_2 + X_3 \ldots + X_n$.

Binomial

Bernoulli (Each is a bernoulli random variable).

→ Mean $\mu = E(X) = \sum_{k=0}^{n} k \cdot p(X=k) = \sum_{k=0}^{n} k\, {}^nC_k\, p^k (1-p)^{n-k}$

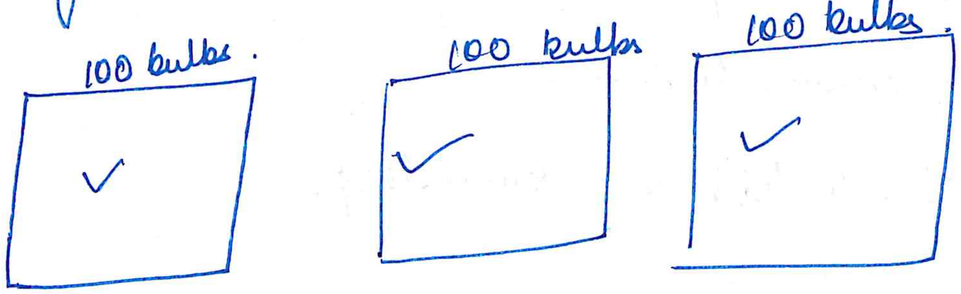
→ Variance $= Var(X_1 + X_2 \ldots + X_n)$ As all $X_1 X_2 \ldots X_n$ are independent.

$= Var(X_1) + Var(X_2) + \ldots Var(X_n)$

$= pq + pq + \ldots\ldots pq$ (n times) $\Rightarrow$ npq

Real World Application of Binomial / Bernoulli R.V.

Manufacturing example of a bulb factory



P(A bulb is faulty in a lot) = 0.01 p.

Bernoulli ($p=0.01$, $n=100$)

P(K out of 100 bulbs in a pack are faulty) = $100_{C_k}\, p^k q^{n-k}$.

– How estimation for p is done.

$n=100$

$P_1 = 2$ faulty n_1

$P_2 = 0$ faulty n_2

⋮

$P_{10} = 2$ faults n_m

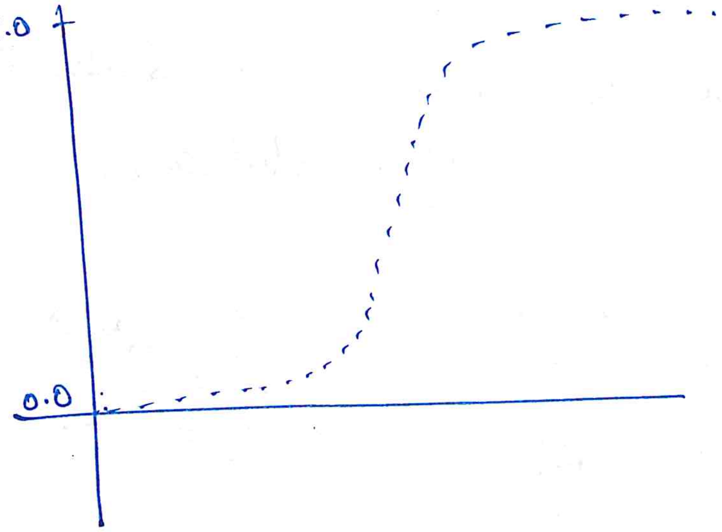
$$\boxed{\hat{p} = \frac{\sum_{i=1}^{m} x_i}{n \times m}}$$

To estimate the value of $\hat{p}$.

→ The PMF of Binomial Distribution.

Bell Shaped but not continuous.

CDF



S - Shaped curve but not continous.

12.14 POISSION DISTRIBUTION

Example: 1. Number of calls recieved at a call centre per hour

(a) Calls happen at a constant rate(λ)

(b) Occurance of one call does not affect another.

(c) Two or more events (calls cannot occour simultaneously

2. No of pateints arriving at the emergency between 10PM – 11PM.
3. No of customers at the counter per hour.
4. ~~No~~ No of insurance claims in a year.
5. No of goals in a sport event in between 2 teams.
6. # of visitors on our website per minute.

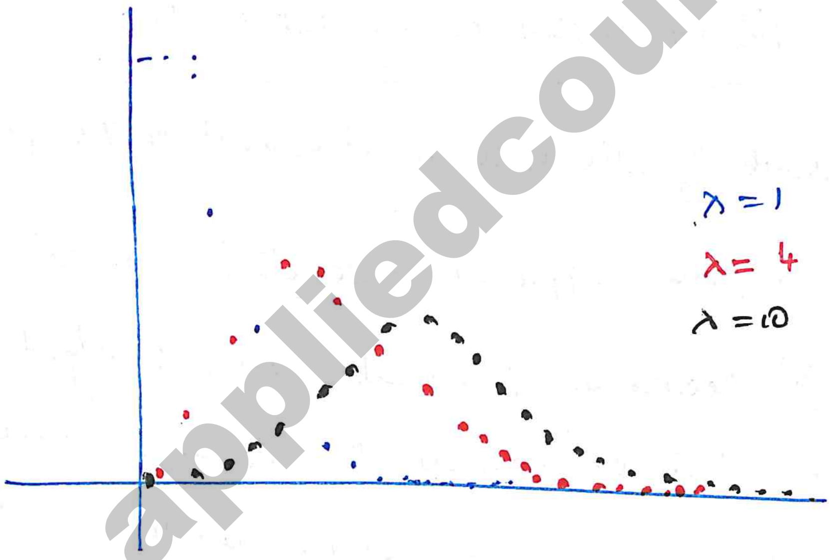
→ The random variable x is said to be poission distributed if

$X \sim \text{Poisson}(\lambda)$ λ is the rate and also the parameter of the distribution.

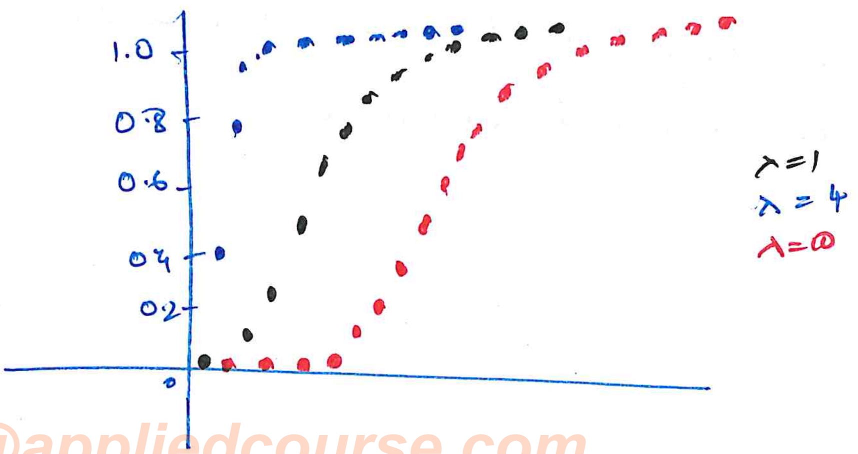
PMF: $P(X=k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$, $\displaystyle\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} e^{\lambda} = 1$

CDF: $P(X <= k) = \displaystyle\sum_{x=0}^{k} \frac{\lambda^x e^{-\lambda}}{x!}$

PMF Curve



CDF Curve



→ Mean of poission random variable $= \mu = E[x] = \sum_{x=0}^{\infty} \frac{x e^{-\lambda} \lambda^{x}}{x!}$

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

$$\Rightarrow \lambda e^{-\lambda} \left[\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} \cdots \right]$$

$$\Rightarrow \lambda e^{-\lambda} e^{\lambda}$$

$$\Rightarrow \underline{\underline{\lambda}}$$

Estimation parameter is the mean of all the observations $\Rightarrow \boxed{\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i}$

→ Variance :-

$$Var(x) = E(x^2) - \mu^2$$

$$\Rightarrow E(x(x-1) + x) - \mu^2$$

$$\Rightarrow E(x(x-1)) + E(x) - \mu^2$$

$E(x) \to \lambda$, $\mu^2 \to -\lambda^2$

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

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

$$\Rightarrow \underline{\underline{\lambda}}$$

∴ Mean and variance of poission R.V are same $= \lambda$.

→ If $X_1 \sim$ Poisson (λ_1)

$X_2 \sim$ Poisson (λ_2)

- X_1 & X_2 are independent

$(X_1 + X_2)$ is also a poisson random variable with

parameter $\lambda_1 + \lambda_2$ i.e $\sim$ Poisson $(\lambda_1 + \lambda_2)$

12.15 UNIFORM (CONTINUOUS DISTRIBUTIONS)

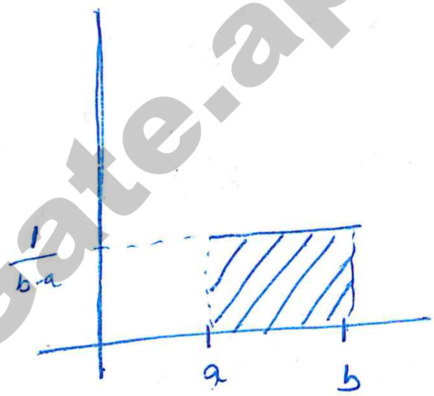
→ $X \sim$ Uniform (a, b) means that $x \in [a, b]$.

a = minimum value.

b = maximum value.

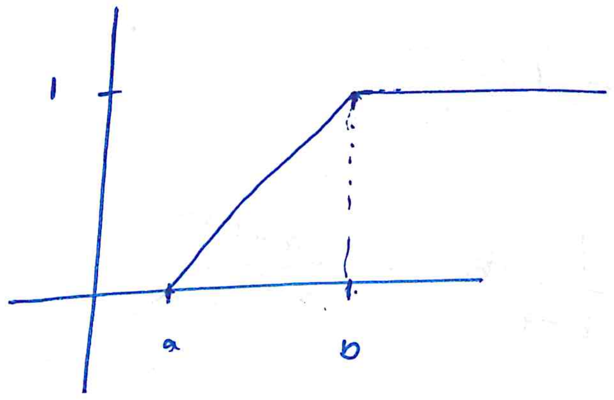
→ X is a continuous random variable which is in between a and b.

Pdf



$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0, & \text{otherwise.} \end{cases}$$

Cummulative distribution function (CDF)



$$F(x) = P(X \le x)$$

$$= \int_{-\infty}^{x} f(u)\,du.$$

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{x-b} & \text{if } x \in [a, b] \\ 1 & \text{if } x \geq b \end{cases}$$

Mean $= \mu = E[x] = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{a} x \cdot 0\,dx + \int_{a}^{b} x \frac{1}{(b-a)}\,dx + \int_{b}^{\infty} x \cdot 0\,dx$

$$\Rightarrow \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{(a+b)}{2}$$

Variance

$$\text{Variance}(x) = E(x^2) - (E(x))^2$$

$$= \int_a^b x^2 \frac{1}{b-a}\,dx - \left(\frac{a+b}{2}\right)^2$$

$$\Rightarrow \left[\frac{x^3}{3}\right]_a^b \times \frac{1}{(b-a)} + \frac{(a+b)^2}{4}$$

$$\Rightarrow \frac{b^3 - a^3}{(b-a)\cdot 3} + \frac{(a+b)^2}{4}$$

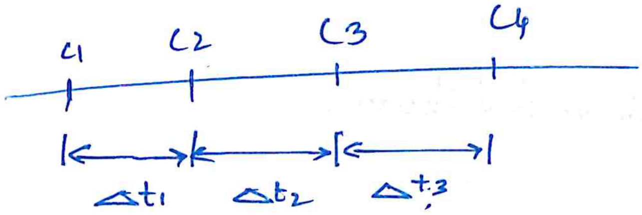
$$\Rightarrow \frac{1}{12}(b-a)^2$$

- Some applications of uniform random variables is in randomized algorithms like Randomized QuickSort.

12.16 Exponentail Distribution

- Continuous distribution

- Lets take an example of a poisson process for example the calls recieved at a call center.



If C_1, C_2, C_3, C_4 etc follow the poisson process/distribution 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.

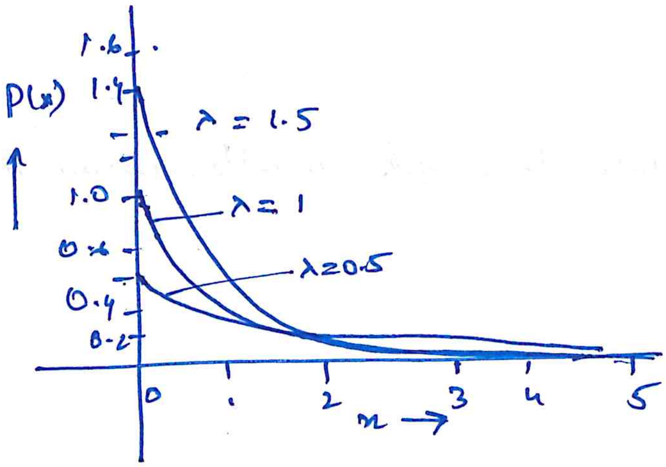
eg we can know $P(\Delta t_2 < 2\text{min})$.

Pdf $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0. \end{cases}$

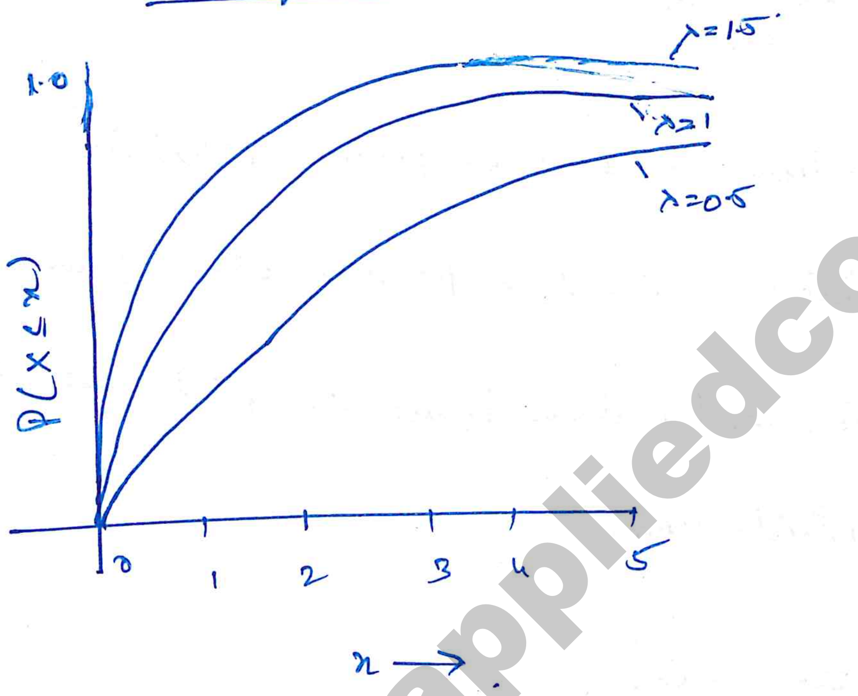
CDF $= F(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = \lambda \left[\frac{1}{-\lambda} e^{-\lambda t} \right]_0^x$

$= \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x \leq 0 - \end{cases}$

PDF for Exponential distribution



CDF for an Exponential distribution



Mean : $\mu = E(x) = \int_0^\infty x(\lambda e^{-\lambda x})\,dx$

applying Integration by parts.

$u = x \quad du = dx$

$v = \lambda e^{-\lambda x} \quad dv = \lambda e^{-\lambda x}$

$$\Rightarrow \left[-x e^{-\lambda x}\right]_0^\infty + \int_0^\infty + e^{-\lambda x} dx$$

$$\Rightarrow 0 + \left[\frac{1}{-\lambda} e^{-\lambda x}\right]_0^\infty$$

$$\Rightarrow \frac{1}{\lambda}$$

$$\boxed{\therefore \mu = \frac{1}{\lambda}}$$

$$\text{Variance} = E(x^2) - [E(x)]^2$$

$$= E(x^2) - \mu^2$$

$$= \int_0^\infty x^2 \lambda e^{-\lambda x} - \frac{1}{\lambda^2}$$ on doing integration by parts we get.

$$\Rightarrow \frac{1}{\lambda^2}$$

$$\boxed{\text{Variance} = \frac{1}{\lambda^2}}$$

→ How to estimate λ given values of outputs $x_1, x_2, x_3 \ldots x_n$,

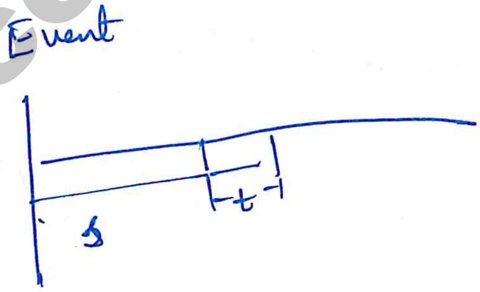
if we are given the observations we know μ (mean) $= \frac{1}{\lambda}$

$$\therefore \quad \frac{x_1 + x_2 \cdots x_n}{n} = \frac{1}{\lambda} \quad \therefore \boxed{\lambda = \frac{n}{\sum_{i=1}^{n} x_i}}$$

→ Memoryless Property of Exponential distribution.

$$P(X > s+t \mid x > s) = P(x > t).$$

$$= \frac{P(X > s+t \cap X > s)}{P(X > s)}$$



$$= \frac{P(X > s+t)}{P(X > s)}$$

$$= \frac{1 - P(X \leq s+t)}{1 - P(X \leq s)} \Rightarrow \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$$

$$= \underline{\underline{P(X > t)}}$$

12.17 Normal Distribution

- Also known as Gaussian distribution, most used and popular distributions.

→ It is a continious distribution

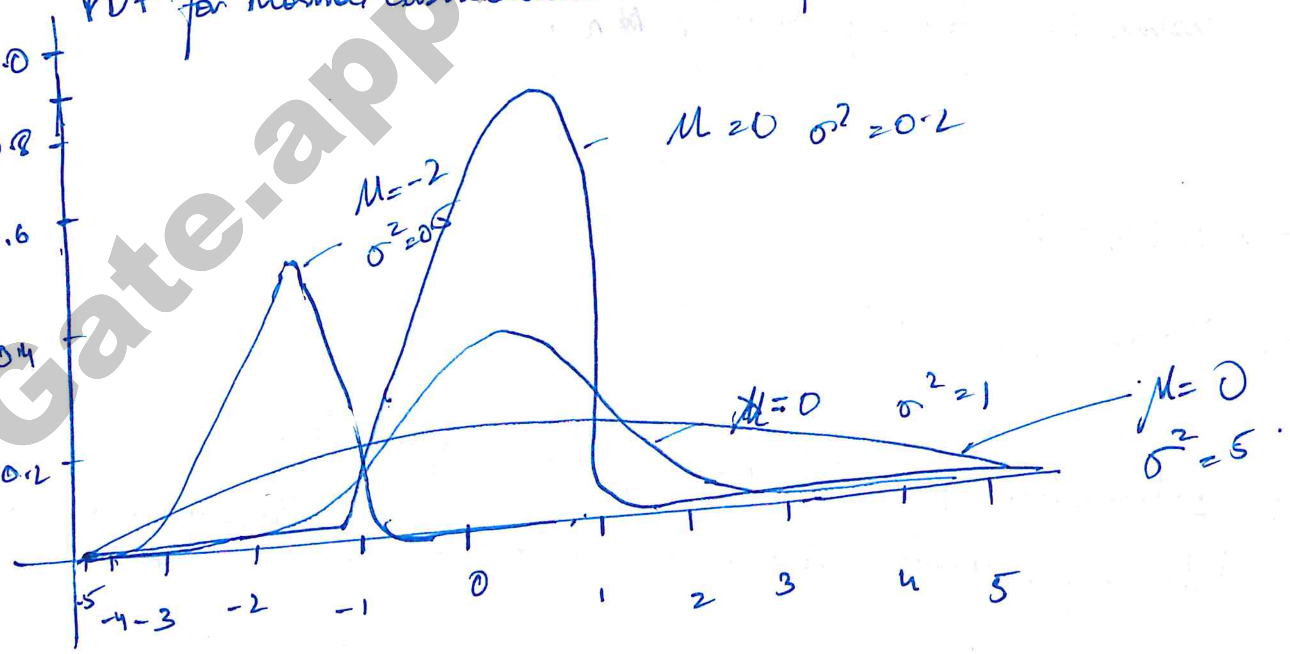
- egs :- heights, lengths of leaves, weights of people in a set of population if not normally distributed they are approximately Normal or Gaussian distributed.

- Represented as $X \sim N(\mu, \sigma^2)$ μ and σ^2 should be finite.

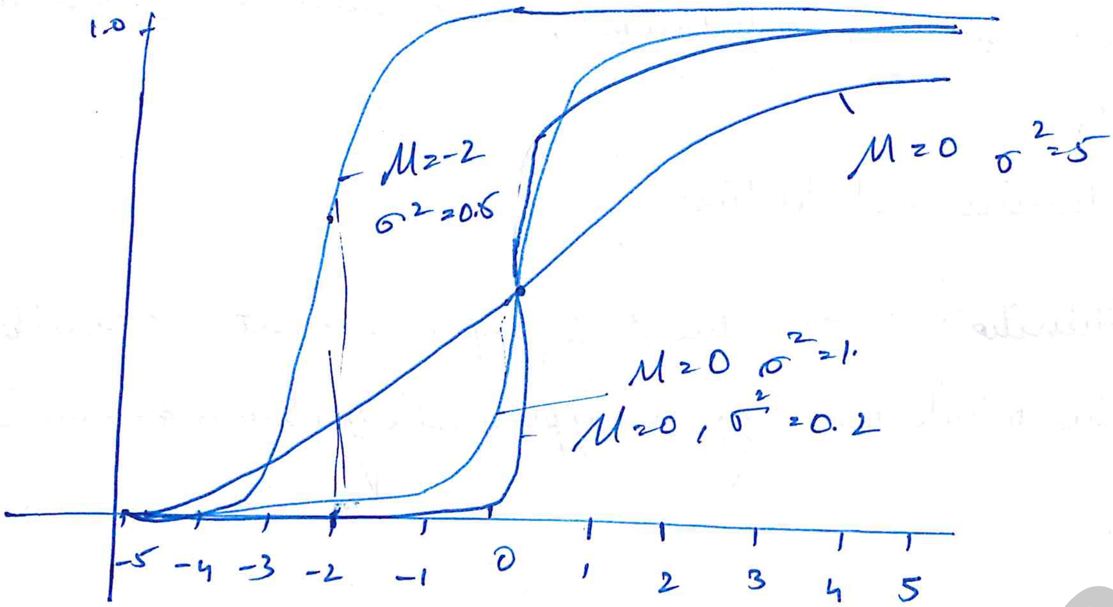
Standard deviation $= +\sqrt{\text{Variance}}$

$$\text{PDF} = f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \sigma \neq 0.$$

PDF for Normal distribution - Bell shaped curve.



→ As variance increases the peakedness of the curve decreases in the PDF.



(More accurate figures can be found on the Wiki Page for Normal distribution)

→ A Normal distribution can be thought of as a logical extension of binomial distribution for large values of n.

→ $E(x) = \mu$

→ $Var(x) = \sigma^2$.

→ Estimation of parameters from given sample data.
We can estimate μ and σ^2 for them using the following formula

$X \sim N(\mu, \sigma^2)$ $\qquad \hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$ $\qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$

Standard Normal distribution

If $X \sim N(\mu, \sigma^2)$ then we can convert it to another normal variable Z with 0 mean and $\sigma^2 = 1$ by the following transformation

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

12.18 Mean, Median & Mode.

Suppose if we are given few observations, for example heights of the students in a classroom.

n-observations 160, 180, 181, 172.6, 150, 120 . . . . . .
$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \; . \; . \; . \; . \; . \; x_n$.

$$\text{Mean} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots x_n}{n}$$

Median = ① Sort all the values and ② select the middle value.

160, 160, 161, 163, 165, 166, 180
1 2 3 4 5 6 7

If we have even no of observations, we need to take average of middle 2 elements

160, 160, 161, 162, 163, 165, 167, 168
1 2 3 4 5 6 7 8

$= (162+163)/2 = \underline{162.5}$

Medain is not much affected by extreme value / or one error.

150, 160, 161, 162, 163, 165, 168 → By mistake taken as 198

still medain remains 162.

Mode : The value that is observed most frequently

Let say following is the list of observation.

Height	Frequency
150	10
155	20
160	70
170	15
180	4

160 - is the Mode of the data set.