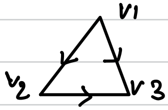
SDM

UNIT -1 Graph Theory :

Directed graph (Digraph) → collection of (V, E), where V is the non empty set and E is the set of edges which contains ordered pairs of elements of V

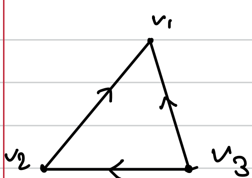
Eg :

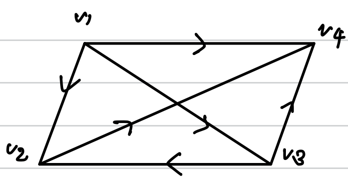


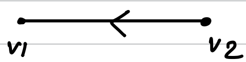
Note : If v_1, v_2 is an edge then v_1 is initial vertex and v_2 is final vertex

In degree and out degree

- let G be a directed graph and v be the vertex number of edges coming to v is called in degree of v ($id(v)$ or $d^-(v)$)
- The no of edges going out from v is called out degree of v [$odv(v)$ or $d^+(v)$]







$id(v_1) = 1$
$id(v_2) = 1$
$id(v_3) = 1$
sum = 3
$od(v_1) = 1$
$od(v_2) = 1$
$od(v_3) = 1$
sum = 3

$id(v_1) = 0$
$id(v_2) = 2$
$id(v_3) = 1$
$id(v_4) = 3$
} 6

$od(v_1) = 3$
$od(v_2) = 1$
$od(v_3) = 2$
$od(v_4) = 0$
} 6

$id(v_1) = 1$
$id(v_2) = 0$
} 1

$od(v_1) = 0$
$od(v_2) = 1$
} 1

→ The sum of the in-degree and out-degree is equal to the no of edges

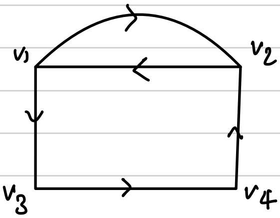
Properties

① sum of indegrees of vertices and sum of out-degrees of vertices is equal to the no. of edges

Let G be the digraph with v_i, $i = 1 \ldots n$ vertices then

$$\sum_{i=1}^{n} id(v_i) = \sum_{i=1}^{n} od(v_i) = \text{no of edges}$$

Eg :



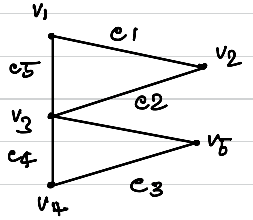
$id(v_1) = 1$ $\quad$ $od(v_1) = 2$
$id(v_2) = 1$ $\quad$ $od(v_2) = 2$
$id(v_3) = 1$ $\quad$ $od(v_3) = 1$
$id(v_4) = 2$ $\quad$ $od(v_4) = 0$

$\Sigma id(v) = 5$ $\quad$ $\Sigma od(v) = 5$ $= |E|$

Graph

Is a pair (V, E) where V is a non-empty set and E is set of unordered pair of elements taken from the set V

Eg:

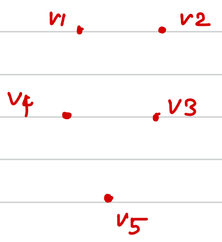


$e_1 = v_1 v_2 = v_2 v_1$

→ Null graph : A graph G is said to be Null graph if $|E| = \phi$

no of edges → cardinality of e

Eg :



$V = \{v_1, v_2, v_3, v_4, v_5\}$
$E = \phi$

no of vertice

→ Trivial Graph:

A null graph is said to be a trivial graph if $|V| = 1$

eg: $\bullet v_1$

→ Finite Graph:

A graph G is said to be finite graph if it contains finite member of vertices and edges

Order and Size

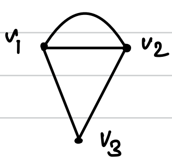
Number of vertices in the graph G is called order of G.

Number of edges in G called size of G

denoted by $|V| = \#(V)$

denoted by $|E| = \#(E)$

Ex:

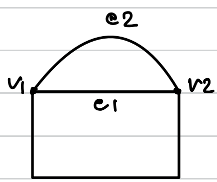


$|V| = 3$

$|E| = 4$

→ parallel edges: Two edges are said to // if they have same end points

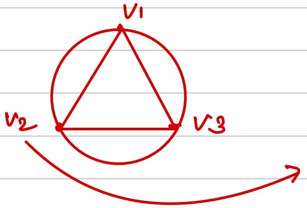
Eg:



e_1 and e_2 are parallel

→ parallel vertices: Two vertices are said to be // if they have an edge b/w them.

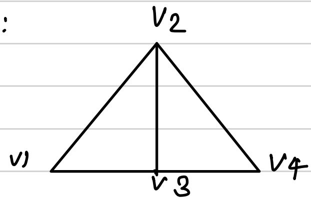
→ Multiple edges : Two or more edges are said to be multiple edges if they have the same end vertices



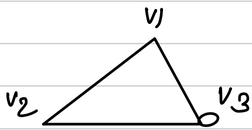
between V1 and V3 ⟶ 2 edges

Simple graph : is said to be a simple graph if it does not contain loops and multiple edges.

Eg :



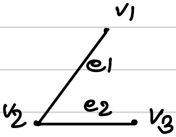
Simple Graph



Not simple graph

→ Incidence : When a V of a graph . G is an end vertex of an edge of the graph G, we say that e is incident to (or on) V

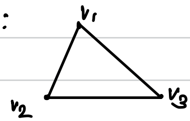
→ Adjacent edges : Two non-parallel edges are said to be adjacent if those two edges incident to same vertex V.

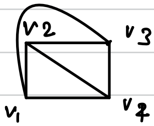


e1 & e2 are adjacent

→ Complete graph : A simple graph of order ≥ 2 is said to be a complete graph if there is an edge b/w every pair of of vertices (k_n) $\{$ only one edge b/w every pair of vertices $\}$

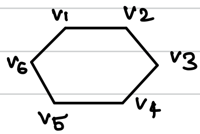
Eg:





→ Bipartite graph: Let G be the simple graph with the vertex set V is the union of two mutually disjoint subsets V_1 and V_2 such that each edge in G has one end vertex in V_1 and other end vertex V_2

Eg:

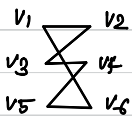


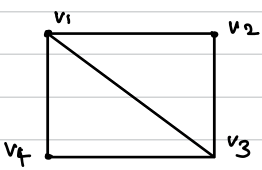
$V = \{v_1, v_2, v_3, v_4, v_5, v_6\}$

$V_1 = \{v_1, v_3, v_5\}$

$V_2 = \{v_2\ v_4\ v_6\}$

$V = V_1 \cup V_2$

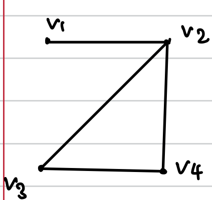




$\longrightarrow$ <u>NOT bipartite</u>

$V_1 = \{v_1 \ldots\ldots\}$ $\quad V_2 = \{v_2 \ldots\ldots\}$

v_3 is neither in V_1 or V_2 hence it is not bipartite graph



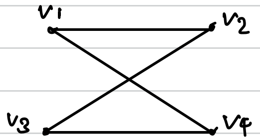
v_1	v_2
v_3	v_4
v_4	

Not a bipartite graph

Since there is an edge between v_3 and v_4 and v_3 to v_2 $\therefore$ v_3 should be in either V_1 or V_2 hence it is not a bipartite graph.

Complete Biparted graph

A biparted graph is said to be complete biparted graph if there is an edge between every vertex V_1 and every vertex V_2



$k(r,s)$

V_1	V_2
V_1	V_2
V_3	V_4

$\boxed{\text{Property 2 :}}$ If G is a simple graph then P.T $2|E|$ (2 times the size) $\le |V|^2 - |V|$
↓
order

Since G is a simple graph, max number of edges is number of pairs of vertices in G.

$$m \le n/2 \le \frac{n!}{(n-2)!\,2!} \le \frac{n(n-1)\cancel{(n-2)!}}{\cancel{(n-2)!}\,2!}$$

$$m \le \frac{n(n-1)}{2} \Rightarrow 2m \le n(n-1)$$
$$2m \le n^2 - n$$

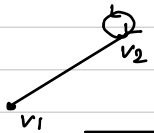
If G is complete graph with n vertices then $2|E| = |V|^2 - |V|$
$2m = n^2 - n$

Note: A complete partite graph is denoted by $G(V_1\ V_2 . E)$ or $k_{r.s}$

→ If G is a complete bipartite $(k_{r,s})$
then a) Number of vertices in G are $r \cdot s$
b) Number of edges is G are rs

Degree of Vertex :

Let G be the graph and v is any vertex in G then the degree of v is the number of vertices incident to v.



$\longrightarrow$ degree of $v_2 = 3$

$\longrightarrow$ degree of $v_1 = 1$

k-Regular Graph

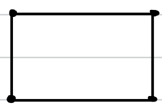
A graph is said to be k-regular graph if degrees of each vertex is k. N_k denote k-regular graph with n-vertices.

K-hypercube :

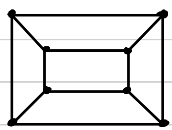
A loop free k regular with 2^k vertices is called k-dim hypercube and denoted by Q_k.

Eg :

1D $\longrightarrow$ 1 degree Q_1



$\longrightarrow$ 2D $\longrightarrow$ 2 degree Q_2



— 3D $\longrightarrow$ 3 degree Q_3

Handshaking property :

Let G be the graph and v_i, $i = 1, 2 \ldots\ldots n$ are vertices of G, then sum of degree of v_i. $\sum_{i=1}^{n} \deg(v_i) = 2|E|$

$\longrightarrow$ no of vertices $= 2$

degree of each vertex $= 1$

no of edge $= 1$

$\therefore$ sum of degrees $= 2$
$= 2 \times$ no of edges

Theorem 1

For any graph G, the no of odd degree of vertices is even.

proof : Let G be the graph with $v_1, v_2 \ldots\ldots v_k$ are odd degree vertices.

$v_{k+1}, v_{k+2} \ldots\ldots v_n$ are even degree vertices

$$\sum_{i=1}^{n} deg(v_i) = \sum_{i=1}^{k} deg(v_i) + \sum_{i=k+1}^{n} deg(v_i)$$

By handshaking property $\sum_{i=1}^{n} deg(v_i)$ is even. Since $deg(v_{k+i})$

$i = 1, 2 \ldots\ldots$ is even, therefore $\sum_{i=k+1}^{n} deg(v_i)$ is even.

Hence $\sum_{i=1}^{k} deg(v_i)$ is even and $deg(v_i)$ is odd for $i = 1, \ldots$ k

Hence number of vertices with odd degree is always even.

Theorem 2 :

Let G be the complete graph with $|V| = n$ (no of vertices) and $|E| = m$ (no of edges) then $4m \leq n^2$.

Since G is a bipartite graph, $V = V_1 \cup V_2$

where $|V_1| = r$ and $|V_2| = s$, $\boxed{r + s = n}$

$\boxed{s = n - r}$

$\boxed{K(r,s)}$ in a complete bipartite graph $\rightarrow$ no of edges $(rs = m)$

W.K.T Arithmetic mean $\geq$ graphical mean

$$\frac{|V_1| + |V_2|}{2} \geq \sqrt{V_1 V_2}$$

$$\frac{r+s}{2} \geq \sqrt{m}$$

$\frac{n}{2} \geq \sqrt{m}$

$\Delta m \leq n^2$

→ For a graph G what is the largest possible value of cardinality V if $|E| = 19$ and degree of each vertex ≥ 4.

$|E| = 19$ $\quad \deg(v_i) \geq 4$

$|V| = ?$

W.K.T $\sum_{i=1}^{n} \deg(v_i) = 2|E| \longrightarrow$ from handshaking property

Given that $\deg(v_i) \geq 4$ for each $i = 1 \ldots \ldots, n$

then $\sum_{i=1}^{n} \deg(v_i) \geq 4n$ $\quad |V| = n$

$2|E| \geq 4n$

$2 \times 19 \geq 4n$

$\boxed{9.5 \geq n}$ Largest integer n is $\boxed{|V| \leq 9}$

2) For a graph with n vertices and m edges if δ is min and Δ is the max of degrees of the vertices then S.T $\delta \leq \frac{2m}{n} \leq \Delta$

$\delta \rightarrow$ lower case delta

$\Delta \rightarrow$ upper case delta

ans Let $v_1, v_2 \ldots \ldots v_n$ are vertices of G

Given $\deg(v_i) \geq \delta \Rightarrow \sum \deg(v_i) \geq n\delta$

$\deg(v_i) \leq \Delta \Rightarrow \sum \deg(v_i) \leq n\Delta$

↓

$n\delta \leq \sum \deg(v_i) \leq n\Delta$

By handshaking property

$n\delta \leq 2m \leq n\Delta$

$\delta \leq \frac{2m}{n} \leq \Delta$

3. S.T k-dim hypercube Q_k has $k2^{k-1}$ edges. No of vertices $= 2^k$ degree at each vertex $= k$. $\boxed{\Sigma \deg(v_i)\ 2k}$

Q_k has 2^k vertices and each vertex has degree k

$\therefore \Sigma \deg(v_i) = 2^k \cdot k$

By handshaking property

$\Sigma \deg(v_i) = 2|E|$

$2^k k = 2|E|$

$\boxed{|E| = k2^{k-1}}$

4. What is the dim of hypercube with 524288 edges?

$|E| = 524288$

$k = ?$

$2^{?}$ (524288)

$= 2^{19}$

$= 2^4 \cdot 2^{15}$

$= 16 \cdot 2^{15}$

$k = 16$

5. Let G be a graph of order 9 such that each vertex has degree 5 or 6.

P.T atleast 5 vertices have degree 6 or atleast 6 vertices have degree 5

Let G be a order of 9

Let there be p vertices of deg 5 and q vertices of 9-p vertices of deg of 6 sum of deg of vertices is,

$\Sigma \deg(v_i) = 5p + 6q = 5p$

$= 5p + 54 - 6p$

$= \boxed{54 - p}$

By handshaking property $54 - p$ is even.

$(0 \le p \le 9$ or $0 \le q \le 9)$

$\therefore$ $P = 0, 2, 4, 6, 8$

P	Q
0	9
2	7
4	5
6	3
8	1

In all possible combination we have either $q \geq 5$ or $p \geq 6$

6. S.T there is no graph with 28 edges and 12 vertices in the following cases.

i) The degree of vertex is either 3 or 4

ii) The degree of a vertex is either 3 or 6.

Let P be vertices with deg 3 then $q = 12 - p$ is vertices of deg 4.

$3p + 4q = 3p + 4(12 - p) = 3p + 48 - 4p$

$\Sigma \deg(v_i) \geq 2|E| \geq 2$

$48 - p \geq 2$

$48 - p \geq 56$

$p = -8$ since $0 \leq p \leq 12$

$\therefore$ there is no such graph

ii $3p + 6q$

$3p + 6(12 - p) \geq 2(28)$

$3p + 72 - 6p \geq 56$

$72 - 5p \geq 56$

$72 - 56 = 3p$

$p = 16/3 = 5.3$

7) Is there a simple graph with 1, 1, 3, 3, 3, 4, 6, 7 as the degree of its vertices

Let $v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$ be the vertices of the graph G and $d(v_1)=1$ $d(v_2)=1$ $d(v_3)=3$ .... $d(v_8)=7$

Since $d(v_8)=7$, v_8 should connect with 7 vertices

Since $d(v_1)=d(v_2)=1$ $\longrightarrow$ It should connect with one degree and one of them being v_8

v_1 and v_2 shouldn't connect with any other vertices other than v_8.

Now v_7 can connect with v_3, v_4, v_5, v_6 but $d(v_7)=6$ $\longrightarrow$ This is contradiction and such a graph is not possible.

8 Let D be a directed graph with n vertices. If the underlying graph of D is K_n P.T $\sum_{EV} [od(v)]^2 = \sum_{EV} [id(v)]^2$

$$\sum_{v \in V} [od(v)]^2 - [id(v)]^2$$

$$\sum_{v \in V} (od(v)+id(v))(od(v)-id(v))$$

since underlying graph is K_n, for any $v \in V$ $d(v)=n-1$

$d(v) = id(v) + od(v) = n-1$

$$\Sigma(od(v))^2 - \Sigma(id(v))^2 = \Sigma(n-1)(od(v)-id(v))$$

$$=(n-1)\sum_{v \in V} od(v) - id(v)$$

$$= n-1 \times 0$$

$$\Sigma(od(v))^2 = \Sigma(id(v))^2 //$$

Isomorphism :

Two graphs $G=(\bar{V}, E)$ and $G'(V', E')$ are said to be isomorphic to each other if for every $f: \bar{V} \to V'$ s.t

(i) f is one-one

ii) $u, u \in \bar{V}$ if (v,u) is an edge in E then $(f(v), f(u))$ to be an edge in E'

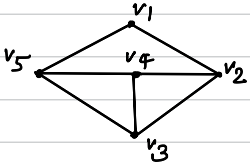
Note : Suppose G and G' are isomorphic to each other then

i) order of G and G' is same

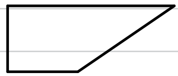
ii) size of G and G' is same

iii) Number of vertices of degree k in G and G' should be same.

Ex1. Verify that the given graphs are isomorphic or not



G



Path and Cycle

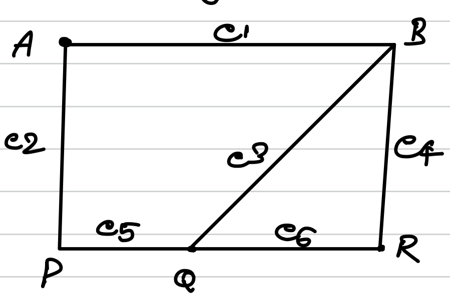
Path: A trail with no vertex repeated is called path

Cycle: A circuit with no vertex except end vertex is called cycle

Ex 1) For the following graph, indicate the nature of the following walks

i) $V_1 e_1 V_2 e_2 V_3 e_2 V_2$ ⟶ open walk but not trial
ii) $V_4 e_7 V_1 e_1 V_2 e_2 V_3 e_3 V_4 e_4 V_5$ → open walk & trial
iii) $V_1 e_1 V_2 e_2 V_3 e_3 V_4 e_4 V_5$ → open walk & path
iv) $V_1 e_1 V_2 e_2 V_3 e_3 V_4 e_7 V_1$ ⟶ cycle
v) $V_6 e_5 V_5 e_4 V_4 e_3 V_3 e_2 V_2 e_1 V_1 e_7 V_4 e_6 V_6$
↳ circuit but not cycle.

2 Find all paths from a vertex A to R and also indicate their lengths.

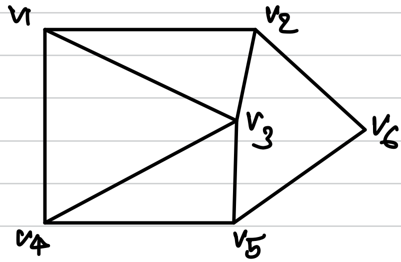


→ $A e_1 B e_4 R$ ⟶ length is 2
→ $A e_2 P e_5 Q e_3 B e_4 R$ ⟶ length is 4
→ $A e_1 B e_3 Q e_6 R$ ⟶ length is 3
→ $A e_2 P e_5 Q e_6 R$ ⟶ length is 3

Distance

→ Let A and B be two vertices of G then shortest path b/w A and B is called distance

Ex 3) Determine number of different paths of length 2 in the graph G

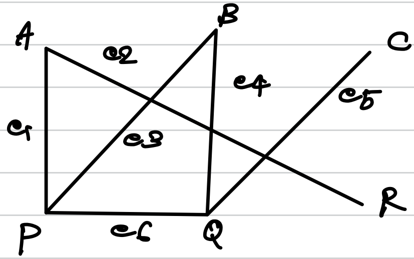


soln: Number of paths passing the vertex V_1 of lengths 2 are

$V_1 \rightarrow 3C_2 = 3$
$V_2 \rightarrow 3C_2 = 3$
$V_3 \rightarrow 4C_2 = 6$
$V_4 \rightarrow 3C_2 = 3$
$V_5 \rightarrow 3C_2 = 3$
$V_6 \rightarrow 2C_2 = 1$

$\overline{19}$

ex 4) Find all the cycles in G



Soln:

$P e_3 B e_4 Q e_6 P$, $P e_6 Q e_4 B e_3 P$ } 1

$Q e_4 P e_3 B e_4 Q$, $Q e_4 B e_3 P e_6 Q$ } 3

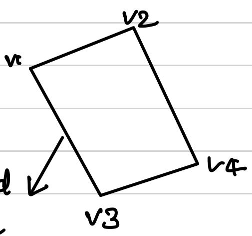
$B e_3 P e_6 Q e_4 B$, $B e_4 Q e_6 P e_3 B$ } 2

Connected Graph

→ A graph G is said to be connected if there exists a path b/w every pair of distinct vertices.

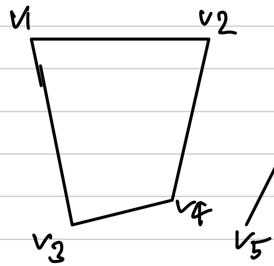
Otherwise the graph is disconnected

eg 1)



connected graph

2)



disconnected graph

Component

→ The subgraph G_1 is said to be component of G if it satisfies the following

i) G_1 is connected

ii) If G_2 is connected subgraph of G and G_1 is subgraph of G_2 then $G_1 = G_2$

Note:

Number of components in the graph in G is denoted by $\kappa(G)$

kappa

Theorem 1

If a graph has exactly two vertices of odd degree, then there must be a path connecting these vertices

proof: Let v_1 and v_2 be two vertices of odd degree. Assume there is no path b/w v_1 and v_2. This implies the graph G is disconnected i.e the graph G contains atleast two components say H_1 and H_2

Now $v_1 \in H_1$ or H_2

$v_2 \in H_2$ or H_1

H_1 & H_2 are connected graphs containing one vertex of odd degree. This contradicts that connected graph contains even number of vertices of odd degree.

Therefore there is a path b/w v_1 and v_2

Theorem 2 :

A simple graph with n vertices and k components can have atmost $\frac{1}{2}(n-k)(n-k+1)$ number of edges.

soln; Given G is simple graph with $|G| = n$

Let n_i be the number of vertices in ith component

$$n_1 + n_2 + \dots\dots + n_k = n$$

$$(n_1 - 1) + (n_2 - 1) + \dots + (n_k - 1) + k = n$$

$$(n_1 - 1) + (n_2 - 1) + \dots + (n_k - 1) = n - k$$

sq on both sides

$$(n_1 - 1)^2 + (n_2 - 1)^2 + \dots\dots + (n_k - 1)^2 + \sum_{i \neq j} 2(n_i - 1)(n_j - 1)$$

$$= (n-k)^2$$

PTO →

$$\Rightarrow (n_1-1)^2+(n_2-1)^2+\ldots\ldots+(n_k-1)^2 \le (n-k)^2$$

$$n_1^2+n_2^2+\ldots\ldots+n_k^2-2(n_1+n_2+\ldots\ldots+n_k)+k \le (n-k)^2$$

$$\sum_{i=1}^{k} n_i^2-2n+k \le n^2+k^2-2nk$$

$$\sum_{i=1}^{k} n_i^2 \le n^2+2n-2nk+k^2-k \longrightarrow ①$$

Max number of edges in ith component is:

$$\frac{1}{2} n_i (n_i-1)$$

Let N be the max number of edges in the graph G then

$$N=\sum_{i=1}^{k} \frac{1}{2} n_i (n_i-1)$$

$$=\frac{1}{2}\sum_{i=1}^{k}(n_i^2-n_i)$$

$$=\left[\sum_{i=1}^{k} n_i^2-\sum_{i=1}^{k} n_i\right]$$

$$N \le \frac{1}{2}\left[n^2+2n-2nk+k^2-k-n\right]$$

$$\le \frac{1}{2}\left[n^2+n-2nk+k^2-k\right]$$

$$\le \frac{1}{2}\left[n^2-2nk+k^2+n-k\right]$$

$$\le \frac{1}{2}\left[(n-k)^2+n-k\right]$$

$$\le \frac{1}{2}(n-k)(n-k+1)$$

Theorem 3

A connected graph with n vertices has at least $n-1$ edges.

Theorem 4

If G is a simple graph with n vertices and in which the degree of every vertex is at least $\frac{n-1}{2}$ then prove that G is a connected graph

proof : Given G is simple graph and $|G| = n$ $\forall$ $v \in G$, $\deg(v) \geq \frac{n-1}{2}$.

Assume that G is not connected. G may contain two components say H_k with k vertices and H_{n-k} with $n-k$ vertices

i) For any vertex v in the max degree is $k-1$

$$① \Rightarrow \quad k-1 \geq \frac{n-1}{2}$$

$$k \geq \frac{n-1}{2} + 1$$

$$k \geq \frac{n+1}{2} \longrightarrow ②$$

ii) For any vertex v in H_{n-k} max degree of u in $n-k-1$

$$① \Rightarrow \quad n-k+1 \geq \frac{n-1}{2}$$

$$-k \geq \frac{n-1}{2} - n + 1$$

$$\geq \frac{n-1-2n+2}{2} = \frac{-n+1}{2}$$

$$k \leq \frac{n-1}{2} \longrightarrow ③$$

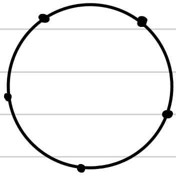
But WKT $\frac{n+1}{2} > \frac{n-1}{2}$ this contradicts ② & ③

$\therefore$ G is connected

Theorem 5

Prove that connected graph G is remains connected after removing an edge e from G iff e is a part of some cycle G.

proof: Assume that e is a part of some cycle C in G by removing e from C the G is still connected



Suppose e is not a part of cycle in G then end vertices of e must have at most one path if we remove e then the graph G is disconnected

Theorem 6

VVI

Let G be a disconnected graph of even order n with two components each of which is complete. Prove that G has a minimum of $n(n-2)/4$ edge.

proof: Let H_1 and H_2 be two components of G. Let x be number of vertices in H_1 then $n-x$ be the number of vertices in H_2.

Since H_1 and H_2 are complete graph number of edges in H_1 are $\frac{x(x-1)}{2}$, number of edges in H_2 are $\frac{(n-x)(n-x-1)}{2}$.

let m be the number edge in G then $m = \frac{x(x-1)}{2} + \frac{(n-x)(n-x-1)}{2}$

$$m = \frac{x^2 - x + n^2 - 2nx - n + x + x^2}{2}$$

Diff w.r.t x

$$\frac{dm}{dx} = \frac{1}{2}\left(2x - \cancel{1} - 2n + 2x + \cancel{1}\right)$$

$$= \frac{1}{2}\left(4x - 2n\right)$$

$$\frac{dm}{dx} = 2x - n$$

C.P is $x = \frac{m}{2}$

$$\frac{d^2m}{dx^2} = 2 > 0$$

Since $\frac{d^2m}{dx^2} > 0$, C-P is point minimum

$$\min(m) = \frac{m}{x} = \frac{n}{2}$$

$$\min(m) = \frac{\frac{n}{2}\left(\frac{n}{2} - 1\right)}{2} + \frac{\left(n - \frac{n}{2}\right)\left(n - \frac{n}{2} - 1\right)}{2}$$

$$= \frac{\frac{n}{2}\left(\frac{n}{2} - 1\right)}{2} + \frac{\frac{n}{2}\left(\frac{n}{2} - 1\right)}{2}$$

$$\min(m) = \frac{n}{2}\left(\frac{n}{2} - 1\right)$$
$$= \frac{n(n-2)}{4}$$

Euler circuit and Euler trial.

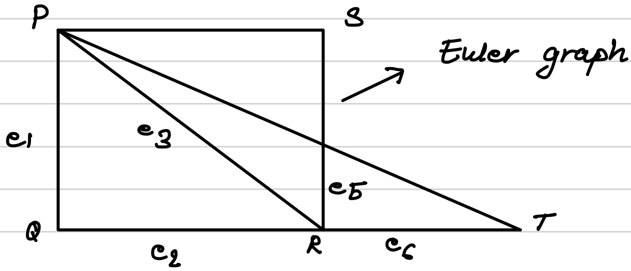
→ A circuit in connected graph is Euler circuit if that contains all the edges of G.

→ If there is a trial in a connected graph G that contains all edges of G then it is called Eulers trial.

→ A connected graph that contains Euler's circuit is called Euler's graph

→ A connected graph that contains Euler's trial is called semi-Euler's Graph

Euler circuit and Euler trial

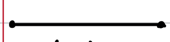


#

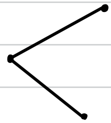
Tree

A graph G is said to be tree if it is connected and has no cycle.

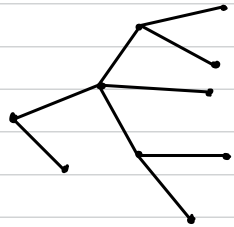
examples : ↓



(a)



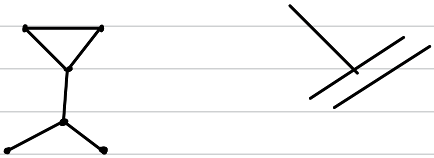
(b)



(c)

Illustration

a)



Note:

1) In the first graph a, containing cycle whereas second is not connected

2) Each component of the second graph is a tree, such a graph is called a "Forest"

Theorem 1:

In a tree there is One and only pair b/w every pair of vertices.

Theorem 2:

If any graph G there is one and only pair b/w every pair of vertices then G is a tree

Theorem 3:

A tree with n vertices has $n-1$ edges

Note

→ A pendant vertex of a tree is also called a leaf

Spanning trees

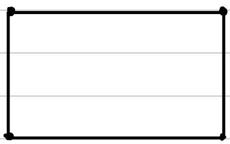
Let G be a connected graph, a sub-graph T of G is called a spanning of G if

i) T is a tree

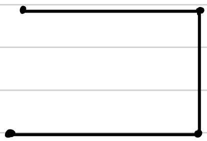
ii) T contains all vertices of G

iii) The edges of a spanning tree are called its branches

Find all the spanning tree of the graph given below :-

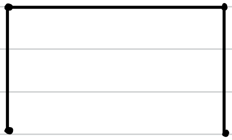


G

i) 

ii) 

iii) 

iv) 

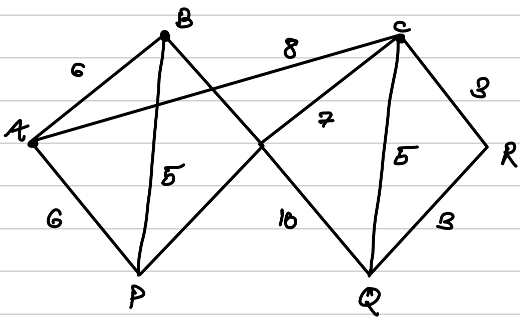
Minimal Spanning Tree

Let G be a graph and suppose there is a +ve real number associated with each edge of G then G is called weighted graph

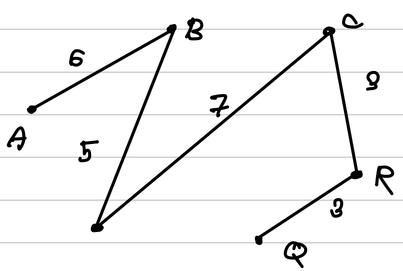
A spanning tree whose weight is the least is called a minimal spanning tree (This tree is not unique)

Kruskal's Algorithm for Minimal Spanning tree

Q. Using KA find a minimal spanning tree of the following weighted graphs



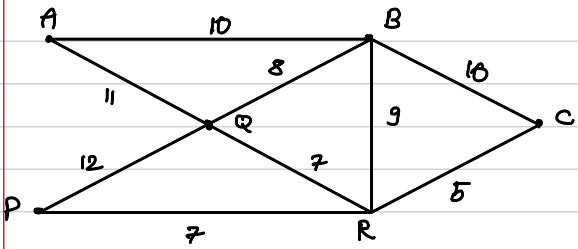
tree

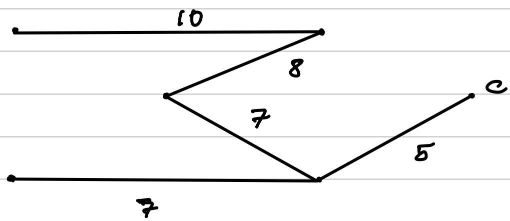


Total weight : 24

Edges	AB	BQ	QR	RC	CP	AP	BP
weight	6	10	3	3	7	6	5
select	✓	✗	✓	✓	✓	✗	✓

2)





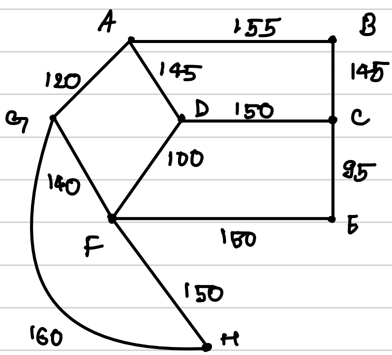
Total weight
= 87

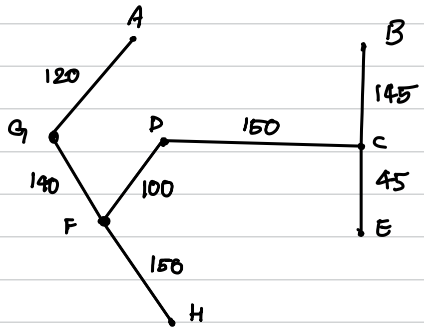
Edges	AB	BC	BR	RC	AQ	QR	PR	PQ	BQ
weight	10	10	9	5	11	7	7	12	8
selected	✓	✗	✗	✓	✗	✓	✓	✗	✓

Q. A, B, C, D, E, F, G, H are required to be connected........ are summarized in the following table.

Track b/w	Cost
A and B	155
A and D	145
A and G	120
B and C	145
C and D	150
C and E	95
D and F	100
E and F	150
F and G	140
F and H	150
G and H	160

Determine a railway network of minimal cost that connects all the cities





total cost = 900

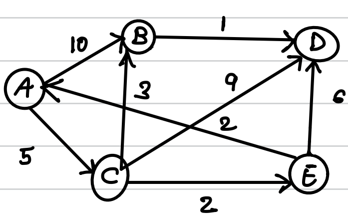
Edges	AB	BC	CD	AD	AG	GF	FH	FE	FD	CE
Weight	155	145	150	145	120	140	150	180	100	95
Selected										

Dijkstra's Algorithm

Let $D = D(V, E)$, $V = \{1, 2, 3 \ldots\ldots n\}$ is a vertex set, D a weighted, directed network in which the weight of every directed edge is non-negative

If $d(i) + c(i,j) < d(j)$ then $d(j) = d(i) + c(i,j)$

Q. Find the shortest path from A to the remaining vertices



P. T. O $\longrightarrow$

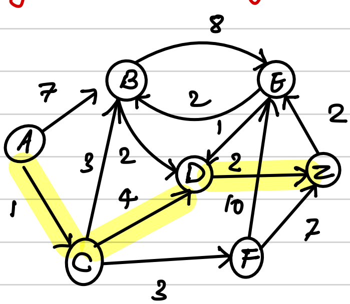
	A	B	C	D	E
A	$\boxed{0}$	∞	∞	∞	∞
C		10	$\boxed{5}$	∞	∞
E		8		14	$\boxed{7}$
B		$\boxed{8}$		13	
D				$\boxed{9}$	

ACB
AC
CBD

Shortest path and shortest distance

		path
A to B	— 8	ACB
A to C	— 5	AC
A to D	— 9	ACBD
A to E	— 7	ACE

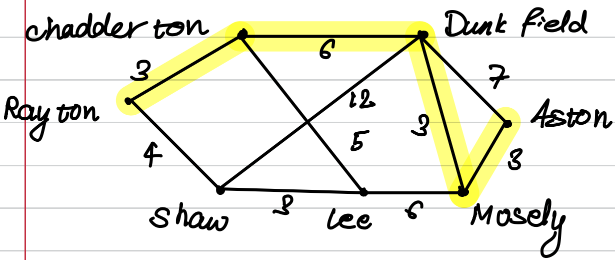
2) Find the shortest path b/w vertices A and Z and it weight using Djikstra's algorithm



	A	B	C	D	E	F	Z
A	$\boxed{0}$	∞	∞	∞	∞	∞	∞
B		7	$\boxed{1}$	∞	∞	∞	∞
C		$\boxed{4}$		5	12	4	∞
D				5	12	$\boxed{4}$	11
E				$\boxed{5}$	12		$\boxed{7}$
F					$\boxed{9}$		
Z							

This weight of the shortest path from A to Z is 7 and the path is A - C - D - Z

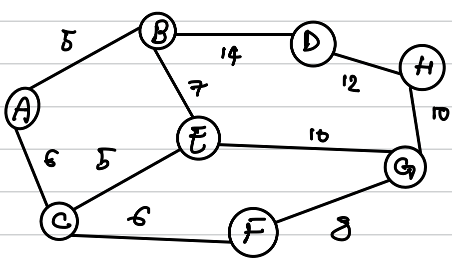
3. The diagram below shows roads connecting villages near Royton. The numbers on each arc /edge represents distance in miles, Leon leaves in royton and works in Aston. Use Djikstra's algorithm to find minimal distance



	R	S	C	D	L	M	A
R	$\boxed{0}$	∞	∞	∞	∞	∞	∞
S		4	$\boxed{3}$	∞	∞	∞	∞
C		$\boxed{4}$		9	8	∞	∞
D				9	$\boxed{7}$	∞	∞
L				$\boxed{9}$		13	∞
M						$\boxed{12}$	16
A							$\boxed{15}$

Shortest path from Royton to Aston is R–C–D–M–A and this distance is 15 miles

4. Find the shortest path b/w A to all other vertices



P. T. O →

	A	B	C	D	E	F	G	H
A	$\boxed{0}$	∞	∞	∞	∞	∞	∞	∞
B		$\boxed{5}$	6	∞	∞	∞	∞	∞
C			$\boxed{6}$	19	12	∞	∞	∞
D				19	$\boxed{11}$	12	∞	∞
E				19		$\boxed{12}$	21	∞
F				$\boxed{19}$			20	∞
G							$\boxed{20}$	31
H								$\boxed{30}$

1) A – B → 5
2) A – C → 6
3) A – B – D → 19
4) A – C – E → 11
5) A – C – F → 12
6) A – C – F – G → 20
7) A – C – F – G – H → 30

Theorem 1: A connected graph G has Euler circuit iff all the vertices of G are of even degree.

Theorem 2: Show that a connected graph with exactly two vertices of odd degree has an euler trail.

proof: Let A and B are the only vertices of odd degree in G.
Add an edge e b/w A & B then A and B be two vertices of even degrees and all other vertices of even degrees.
Now, we have graph
$G_1 = G \vee e$

∴ G is connected graph with even vertex is of even degree

⇒ G is a Euler graph

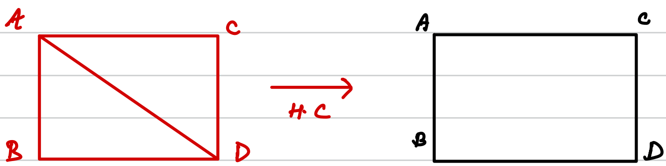
⇒ G contains Euler's circuit

If we remove e from Euler's circuit we get Euler's trail and that is the proof of G

Hamiltonian Cycle

A cycle of a connected graph G is said to be Hamiltonian cycle if it contains all the vertices of G.

Example:



Hamiltonian graph

A connected graph G is said to be Hamiltonial graph if it contains hamiltonian graph.

Theorem 1:

If in a simple connected graph with n vertices $(n \geq 3)$ the sum of degree of every pair of non-adjacent vertices is $\geq n$, then the graph is Hamiltonian

Theorem 2:

If in a simple connected graph with n vertices $(n \geq 3)$ the degree of every vertex $\geq n/2$ then the graph is Hamiltonian.

Theorem 3 :

Let G be a simple graph with n vertices and m edges $(m \geq 3)$. If $m \geq \frac{1}{2}(n-1)(n-2)+2$ then G is Hamiltonian graph.

soln : Let u and v be two non-adjacent vertices and $\deg(u) = x$ and $\deg(v) = y$

If we delete u & v from G, we get a graph G_1 with q edges (say)

$\therefore G_1$ is a simple graph with $n-2$ vertices then $q \leq {}^{n-2}C_2 \leq \frac{1}{2}(n-2)(n-3)$ $-q \geq -\frac{1}{2}(n-2)(n-3) \longrightarrow (1)$

Since u and v are non-adjacent vertices

$$m = q + x + y$$

$$\Rightarrow x+y = m - q$$

$$x+y \geq \frac{1}{2}(n-1)(n-2) + 2 - \frac{1}{2}(n-2)(n-3)$$

$$\geq \frac{1}{2}(n-2)(\cancel{n}-1-\cancel{n}+3)+2$$

$$\geq \frac{1}{\cancel{2}}(n-2)(\cancel{2})$$

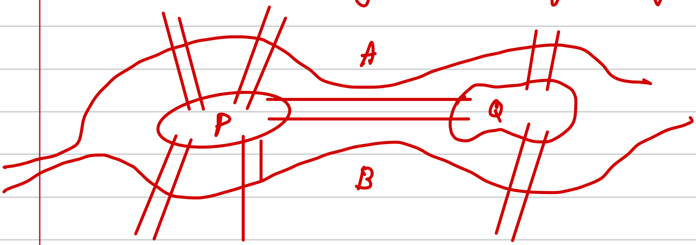
$$\geq n-2+2$$

$$x+y \geq n$$

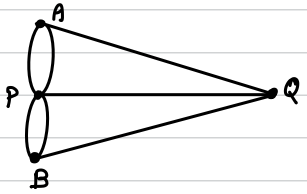
Sum of degree of two non-adjacent vertices $\geq n$

$\Rightarrow G$ is Hamiltonian.

Q. Consider a diagram (konigs bridge)

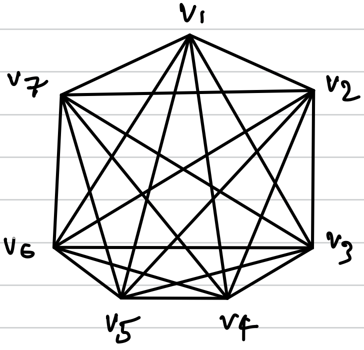


Is it possible to travel through each bridge exactly once by starting and ending at same point.

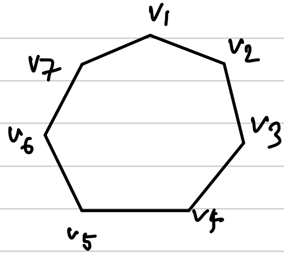


Since G is connected and deg (A) = deg (B) = 3 (odd)
Therefore there is no Euler's circuit in G

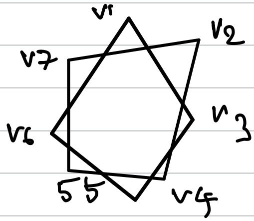
Q. Suppose a committee member has 7 members, these members meet each-day at a round table for lunch. They decide to sit in such a way that, every member has different neighbours at each lunch. How many day can the arrangement last?



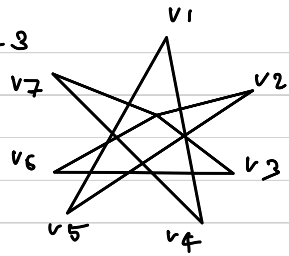
case - 1



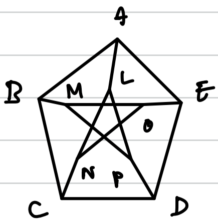
case 2 :



case 3



3. Show that petersoo graph has no Hamilton cycle



$p \to q$

$\sim p \not\to \sim q$

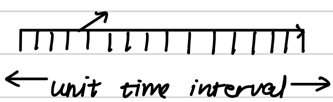
$\sim q \to \sim p$

Unit - 2

Probability Distribution

Poisson Distribution

$P\begin{pmatrix}\text{occurence of event} \\ \text{at this time}\end{pmatrix} = p$



- W Δt be the length each sub-interval $n \cdot \Delta t = 1$
- Let k be the number of events occured
- Average number of events in the unit interval is $\lambda = np$
- probability of occuring of an event is not depending on probability of occuring of other event

From the binomial distributions

$$p(k) = {}^{n}C_{k}\, p^{k} (1-p)^{n-k}$$
$$= \frac{n!}{(n-k)!\,k!} \left(\frac{\lambda}{n}\right)^{k} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n(n-1)\cdots\cdots(n-(k-1))(n-k)!}{(n-k)!\,k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{1}{n}\right)^{n-k}$$

$$P(k) = \frac{\cancel{n^k}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{k!}\left(\frac{\lambda}{\cancel{n}}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

as $n \longrightarrow k$

$$P(k) = \lim_{n\to\infty}\frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\lambda^k\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}}{k!}$$

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

This is called poisson distribution. | $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n \cdot e^{-\lambda}$

Note that,

$$\sum_{k=0}^{\infty} P(k) = \sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}\times e^{\lambda} = \underline{1}$$

Note: The interval that was partitioned is a time interval. The same method can be applied to any length, an area, or volume.

Generally.

If T is the time interval and k is the random variable then;

$$P(k:\lambda T) = \frac{(\lambda T)^k e^{-\lambda T}}{k!}\ ,\ k = 0,1,2\ldots\ldots$$

Here also,

$$\sum_{k=0}^{\infty} P(k;\lambda T) = \sum_{k=0}^{\infty}\frac{(\lambda T)^k e^{-\lambda T}}{k!} = e^{-\lambda T}\sum_{k=0}^{\infty}\frac{(\lambda T)^k}{k!}$$

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

$$= 1$$

Let us consider the time interval $T=1$ then

1) $$\text{variance}(k) = \sum_{k=0}^{\infty} k^2 p(k;\lambda) - \lambda^2$$

$$= \sum_{k=0}^{\infty} \frac{k^2 \lambda^k e^{-\lambda}}{k!} - \lambda^2$$

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

$$= \sum_{k=1}^{\infty} \frac{(k-1+1)\lambda^k e^{-\lambda}}{(k-1)!} - \lambda^2$$

$$= \sum_{k=1}^{\infty} \frac{k-1}{(k-1)!} \lambda^k e^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} - \lambda^2$$

$$\text{var}(k) = \sum_{k=2}^{\infty} \frac{(k-1)\lambda^k e^{-\lambda}}{(k-2)!\,(k-1)} + \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} - \lambda^2$$

$$= \sum_{k=0}^{\infty} \frac{\lambda^{k+2} e^{-\lambda}}{k!} + \sum_{k=0}^{\infty} \frac{\lambda^{k+1} e^{-\lambda}}{k!} - \lambda^2$$

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

$$= \lambda^2 \sum P(k) + \lambda \sum P(k) - \lambda^2$$

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

$$= \lambda$$

2. Expectations

$$E[k] = \sum_{k=0}^{\infty} k\, P(k;\lambda)$$

$\vdots$

$$= \lambda$$

$$E[k] = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!}$$

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

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

Probability Mass Function.

Is a function that gives probability of discrete random variable.

For the poisson distribution PMF is given by $P(k=x) = \frac{e^{-\lambda} \lambda^x}{x!}$

$P(k=x) = P(x)$

where $\lambda = np$

Consider a frequency distribution table

x	0	1	2	3	.	.	.	.	.	.
freq	f_1	f_2	f_3	f_4	.	.	.	.	.	.

The theoretical frequency of x is given by $P(x)$.

$$\text{Total freq} = P(x) \sum f_i = \frac{e^{-\lambda} \lambda^x}{x!} \sum f_i$$

where $\text{mean} = \lambda = \frac{\sum x_i f_i}{\sum f_i}$

Q. Given average accidents/week is $\lambda = 3$

WKT probability mass function.

$$p(k=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$P(k=2) = \frac{e^{-3}3^{2}}{2!}$$

$$= 0.224$$

i.e 22.4%

Q. A distributor of bean seeds determines from extensive tests that 5% of large batch of seeds will not germinate. He sells the seeds in pkts of 200 and guarantees 98% germination. Determine the prob that a particular pkt will violate the guarantee.

→ Given that $p = \frac{5}{100}$

and $n = 200$

$\therefore$ Mean $= \lambda = n \cdot p = 200 \cdot \frac{5}{100}$

$$\lambda = 100$$

Thus PMF is:

$$p(k=x) = \frac{e^{-\lambda}\lambda^{x}}{x!} = p(x)$$

$$\therefore p(x>2) = 1 - P(x \leq 2)$$

$$= 1 - (P(0) + P(1) + P(2))$$

$$= 1 - (e^{-10} + e^{-10} \times 10 + e^{-10} \times \frac{10^{2}}{2})$$

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

$$= p(x>2) = 0.9972$$

Q. A book contains 100 misprints distributed randomly throughout its 100 pages, assuming Poisson distribution find the prob. at random contain at least 2 misprints in a Page

Given 100 misprints in 100 pages

$$\therefore \text{Mean} = \lambda = \frac{100}{100} = 1$$

PMF is given by

$$P(k=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$P(x \geq 2) = 1 - p(x<2)$$
$$= 1 - (P(0) + P(1))$$

$$P(x \geq 2) = 1 - (e^{-1} + e^{-1})$$
$$= 1 - 2e^{-1}$$
$$= 0.264$$

probability at atleast two misprints /page $= 0.264$.

Q. The probability that a news reader commits no mistake in reading the news is e^{-3}. Find the prob. that on particular news broadcast he commits :

i) only 2 mistakes

ii) more than 3 mistakes

iii) utmost 3 mistakes

Given $P(x=0) = e^{-3}$

$$e^{-\lambda} = e^{-3}$$
$$\lambda = 3 = \text{Mean}$$

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

i) probability of only two mistakes is

$$P(2) = \frac{e^{-3}\lambda^{2}}{2!} = 0.2240$$

ii) probability of more than 3 mistakes

$p(x > 3) = 1 - P(x \leq 3)$

$p(x>3) = 1 - (p(0) + p(1) + p(2) + p(3))$

$= 1 - \left(e^{-3} + e^{-3} \times 3 + e^{-3} \cdot \frac{3^2}{2} + e^{-3} \cdot \frac{3^3}{3!}\right)$

$= 1 - e^{-3}\left(1 + 3 + \frac{3^2}{2} + \frac{3^3}{6}\right)$

$= 1 - 0.6472$

$= 0.3527$

iii) probability of atmost 3 mistakes

$p(x \leq 3) = 1 - P(x>3)$

$= 0.6472$

Q. 2% of the fuses manufactured by a firm are found to be defective. Find the prob that a box containing 200 fuses contains:

i) No defective fuses

ii 3 or more defective fuses.

Given $p = \frac{2}{100}$ & $n = 200$

$\therefore$ Mean $= \lambda = np = 4$

i) probability of no defective fuses is:

$P(x=0) = e^{-4} = 0.018$

ii) probability of 3 or more defective fuses

$p(x \geq 3) = 1 - P(x<3)$

$p(x \geq 3) = 1 - (p(0) + p(1) + p(2))$

$= 1 - \left(e^{-4} + e^{-4} \cdot 4 + e^{-4} \cdot \frac{4^2}{2}\right)$

$$= 1 - e^{-4}\left(1 + 4 + \frac{4^2}{2}\right)$$
$$= 0.7618$$

6. Given $P(x=2) = \frac{2}{3}\, p(x=1)$ then find $p(x=0)$ & $p(x=3)$

The probability mass function is:
$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Given;
$$p(x=2) = \frac{2}{3}\, p(x=1)$$
$$\frac{\cancel{e^{-\lambda}}\lambda^2}{2!} = \frac{2}{3}\,\cancel{e^{-\lambda}}\frac{\lambda}{1!}$$
$$\lambda = \frac{4}{3}$$

i) $P(x=0) = \dfrac{e^{-\lambda}\lambda^0}{0!} = e^{-\frac{4}{3}} =$

ii) $p(x=3) = \dfrac{e^{-4/3}\left(\frac{4}{3}\right)^3}{3!}$
$$= 0.104$$

Q. If x is the poisson variable s.t $p(x=2) = 9\, p(x=4) + 90\, p(x=6)$ then find mean and variable of the P.D.

$$p(x=2) = 9\,P(x=4) + 90\,P(x=6)$$
$$\frac{e^{-\lambda}\lambda^2}{2!} = \frac{9e^{-\lambda}\lambda^4}{4!} + \frac{90 \cdot e^{-\lambda}\lambda^6}{6!}$$
$$\frac{\lambda^2}{2!} = \lambda^4\left(\frac{9}{4!} + \frac{90}{6!}\lambda^2\right)$$
$$\frac{1}{2!} = \lambda^2\left(\frac{9}{4!} + \frac{90}{8!}\lambda^2\right)$$

$\lambda = 1 = \text{mean}$

variance $= \lambda = 1$ //

9.

x	0	1	2	3	4
f	111	63	22	3	1

ans :

$$\text{Mean} = \lambda = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{12\cancel{0}}{20\cancel{0}}$$

$$\lambda = 0.6$$

x	f	xf
0	111	0
1	63	63
2	22	44
3	3	9
4	1	4

$\Sigma f_i = 200$ $\quad \Sigma x_i f_i = 120$

The poisson fit is PMF

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-0.6}(0.6)^{x}}{x!}$$

Theoretical frequency of

$x = p(x).\Sigma f_i$

Theoretical frequency of x :

when $x=0 \longrightarrow p(0).\Sigma f_i = 109.76$

$x=1 \longrightarrow p(1).\Sigma f_i = 65.85$

$x=2 \longrightarrow p(2).\Sigma f_i = 19.75$

$x=3 \longrightarrow p(3).\Sigma f_i = 3.45$

$x=4 \longrightarrow P(4)\,\Sigma f_i = 0.59$

10. The frequency of accidents per shift in a factory is as shown in the following.

Accidents

Geometric Distribution.

→ Geometric distribution models the number of trials required for the first success in a series of independent Bernoulli's trials

1) Let x be the number of trials
2) p be the probability of success
3) q be the probability of failure

Thus the PMF for the geometric distribution is :

$$P(X=x) = q^{x-1} \cdot p = p(1-p)^{x-1}$$

where $p+q = 1$.

Q. The probability that a phone call leads to a sale is 0.4. Calculate the prob. that first sale occurs in the 5^{th} call.

→ Given, $p = 0.4$

$\therefore q = 1-p = 1-0.4 = 0.6$

The probability that such occurs in the 5^{th} call is $P(x=5)$

WKT,

PMF

$$p(x) = q^{x-1} \cdot p$$

$$P(x=5) = (0.6)^{5-1}(0.4) = (0.6)^4 (0.4) = 0.051$$

2Q. In a certain town, 30% of the people own an iphone, a researcher asks people at random whether they own an iphone. The random variable X represents the number of people asked upto and including the first person who owns an iphone. Determine the chances that:

i) 4 person owns a iphone

ii) at least 3 people own an iphone

iii) at most 4 people own an iphone

WKT PMF is $p(x) = q^{x-1} p$

Given $p = 30\% = \frac{30}{100} = 0.3$

$\therefore q = 1 - p = 0.7$

i) $p(x=4) = (0.7)^3 (0.3) = 0.1025$

5Q. A person decides to continue placing a bet of Rs 5000 on the number 5 in consecutive spins of a roulette wheel until he wins. On any spin there is a 1 on 50 chances that the ball would land on number 5.

i) How many spins do you expect until he wins?

ii) What is the amount he is expected to spend until he has his first win?

iii) What are the chances that it takes 5 spins before he wins?

iv) What are the chances that it would take him more than 50 chances to win.

Note: For geometric distribution

Expectan of x is

$E(x) = 1/p$

variance of x is $var(x) = \frac{q}{p^2}$

Sol:- Given $p = \frac{1}{50}$

Number of spins expect until he wins is :

$$E(x) = \frac{1}{p} = \frac{1}{1/50} = 50$$

$$E(x) = 50$$

ii) The amount he is expected to spend until he win is :

$$50 \times 5000 = 250000$$

Note: The PMF is :

$$P(x) = q^{x-1} \cdot p = (1-p)^{x-1} \cdot p$$
$$= (1 - 1/50)^{x-1} \frac{1}{50}$$

$$P(x) = \left(\frac{49}{50}\right)^{x-1} \times \frac{1}{50}$$

iii) This probability that he wins in 6th attempt is :

$$P(x=6) = \left(\frac{49}{10}\right)^{6-1} \frac{1}{50}$$

$$= 0.0180$$

iv) $P(x > 50) = 1 - P(x \leq 50)$

$$= 1 - (P(x=1) + p(x=2) + \ldots\ldots + p(x=50))$$
$$= 1 - \left(\frac{1}{50} + \frac{49}{50} \times \frac{1}{50} + \left(\frac{45}{50}\right)^2 \times \frac{1}{50} + \ldots\ldots \left(\frac{49}{50}\right)^{49} \times \frac{1}{50}\right)$$

This is G P with common ratio $r = \frac{49}{50}$

$$4a = \frac{1}{50}$$

$$\text{sum} = a\left(\frac{1-r^{n-1}}{1-r}\right)$$

$$= 1 - (0.6283) = 0.371$$

Q. A swing trader has a trading set of width 55% chance of success. He started trading with initial capital 1 lakh rupees by making 5% . . . . Find the no. of consecutive failed trades.

Given $P = 55\% = 0.55$

Number of consecutive failed trades required to loose 40% initial capital is 10

```
x = 1 lakh
for i = 1:100
   y = x - 5%(x)
   x = y
if abs(x-y) < 10^-5
   break
end
   point(i)
```

ii) WKT $p(x) = p \cdot q^{x-1}$

$= p(1-p)^{x-1}$

$\therefore p(x=10) = 0.55(1-0.55)^{10}$

$= 1.87 \times 10^{-4}$

Q. To finish a board game A needs a sum of 4 with two dice. What is the prob. that it takes 4 under 5 tries to win? How many rolls would you expect A to take until she wins?

Continuous probability distribution :

For a continuous R.V. x, the continuous function $f(x)$ is said to be probability density if it satisfies the following

i) $f(x) \geq 0 \quad \forall x$

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

iii) $P(a \leq x \leq b) = \int_a^b f(x)\, dx$

area under the curve $y = f(x)$ b/w ordinates a & b

Note :

i) $P(x = a)$

$$= \lim_{\Delta x \to 0} \int_{a-\Delta x}^{a+\Delta x} f(x)\, dx$$

$$= \lim_{\Delta x \to 0} \left(F(x) \Big|_{a-\Delta x}^{a+\Delta x} \right) \quad \Big/ \int f(x)\, dx = F(x)$$

$$= \lim_{\Delta x \to 0} \left(F(a+\Delta x) - F(a-\Delta x) \right)$$

$$= F(a) - F(a)$$

$$= 0$$

ii) $P(a \leq x \leq b) = P(a \leq x < b) = P(a < x \leq b) = P(a < x < b)$

$$\int_a^b f(x)\, dx \quad x \in [a, b]$$

$$\int_a^b f(x)\, dx \quad x \in (a, b)$$

Cumulative distributive function (CDF)

For a random variable x with the PDF $f(x)$ the CDF $(F(x))$ is defined as :

P.T.O $\longrightarrow$

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

$$F(-\infty) = 0, \quad F(+\infty) = 1$$

$$\therefore 0 \leq F(x) \leq 1$$

Expectation of x

→ Expectation of x or mean is given by

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

Variance

Variance is given by :

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

Uniform Distribution (Rectangular Distribution)

The function $f(x)$ is defined as

$$f(x) = \begin{cases} k & x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

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

$$\int_a^b f(x)\,dx = 1$$

$$\int_a^b k\,dx = 1$$

$$k\,(x)_a^b = 1$$

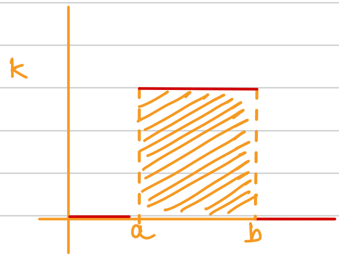
$$k(b-a) = 1$$

$$k = \frac{1}{(b-a)}$$

Thus,

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

This is called PDF for the uniform distribution



Expectation (Mean) for the uniform Distribution

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

$$= \frac{1}{b-a} \int_a^b x \, dx$$

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

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

$$= \frac{a+b}{2}$$

Variance for uniform distribution

$$\text{variance} = \sigma^2 = \int_a^b (x-\mu)^2 f(x)\, dx$$

$$= \frac{1}{b-a} \int_a^b (x-\mu)^2 dx$$

$$= \frac{1}{b-a} \left(\frac{(x-\mu)^3}{3} \right)_a^b$$

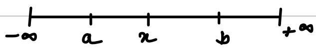
$$= \frac{1}{3(b-a)} \left((b-\mu)^3 - (a-\mu)^3 \right)$$

$$= \frac{1}{3(b-a)} \left(\left(b - \frac{a+b}{2}\right)^3 - \left(a - \frac{a+b}{2}\right)^3 \right) \quad = \frac{(b-a)^2}{12}$$

Cumulative distribution

i) when $x \le a$, $\int_{-\infty}^{x} f(x)dx = 0$

ii) when $x \in (a,b)$



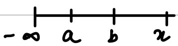
$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

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

$$F(x) = P(X \le x) = \Big|_{a}^{x} f(x)dx$$

$$= \frac{1}{b-a}\int_{a}^{x} dx = \frac{x-a}{b-a}$$

when $x \ge b$



$$F(x) = P(x \ge b) = \cancel{\int_{-\infty}^{a} f(x)dx} \rightarrow 0 + \int_{a}^{b} f(x)dx + \int_{b}^{x} f(t)\,dt$$

$$= \frac{1}{b-a}\Big|_{a}^{b} dx + 0 = 1$$

Q 1) If x is uniformly distributed in $-2 \le x \le 2$ then find:

i) $p(x < 1)$

ii) $p(|x-1| \ge 1/2)$

soln i) WKT $P(X \le x) = \frac{x-a}{b-a}$

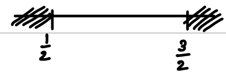
$$P(X < 1) = \frac{1-(-2)}{2+2}$$

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

ii) $|x-1| \ge \frac{1}{2} \Rightarrow x-1 \ge \frac{1}{2}$ or

$x-1 \le -1/2$

$x \ge \frac{3}{2}$ or $x \le \frac{1}{2}$



$P\left(x \ge \frac{3}{2} \text{ or } x \le \frac{1}{2}\right)$

$$= P\left(x \geq \frac{3}{2}\right) + P\left(x \leq \frac{1}{2}\right)$$

$$= 1 - P\left(x \leq \frac{3}{2}\right) + P\left(x \leq \frac{1}{2}\right)$$

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

$$= 1 - \frac{\frac{3}{2}+2}{4} + \frac{\frac{1}{2}+2}{4} = 1 - \frac{7}{8} + \frac{5}{8} = \frac{6}{8} = \frac{3}{4}$$

Q 2) If X is U.D with mean 1 and variance 4/3 then find $P(X < 0)$.

soln: WKT, mean $= \frac{b+a}{2} = 1 \Rightarrow a+b = 2 \longrightarrow$ ①

variance $= \sigma^2 = \frac{(b-a)^2}{12} = \frac{4}{3}$

$$(b-a)^2 = 4^2$$

$$b - a = \pm 4 \longrightarrow ②$$

solving ① & ②, we get

$a = -1$ and $b = 3$

$$\therefore P(X < x) = \frac{x-a}{b-a}$$

$$P(X < 0) = \frac{0-(-1)}{4} = \frac{1}{4}$$

Q 3) The driving time X from a person's home to the train station is U.D as $U[10, 50]$. If it takes 2 minutes to board the train determine the prob. that the person catches the 7 am train if he starts at 6:43 am from his home.

soln: $\longrightarrow$ P.T.O

soln; $P(x \le 15) = \dfrac{15-a}{b-a}$

$$= \frac{15-10}{40}$$

$$= \frac{5}{40} = \frac{1}{8}$$

Q 11) The amount charged for a visit to a dental clinic is U.D from 0 to 1000 (in INR). Given that the amount charged for a visit exceeds Rs 500, calculate the probability that it exceeds Rs 750.

soln; WKT the conditional probability of happening of an event A when B is already happened is:

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

$$P\left(\text{exceeds } 700/_{500}\right) = \frac{P(\text{exceeding } 750 \cap \text{exceeding } 500)}{P(\text{exceeding } 500)}$$

$$= \frac{P(x \ge 750)}{P(x \ge 500)}$$

$$= \frac{1 - P(x \le 750)}{1 - P(x \le 500)}$$

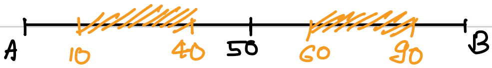
$$= \left(1 - \frac{750-0}{1000}\right) \Big/ \left(1 - \frac{500}{1000}\right)$$

$$= 0.5$$

6. A bus travels between two cities A and B which are 100 miles apart. If the bus has a breakdown, the distance X of the point of breakdown from the city A has a uniform distribution $U[0,100]$. There are service garages in the city A, city B and midway between the two cities such that in case of a breakdown a tow truck is sent from the garage nearest to the point of breakdown.
   (i) What is the probability that the tow truck has to travel more than 10 miles to reach the bus?
   (ii) Would it be more "efficient" if the three service garages were placed at 25, 50 and 75 miles from city A, apart from service garages at city A and city B?

ans

i)



$P(10 < x < 40 \text{ or } 60 < x < 90)$

$= P(10 < x < 40) + P(60 < x < 90)$

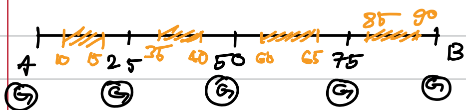
$= \int_{10}^{40} f(x)\,dx + \int_{60}^{90} f(x)\,dx \quad \Big| \quad f(x) = \frac{1}{b-a} = \frac{1}{100}$

$= \frac{1}{100}\int_{10}^{40} dx + \frac{1}{100}\int_{60}^{90} dx$

$= \frac{1}{100} \times 30 + \frac{1}{100} \times 30$

$= \frac{60}{100} = \frac{3}{5}$ //

ii)



$= P(10 < x < 15) + P(35 < x < 40) + P(60 < x < 65) + P(85 < x < 90)$

$= \int_{10}^{15} f(x)\,dx + \int_{35}^{40} f(x)\,dx + \int_{60}^{65} f(x)\,dx + \int_{85}^{90} f(x)\,dx$

$= \frac{1}{100}\left[\int_{10}^{15} dx + \int_{35}^{40} dx + \int_{60}^{65} dx + \int_{85}^{90} dx\right]$

$= \frac{1}{100}\left[5+5+5+5\right]$

$= \frac{20}{100} = \frac{1}{5}$ //

$\therefore$ Since $\frac{1}{5} < \frac{3}{5}$,

hence secondroute (ii) is more efficient than (i). //

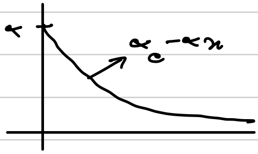
Exponential Distribution

For the continuous R.V x, the probability density function for the exponential distribution is

$$f(x) = \begin{cases} \alpha e^{-\alpha x}, & x \geq 0 \\ 0 & \text{Otherwise} \end{cases}$$

where

$\alpha > 0$



Mean (μ) or $E[x]$

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

$$= \int_{-\infty}^{0} x f(x)\,dx + \int_{0}^{\infty} x f(x)\,dx$$

$$= 0 + \int_{0}^{\infty} x\, \alpha e^{-\alpha x}\,dx$$

$$= \int_{0}^{\infty} x\, \alpha e^{-\alpha x}\,dx$$

$$\mu = \alpha \int_{0}^{\infty} \underset{u}{x}\, \underset{v}{e^{-\alpha x}}\,dx$$

$$= \alpha \left[\frac{x e^{-\alpha x}}{-\alpha} - \frac{1 \times e^{-\alpha x}}{(-\alpha)^2} \right]_0^{\infty}$$

$$= \alpha \left[0 - \left(0 - \frac{1}{\alpha^2} \right) \right]$$

$$\mu = \frac{1}{\alpha}$$

//

Variance

$$\sigma^2 = \int_{0}^{\infty} (x-\mu)^2 f(x)\,dx$$

$$= \int_0^\infty \left(x - \frac{1}{\alpha}\right) \alpha e^{-\alpha x} \, dx$$

$$= \alpha \left[\left(x - \frac{1}{\alpha}\right) \frac{e^{-\alpha x}}{-\alpha} - 1 \times \frac{e^{-\alpha x}}{(-\alpha)^2} \right]_0^\infty$$

$$= \alpha \left[-\frac{(x - 1/\alpha)^2 e^{-\alpha x}}{-\alpha} - 2\left(x - \frac{1}{\alpha}\right) \frac{e^{-\alpha x}}{(-\alpha)^2} + \frac{2 e^{-\alpha x}}{(-\alpha)^3} \right]_0^\infty$$

$$= \alpha \left[0 - \left(-\frac{1}{\alpha^3} + \frac{2}{\alpha^3} - \frac{2}{\alpha^3} \right) \right]$$

$$= \frac{1}{\alpha^2}$$

Cumulative distributions

The CDF of exponential distribution is:

$$F[x] = p(x \le x) = \int_0^x f(x) \, dx$$

$$= \int_0^x \alpha e^{-\alpha x} dx$$

$$= \alpha \left[\frac{e^{-\alpha x}}{-\alpha} \right]_0^x$$

$$= -\left[e^{-\alpha x} - 1 \right] =$$

1. In a certain town, the duration of shower is exponentially distributed with mean 5 minutes. What is the probability that the shower will last for (i) less than 10 minutes (ii) 10 minutes or more and (iii) between 10 and 12 minutes.

Given, $\mu = 5$ min

WKT, $\mu = \frac{1}{\alpha} = 5 \Rightarrow \alpha = \frac{1}{5}$

Let x is the duration of shower then;

i) $p(X < x) = 1 - e^{-\alpha x}$

$p(x < 10) = 1 - e^{-\frac{10}{5}}$

$= 1 - e^{-2}$

$= 0.8646$

ii) $p(x \geq 10) = 1 - P(x \leq 10)$

$= 1 - 0.8646$

$= 0.1353$

iii) $P(10 \leq x \leq 12)$

$$= \int_{10}^{12} f(x)\,dx$$

$$= \int_{10}^{12} \alpha e^{-\alpha x}\,dx$$

$$= \frac{1}{5}\int_{10}^{12} e^{-\frac{x}{5}}\,dx$$

$$= \frac{1}{5}\left[e^{-\frac{x}{5}} \right]_{10}^{12}$$

$$= -\left[e^{-\frac{12}{5}} - e^{-\frac{10}{5}} \right]$$

$$= e^{-2} - e^{-\frac{12}{5}}$$

$$= 0.0446$$

Q 5. The daily turnover in a medical shop is exponentially distributed with Rs.6000 as the average with a net profit of 8%. Find the probability that the net profit exceeds Rs.500 on a randomly chosen day.

soln; Average net profit is :

$\mu = 8\% \,(6000)$

$\mu = 480 = \frac{1}{\alpha}$

$\alpha = 1/480$

$P(x \geq 500) = 1 - P(x \leq 500)$

$= 1 - (1 - e^{-\alpha x})$

$= e^{-\frac{500}{480}}$

$= 0.3528$

Q 4. The sales per day in a shop is exponentially distributed with average sale amounting to Rs100/- and net profit is 8%. Find the probability that the net profit exceeds Rs. 30/- on 2 consecutive days.

soln; Given that on average net profit of each day is 8% (100) = 8

Average net profit on 2 consecutive days is 16.

Let x be the net profit on two consecutive days.

We need to find $p(x \geq 30)$.

$\lambda = 1/16$

WKT

$$p(X \geq x) = 1 - p(x \leq x) = 1 - (1 - e^{-\lambda x})$$

$$p(x \geq x) = e^{-\lambda x}$$

$$\therefore p(x \geq 30) = e^{-\frac{1}{16} \times 30}$$

$$= 0.156$$

Gamma distribution

For the continuous R.V. x the PDF is given by

$$f(x) = \begin{cases} \dfrac{x^{\alpha - 1} e^{-\frac{x}{\beta}}}{\beta^{\alpha} \Gamma(\alpha)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where,

$$\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx$$

α is called slope variable and β is called growth variable

Mean $= \mu = \alpha\beta$

variance $= \sigma^2 = \alpha\beta^2$

Cumulative distribution

$$p(x \leq x) = \int_0^{x} \frac{x^{\alpha - 1} e^{-x/\beta}}{\beta^{\alpha} \Gamma(x)} dx$$

put $\frac{x}{\beta} = t \Rightarrow dx = \beta dt$

$$p(x \leq x) = \int_0^{\frac{x}{\beta}} \frac{\beta^{\alpha-1} t^{\alpha-1} e^{-t}}{\beta^{\alpha} \Gamma(\alpha)} \beta dt$$

$$= \int_0^{\frac{x}{\beta}} \frac{t^{\alpha-1} e^{-t}}{\Gamma(\alpha)} dt$$

$$= \frac{1}{\Gamma(\alpha)} \Gamma\left(\alpha, \frac{x}{\beta}\right)$$

where $\Gamma\left(\alpha, \frac{x}{\beta}\right) = \int_0^{\frac{x}{\beta}} t^{\alpha-1} e^{-t} dt$

$\hookrightarrow$ incomplete gamma function.

Questions :

The daily consumption of electric power (in million Kw-hours) in a certain city is a random variable X having the probability density

$$f(x) = \begin{cases} \frac{1}{9} x e^{-x/3}, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases}$$

Find the probability that the power supply is adequate on any given day if the capacity of the power plant is 12 million Kw hours.

Soln; Comparing given PDF with standard PDF, we get $\beta = 3$ and $\alpha = 2$

$$P(x \leq x) = \frac{\Gamma(\alpha, x/\beta)}{\Gamma(\alpha)}$$

$$P(x \leq 12) = \frac{\Gamma(2, 12/3)}{\Gamma(2)}$$

$\Gamma(n+1) = n!$

$$= \frac{\Gamma(2, 4)}{1}$$

$\Gamma(2) = 1$

$$P(x \leq 12) = \Gamma(2, 4)$$

$$= \int_0^4 t^{2-1} e^{-t} dt$$

$$= \int_0^4 t e^{-t} dt$$

$$= \left(t \frac{(e^{-t})}{-1} - 1 \cdot \left(e^{-t}\right) \right)_0^4$$

$\Gamma(n, x)$

$$= \int_0^x t^{n-1} e^{-t} dt$$

$$\int uv = u \int v - u' \iint v$$

$$+ u'' \iiint v \ldots \ldots$$

R.V = random variable

$$= \frac{4e^{-4}}{-1} - e^{-4} - (0-1) = -5e^{-4}+1$$
$$= 0.908$$

Daily consumption of milk in a town in excess of 20,000 litres is approximately given by Gamma distribution with alpha = 3 and beta= 10,000. The town has a daily stock of 30,000 litres. Find the probability that the stock is insufficient on a given day.

soln; Let x (be R.V) daily consumption of milk in excess of 20,000 Litre.

Given $\alpha = 3$ and $\beta = 10,000$

If $x \geq 10,000$ thus stock is insufficienct

$p(x \geq 10000) = ?$

Now,

$$P(x \geq 10000) = 1 - P(x \leq 10000)$$
$$= 1 - \frac{\Gamma\left(\alpha, \frac{10000}{\beta}\right)}{\Gamma(\alpha)}$$
$$= 1 - \frac{\Gamma(3,1)}{\Gamma(3)}$$
$$= 1 - \int_0^1 \frac{t^{3-1} e^{-t} dt}{2!}$$
$$= 1 - \frac{1}{2}\int_0^1 t^2 e^{t} dt = 0.999$$

If a random variable has gamma distribution with alpha = 2 and Beta = 2 . Find (i) mean (ii) standard deviation (iii) the probability that *X* will take a value less than 4.

⇒ Given $\alpha = 2$ and $\beta = 2$

i) Mean $= \alpha\beta = 4$

ii) S.D $= \sqrt{var(x)} = \sqrt{\alpha\beta^2}$
$= \sqrt{8} = 2\sqrt{2}$

iii) $p(x<4) = \Gamma(\alpha, 4/\beta)$

$$= \frac{\Gamma(2,2)}{\Gamma(2)} = \Gamma(2,2)$$

$$p(x<4) = \int_0^2 t^{2-1} e^{-t} dt$$

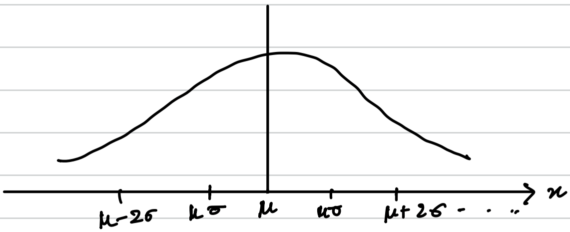
$$= \int_0^2 t e^{-t} dt$$

$$p(x<4) = 0.594$$

$$\Gamma(n,x) = \int_0^x e^{-t} t^{x-1} dt$$

Normal distribution

Thus R.V x with bell shaped distribution is called normal R.V.



Thus probability density function for the R-V x is $N(x:\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

where μ is mean & σ is S.D.

The probability of x lies b/w μ and x_2 is given below:

$$p(\mu \le x \le x_2) = \int_{\mu}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \longrightarrow ①$$

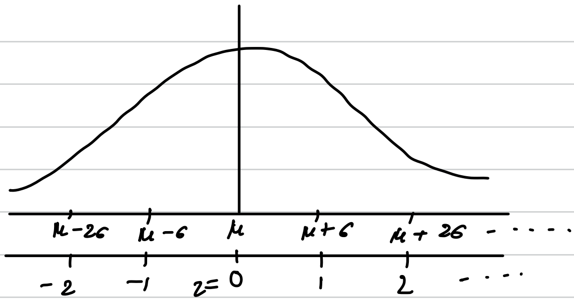
Thus R.V z with mean 0 and S.D 1 is called standard normal R.V

$z = \frac{x-\mu}{\sigma}$

$dz = \frac{dx}{\sigma}$

$dx = \sigma dz$

P.T.O $\longrightarrow$



① ⇒

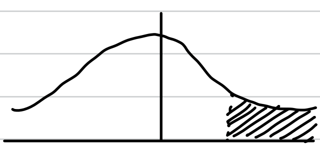
$$P(z_1 \le z \le z_2) = \int_{z_1}^{z_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \, dz$$

$$P(z_1 \le z \le z_2) = \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dz$$

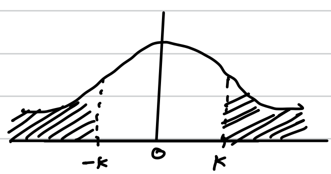
$$\boxed{z_1 = \frac{x_1-\mu}{\sigma} \qquad z_2 = \frac{x_2-\mu}{\sigma}}$$

Let $k > 0$, then

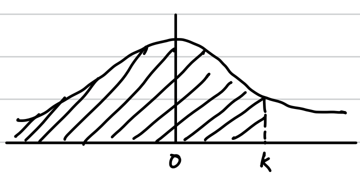
1) $P(z > k) = 0.5 - P(0 \le z \le k)$



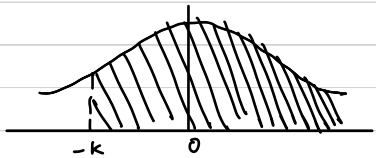
2) $P(z \le -k) = 0.5 - P(-k \le z \le 0)$
$= 0.5 - P(0 \le z \le k)$



3. $P(z \le k) = 0.5 + P(0 \le z \le k)$

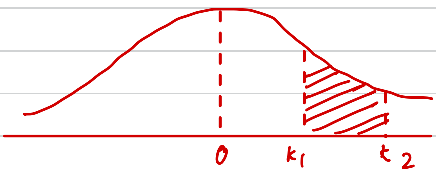


4 $p(-k \leq z) = 0.5 + p(-k \leq z \leq 0)$
$= 0.5 + p(0 \leq z \leq k)$

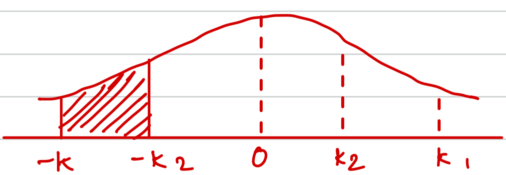


let $k_1 > 0$ and $k_2 > 0$ then

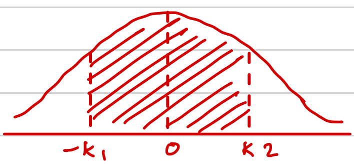
① $p(k_1 \leq z \leq k_2) = p(0 \leq z \leq k_2) - P(0 \leq z \leq k_1)$



2. $p(-k_1 \leq z \leq -k_2) = p(-k_1 \leq z \leq 0) - p(-k_2 \leq z \leq 0)$
$= p(0 \leq z \leq k) - p(0 \leq z \leq k_2)$

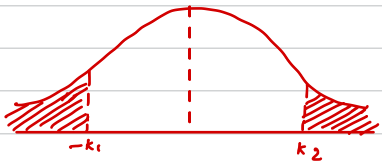


3. $p(-k_1 \leq z \leq k_2) = p(-k_1 \leq z \leq 0) + P(0 \leq z \leq k_2)$
$= p(0 \leq z \leq k_1) + p(0 \leq z \leq k_2)$



4. $P(z \le -k_1 \text{ and } z \ge k_2) = 1 - P(-k_1 \le z \le 0) - P(0 \le z \le k_2)$

$$= 1 - P(0 \le z \le k_1) - P(0 \le z \le k_2)$$

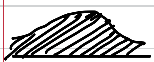


Properties of N·D

1) Mean $= \mu$

variance $= \sigma^2$

S·D $= \sigma$

2)  Area under the normal curve in the interval $(-\infty, \infty)$ is one

3) The curve is symmetric about $x = \mu$ $(z = 0)$

4) Area b/w $(-\infty, 0)$

$= $ Area b/w $(0, \infty) = 0.5$

Questions

1. In a normal distribution 31% of the items are under 45 and 8% are over 64. Find the mean and S·D of the distribution.

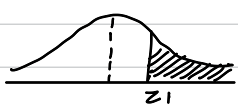
Soln; WKT, $z = \dfrac{x - \mu}{\sigma}$

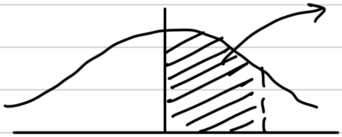
Given $P(x < 45) = 0.31$ $\quad P(x > 64) = 0.08$

Consider $P(x \le 45) = 0.31 < 0.5$ we need to find the value of z for which area b/w 0 to z is

$$= 0.5 - 0.31$$

$$= 0.19$$

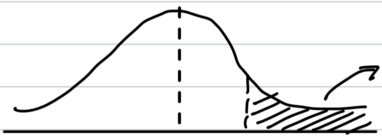




By table, $z_1 = 0.5$

Consider;

$$P(x > 64) = 0.08 < 0.5$$



We can find z_2 such that area b/w 0 to z_2 is $0.5 - 0.08$
$= 0.42$

By table,

$z_2 = 1.41$

WKT $z = \frac{x - \mu}{\sigma}$

at $z_1 \Rightarrow z_1 = \frac{x_1 - \mu}{\sigma} \Rightarrow 0.5 = \frac{45 - \mu}{\sigma}$

$\boxed{\mu + 0.5\sigma = 45} \longrightarrow$ 1

at $z_2 \Rightarrow z_2 = \frac{x_2 - \mu}{\sigma} \Rightarrow 1.41 = \frac{64 - \mu}{\sigma}$

$\boxed{\mu + 1.41\sigma = 64} \longrightarrow$ 2

solving 1 and 2, we get

$\mu = 34.56$ and $\sigma = 20.87$

2. In a university scholarship program, anyone with a grade point average over 3.5 receives a \$1,000 scholarship, anyone with an average between 3.0 and 3.5 receives \$500, anyone with an average between 2.5 and 3.0 receives \$100, and all others receive nothing. A particular class of 500 students has an overall average of 2.8 with a standard deviation of 0.6. Assuming the grade point averages are distributed normally, calculate the expected cost to the university of supplying scholarships for this class.

soln; Let x be the average grade point of a student

Given $\mu = 2.8$, $\sigma = 0.6$

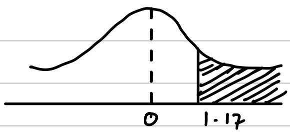
i) $P(x > 3.5) = P\left(z > \frac{3.5 - 2.8}{0.6}\right)$

$= P(z > 1.166) \simeq P(z > 1.17)$

$= 0.5 - P(0 < z < 1.17)$

$= 0.5 - 0.379$

$P(x > 3.5) = 0.121$



$\therefore$ # students average grade > 3.5

$= 0.12 \times 500 = 60.5 \simeq 61$

ii) $P(3 < x < 3.5)$

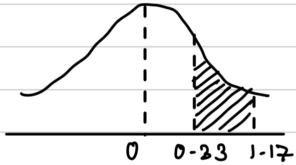
$= P\left(\frac{3 - 2.8}{0.6} < z < \frac{3.5 - 2.8}{0.6}\right)$

$= P(0.333 < z < 1.166)$

$\simeq P(0.33 < z < 1.17)$

$= P(0 < z < 1.17) - P(0 < z < 0.33)$

$= 0.379 - 0.129 = 0.25$



students average grade point b/w 3 and 3.5 $= 0.25 \times 500 = 125$

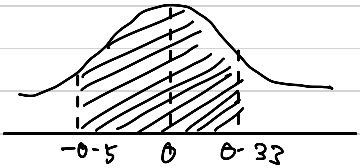
iii) $P(2.5 < x < 3) = P\left(\frac{2.5 - 2.8}{0.6} < z < \frac{3 - 2.8}{0.6}\right)$

$= P(-0.5 < z < 0.33)$

$= P(0 < z < 0.5) + P(0 < z < 0.33)$

$= 0.1915 + 0.1293$

$= 0.3208$



Students average grade point b/w 2.5 and 8 = 0.3208 × 500
= 160.4 ≃ 161

Expected cost = 61×1000 + 125×500 + 161×100
= $ 139600

3. The mean height of 500 students is 151 cm and the S.D. is 15 cm. Assuming that the heights are normally distributed, find how many student's heights lie between 120 and 155 cm. Also find 2 symmetrical values an and b such that P(a<=X<=b)=0.95.

soln; Let x be the height of a student

Given $\mu = 151$ cm and $\sigma = 15$ cm

$P(120 < x < 155)$

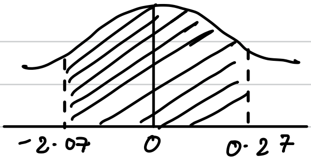
$= P\left(\frac{120-151}{15} < z < \frac{155-151}{15}\right)$

$P(120 < x < 155) = P(-2.07 < z < 0.27)$

$= P(0 < z < 2.07) + P(0 < z < 0.27)$

$= 0.4808 + 0.1064$

$= 0.5872$



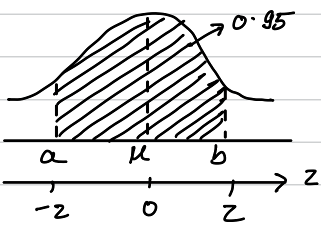
Students whom height is b/w 120 cm and 155 cm = 0.5872 × 500
= 294

We need to find a and b (symmetric)

$P(a < x < b) = 0.95$

$P(a < x < b)$

$= P\left(\frac{a-151}{15} < z < \frac{b-151}{15}\right)$



$$= 2P\left(0 < z < \frac{b-151}{15}\right) \qquad \left(\because\ a \text{ and } b \text{ are symmetric}\right)$$

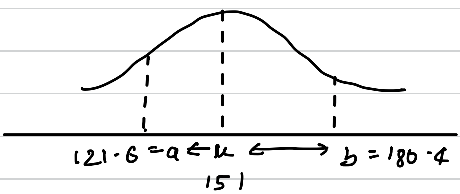
Given that

$$2P\left(0 < z < \frac{b-151}{15}\right) = P(a < x < b) = 0.95$$

$$P\left(0 < z < \frac{b+51}{15}\right) = \frac{0.95}{2} = 0.475$$

From the table

$$\frac{b-151}{15} = 1.96 \Rightarrow b = 180.4$$



WKT;

$$\mu - a = b - \mu$$

$$2\mu - b = a$$

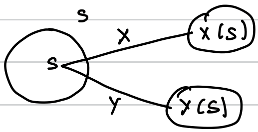
$$\therefore a = 2(151) - 180.4$$

$$\therefore a = 121.6$$

4. The mean inside diameter of a sample of 200 washers produced by a machine is 5.02 mm and the S.D. is 0.05 mm. the purpose for which these washers are intended allows a maximum tolerance in the diameter of 4.96 to 5.08 mm, otherwise the washers are considered defective. Determine the percentage of defective washers produced by the machine, assuming the diameters are normally distributed.

Unit - 3 : Joint Probability

Consider a probabilistic model with two discrete R.V X and Y with PMF's $P_X(X=x)$ and $P_Y(Y=y)$, respectively.



$$P_{X,Y}(X=x, Y=y)$$

For any $x \in X$ and $y \in Y$, the probability of occurence x and y denoted by $P_{X,Y}(x=X \text{ and } Y=y) = P_{X,Y}(X=x, X=y)$ is called Joint probability mass function (JPMF)

WKT,

i) $P_{X,Y}(x,y) \geq 0$ ii) $\sum_{i \geq 1}\sum_{j \geq 1} P_{X,Y}(x_i, y_j) = 1$

Suppose

$X = \{x_1, x_2, x_3, \ldots\ldots\}$

$Y = \{y_1, y_2, y_3, \ldots\ldots\}$

The joint probability table is given by:

X \ Y	y_1	y_2	y_3	y_4	...
x_1	P_{11}	P_{12}	P_{13}	P_{14}	....
x_2	P_{21}	P_{22}	P_{23}	P_{24}	.....
x_3	P_{21}	P_{32}	P_{33}	P_{34}	.....

where $P_{ij} = P_{X,Y}(x_i, y_j)$

once we have JPMF $P_{X,Y}(X=x, Y=y)$, the individual PMF's $P_X(x)$ and $P_Y(y)$ is called marginal PMF's.

$$P_X(x_i) = P_{i1} + P_{i2} + \ldots\ldots = \sum_{j \geq 1} P_{ij} = \sum_{j \geq 1} P_{X,Y}(x_i, y_j)$$

$$P_Y(y_j) = P_{1j} + P_{2j} + P_{3j} + \ldots\ldots = \sum_{i \geq 1} P_{ij} = \sum_{i \geq 1} P_{X,Y}(x_i, y_j)$$

Note : If x and y are independent R.V then :

$$P_{X,Y}(X=x, Y=y) = P_X(X=x) = P_Y(Y=y)$$

Expected values of $g(x,y)$:

$$E[g(x,y)] = \sum_x \sum_y g(x,y)\, P_{X,Y}(X=x, Y=y)$$

Expected values of X+Y

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

Covariance :

Covariance of X and Y is denoted by $\mathrm{cov}(x,y)$ and is defined by

$$\mathrm{Cov}(X,Y) = \sum_X \sum_Y (x-\mu_x)(y-\mu_y)\, P_{X,Y}(x,y)$$

where $\mu_x = E[X=x]$, $\mu_y = E[Y=y]$

$$\mathrm{Cov}(x,y) = E[(X-\mu_x)(Y-\mu_y)]$$

1) If $X=Y$ then

$$\mathrm{Cov}(X,x) = E[(X-\mu_x)^2] = \mathrm{var}(x)$$

2) $\mathrm{Cov}(X,Y) = E[X,Y] - E[X]E[Y]$

Correlation co-efficient of X and Y

For the R.V X and Y the correlation co-efficient is :

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_x \sigma_y}$$

where σ_x is S.D of x, σ_y is S.D of y.

1. The joint probability distribution of two random variables X and Y is given below:

X \ Y	2	3	4
1	0.06	0.15	0.09
2	0.14	0.35	0.21

Find the marginal distributions of X and Y. Also verify whether X and Y are stochastically independent.

Marginal distributions are :

$P_X(X=1) = 0.06 + 0.15 + 0.09 = 0.3$

$P_X(X=2) = 0.14 + 0.35 + 0.21 = 0.7$

$P_X(y=2) = 0.06 + 0.14 = 0.2$

$P_Y(Y=3) = 0.15 + 0.35 = 0.5$

$P_Y(Y=4) = 0.09 + 0.21 = 0.3$

$\forall\ X = x$ and $Y = y$

$P_{X,Y}(x,y) = P_X(x) \times P_X(y)$

$\therefore$ x and y are independent

1. $f(x,y) = c(2x+y)$

$0 \leq x \leq 2,\ 0 \leq y \leq 3$

$x, y \in Z$

X \ Y	0	1	2	3
0	0	C	2C	3C
1	2C	3C	4C	5C
2	4C	5C	6C	7C

WKT,

i) $\Sigma \Sigma P_{xy}(x,y) = 1$

$0 + C + 2C + 3C + 2C + 3C + 4C + 5C + 4C + 5C + 6C + 7C$

$42C = 1$

$C = \frac{1}{42}$

ii) $P(x \geq 1, y < 2)$

$= (2+3+4+5)C$

$= 14C = \frac{14}{42}$

iii) Marginal distribution of x and y

X	0	1	2
$P_X(x)$	6/42	14/42	22/42

Y	0	1	2	3
$P_Y(y)$	6/42	9/42	12/42	15/42

iv) x and y are independent if $P_{X,Y}(x,y) = P_X(x) P_Y(y) \ \forall x, y$

since,

$P_{X,Y}(0) = 0$ but

$P_X(0) \cdot P_Y(0) = \frac{6}{42} \cdot \frac{6}{42} \neq 0$

$\therefore$ X and Y are not independent.

6. A coin is tossed three times. Let X denote 0 or 1 according as tail or head occurs on the first
toss. Let Y denote the total number of tails which occur. Determine:
(i) the marginal distributions of X and Y
(ii) the joint distributions of X and Y. Also find the expected values of $X + Y$ and XY.

$x \to 0 \to$ Tail in first toss

$1 \to$ head in first toss

$Y \to$ no of tails

occurances	X	Y
HHH	1	0
HHT	1	1
HTH	1	1
HTT	1	2
THH	0	1
THT	0	2
TTH	0	2
TTT	0	3

joint probability

X \ Y	0	1	2	3
0	0	1/8	2/8	1/8
1	1/8	2/8	1/8	0

i) Marginal distribution of X and Y

X	0	1
$P_X(x)$	4/8	4/8

Y	0	1	2	3
$P_Y(y)$	1/8	3/8	3/8	1/8

ii) $E(X+Y) = \sum\sum (x+y) P_{x,y}(x,y)$

$$E[X+Y] = 0 + 1 \cdot \frac{1}{8} + 2 \cdot \frac{2}{8} + 3 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 2 \times \frac{2}{8} + 3 \cdot \frac{1}{8} + 4 \cdot 0$$

$$E[X+Y] = 2$$

$$E[X,Y] = \sum\sum xy\, P_{x,y}(x,y)$$

$$= 0 + 1 \cdot \frac{2}{8} + 2 \cdot \frac{1}{8} + 3 \cdot 0$$

$$= \frac{4}{8} = 0.5$$

7. Two flashcards are selected at random from a box which contains five cards numbered 1, 1, 2, 2 and 3. Find the joint distributions of X and Y where X denotes the sum of two numbers and Y denote the maximum of two numbers drawn. Also determine *Cov* (*X*, *Y*), and correlation coefficient of X and Y.

$\boxed{1}$ $\boxed{1}$ $\boxed{2}$ $\boxed{2}$ $\boxed{3}$

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

$X \rightarrow$ sum of two numbers

$Y \rightarrow$ Max of two numbers

X \ Y	1	2	3
2	1/10	0	0
3	0	4/10	0
4	0	1/10	2/10
5	0	6	2/10

ii) Marginal distributions of x and y

X	2	3	4	5
$P_X(x)$	1/10	4/10	2/10	2/10

y	1	2	3
$P_Y(y)$	1/10	5/10	4/10

$$E[x] = \Sigma x P_X(x)$$

$$E[x] = 2 \cdot \frac{1}{10} + 3 \cdot \frac{4}{10} + 4 \cdot \frac{3}{10} + 5 \cdot \frac{2}{10}$$

$$E[x] = \frac{36}{10} = 3.6$$

$$E[Y] = \Sigma y P_y(y) = 1 \cdot \frac{1}{10} + 2 \cdot \frac{5}{10} + 3 \cdot \frac{4}{10}$$

$$E[Y] = \frac{23}{10} = 2.3$$

$$E[XY] = \Sigma\Sigma xy \, P_{x,y}(x,y)$$

$$E[XY] = 2 \cdot \frac{1}{10} + 6 \cdot \frac{4}{10} + 8 \cdot \frac{1}{10} + 12 \cdot \frac{2}{10} + 15 \cdot \frac{2}{10}$$

$$E[XY] = \frac{88}{10} = 8.8$$

$$\therefore Cov(x,y) = E[XY] - E[X]E[Y]$$

$$= 8.8 - (3.6)(2.3)$$

$$Cov(X,Y) = 0.52$$

iii) $E[X^2] = \Sigma x^2 P_X(x)$

$$E[X^2] = 4 \cdot \frac{1}{10} + 9 \cdot \frac{4}{10} + 16 \cdot \frac{3}{10} + 25 \cdot \frac{2}{10}$$

$$E[x^2] = \frac{138}{10} = 13.8$$

$E[Y^2] = \sum y^2 P_Y(y)$

$E[Y^2] = 1 \cdot \frac{1}{10} + 4 \cdot \frac{5}{10} + 9 \cdot \frac{4}{10}$

$E[Y^2] = \frac{57}{10} = 5 \cdot 7$

$Var(x) = Cov(X,X) = E[X^2] - [E[X]]^2$

$Var(x) = 13 \cdot 8 - (3 \cdot 6)^2 = 0 \cdot 84$

$\sigma_x = \sqrt{var(x)} = \sqrt{0 \cdot 84} = 0 \cdot 91$

$Var(Y) = cov(Y, y) = E[Y^2] - [E[Y]]^2$

$var(Y) = 5 \cdot 7 - (2 \cdot 5)^2$

$var(Y) = 0 \cdot 4$

$\sigma_y = \sqrt{var(y)} = \sqrt{0 \cdot 4}$

$\sigma_y = 0 \cdot 64$

$f(x,y) = \frac{Cov(x,y)}{\sqrt{x}\sqrt{y}}$

$= \frac{0 \cdot 52}{0 \cdot 91 \times 0 \cdot 64}$

$f(x,y) = 0 \cdot 89$

9. Two fruits are selected at random from a bag containing 3 Apples, 2 Oranges and 4 Mangoes. If X and Y are respectively, the number of Apples and the number of Oranges included among the two fruits drawn from the bag, find the probability associated with all possible pair of values (x, y). Also find the correlation between the variables X and Y.

3 → Apples

2 → oranges

4 → Mangoes

9 fruits

X → no of apples
X → no of oranges

occurrence	2_A	2_O	2_M	1_A 1_O	1_A 1_M	1_O 1_M
X	2	0	0	1	1	0
Y	0	2	0	1	0	1

X \ Y	0	1	2
0	$\frac{4C2}{9C2}$	$\frac{2C_1 \times 4C_1}{9C_1}$	$\frac{2C2}{9C_2}$
1	$\frac{3C_1 \times 2C_1}{9C_2}$	$\frac{3C_1 \times 2C_1}{9C_2}$	0
2	$\frac{3C2}{9C2}$	0	0

X \ Y	0	1	2
0	6/36	8/36	1/36
1	12/36	6/36	0
2	3/36	0	0

Marginal distribution of x and Y

X	0	1	2
$P_X(X)$	15/36	18/36	3/36

Y	0	1	2
$P_Y(y)$	21/36	14/36	1/36

$E[X] = \Sigma x P_x(x) = 0 + 18/36 + 6/36 = 24/36$

$E[Y] = \Sigma y P_y(y) = 0 + 14/36 + 2/36 = 16/36$

$E[X,Y] = \Sigma\Sigma xy \, P_{x,y}(x,y)$

$= 0 + \frac{6}{36} + 0$

$E[X,Y] = \frac{6}{36}$

$E[Y^2] = \Sigma x^2 P_x(x) = 0 + \frac{18}{36} + \frac{12}{36} = \frac{30}{36}$

$E[Y^2] = \Sigma y^2 P_y(y)$

$= \frac{14}{36} + \frac{4}{36}$

$E[Y^2] = \frac{18}{36}$

$\therefore \text{Cov}(X, Y) = E[X, Y] - E[X]E[Y]$

$$= \frac{6}{36} - \frac{24}{36} \times \frac{16}{36}$$

$\text{cov}(x, y) = -0.129$

$\text{Var}(x) = \text{cov}(x, x) = E[x^2] - (E[x])^2$

$\text{Var}(x) = \frac{30}{36} - \left(\frac{24}{36}\right)^2 = 0.38$

$\Rightarrow \sigma_x = \sqrt{\text{var}(x)} = \sqrt{0.38} = 0.616$

$\text{Var}(y) = \text{Cov}(X, Y) = E[y^2] - [E[y]]^2$

$\text{var}(y) = \frac{18}{36} - \left(\frac{16}{36}\right)^2 = 0.302$

$\sigma_y = \sqrt{\text{var}(y)} = \sqrt{0.302} = 0.549$

$\therefore \rho(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$

$$= \frac{-0.129}{0.616 \times 0.549}$$

$\rho(x, y) = -0.38$

Continuous joint distribution

→ Let X and Y are two continuous R.V, the probability density function $f_{x,y}(x,y)$ $(f(x,y))$ satisfy the following:

i) $f_{x,y}(x,y) \geq 0 \ \forall (x,y)$

ii) $$\int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} f_{x,y}(x,y)\, dx\, dy = 1$$

iii) $$P((x,y) \in R) = \iint_R f_{x,y}(x,y)\, dA$$

where R is the region in 2-D

Note: If Area(R) = 0 the probability $P((x,y) \in R) = 0$
i.e 1) $P((x,y)) = 0$ (i.e probability at a point is zero)
2) probability on a line is zero

Expectation

Let X and Y are continuous R.V and $g(x,y)$ be the function of x and y thus

$$E[g(x,y)] = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} g(x,y)\, f_{x,y}(x,y)\, dx\, dy$$

Linearity

1) $E[ax+by] = aE[x] + bE[y]$
2) $E[x+y] = E[x] + E[y]$

Cumulative PDF

Let x and y are continuous R.V's then the cumulative PDF is defined as:

$$P(X \leq x, Y \leq y) = \int_{y=-\infty}^{y} \int_{x=-\infty}^{x} f_{x,y}(x,y)\, dx\, dy$$

Marginal Distribution :

let x and y are continuous R.V's then marginal distribution of x and y are

$$f_X(x) = \int_{y=-\infty}^{\infty} f_{X,Y}(x,y)\, dy \;:\; f_Y(y) = \int_{x=-\infty}^{\infty} f_{X,Y}(x,y)\, dx$$

Co-variance and Correlation

$$Cov(x,y) = E[xy] - E[x]E[y]$$

$$\rho = \frac{Cov(x,y)}{\sqrt{var(x) \cdot var(y)}}$$

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

$var(y) = E(y^2) - (E[y])^2$

Independence

Two joint RV's are independent if $f_{x,y}(x,y) = f_x(x) \cdot f_y(y)$

Note: If x and y are independent then:

i) $E[XY] = E[X]E[Y]$

ii) $Cov(X,Y) = 0$

Q1) If joint probability function for (x,y) is $f(x,y) = \begin{cases} c(x^2+y^2) & 0 \le x \le 1,\ 0 \le y \le 1,\ c \ge 0 \\ 0 & \text{otherwise} \end{cases}$

Determine the following

i) Find the value of c

ii) Marginal distributions of x and y

iii) $P\left(x < \frac{1}{2}, y > \frac{1}{2}\right)$

iv) $P\left(\frac{1}{4} < x < \frac{3}{4}\right)$

v) $P(y < 1/2)$

1) For PDF

$$\int_{y=-\infty}^{\infty}\int_{x=-\infty}^{\infty} f_{x,y}(x,y)\,dx\,dy = 1$$

$$\int_{y=0}^{1}\int_{x=0}^{1} c(x^2+y^2)\,dx\,dy = 1$$

$$\int_{y=0}^{1} c\left(\frac{x^3}{3}+y^2x\right)_0^1 dy = 1$$

$$\int_0^1 c\left(\frac{1}{3}+y^2\right)dy = 1$$

$$c\left[\frac{1}{3}y+\frac{y^3}{3}\right]_0^1 = 1$$

$$c\left(\frac{1}{3}+\frac{1}{3}-0\right) = 1 \quad\Rightarrow\quad c\left(\frac{2}{3}\right)=1$$

$$\Rightarrow c = \frac{3}{2}$$

2. $$f_x(x) = \int_{y=0}^{1} f_{x,y}(x,y)\,dy$$

$$= \int_{y=0}^{1} \frac{3}{2}(x^2+y^2)\,dy$$

$$= \frac{3}{2}\left(x^2y+\frac{y^3}{3}\right)_0^1$$

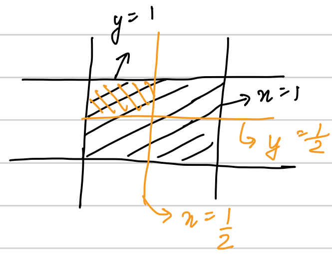
$$= \frac{3}{2}\left(x^2+\frac{1}{3}-0\right) = \frac{3x^2}{2}+\frac{1}{2}$$

$$f_Y(y) = \int_{x=0}^{1}\frac{3}{2}(x^2+y^2)\,dx$$

$$= \frac{3}{2}\left(\frac{x^3}{3}+y^2x\right)_0^1 = \frac{3y^2}{2}+\frac{1}{2}$$

P.T.O →

3. $P\left(x < \frac{1}{2},\ y > \frac{1}{2}\right)$



$$= \int_{y=\frac{1}{2}}^{1}\int_{x=0}^{1/2} \frac{3}{2}(x^2+y^2)\,dx\,dy$$

$$= \frac{3}{2}\int_{y=\frac{1}{2}}^{1} \left(\frac{x^3}{3} + y^2 x\right)_{x=0}^{1/2} dy$$

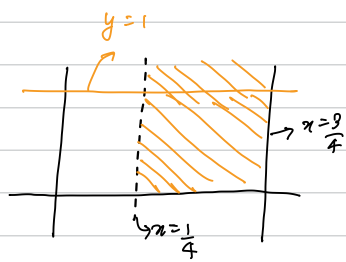
$$= \frac{3}{2}\int_{y=\frac{1}{2}}^{1} \left(\frac{1}{24} + \frac{y^2}{2}\right) dy$$

$$= \frac{3}{2}\left(\frac{y}{24} + \frac{y^3}{6}\right)_{y=1/2}^{1}$$

$$P\left(x < \frac{1}{2},\ y > \frac{1}{2}\right) = \frac{3}{2}\left(\frac{1}{24} + \frac{1}{6} - \left(\frac{1}{48} + \frac{1}{48}\right)\right)$$

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

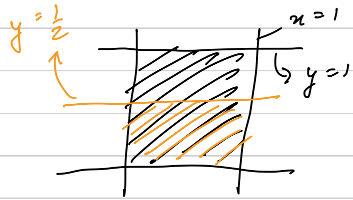
4) $P\left(\frac{1}{4} < x < \frac{3}{4}\right)$



$$= \int_{y=0}^{1}\int_{x=\frac{1}{4}}^{3/4} \frac{3}{2}(x^2+y^2)\,dx\,dy$$

$$= \frac{29}{64}$$

5. $P(y < 1/2)$



$$= \int_{y=0}^{1/2}\int_{x=0}^{1} \frac{3}{2}(x^2+y^2)\,dx\,dy$$

$$P(Y < 1/2) = \frac{5}{16}$$

Q 2. The joint PDF is given by $f(x,y) = \begin{cases} kxy & x \in (0,2) \\ & y \in (0,x) \\ 0 & \text{otherwise} \end{cases}$

Determine k, $P(1/2 < x < 1)$, $E[x]$, $E[Y]$, $E[XY]$, $E[x+3y]$.

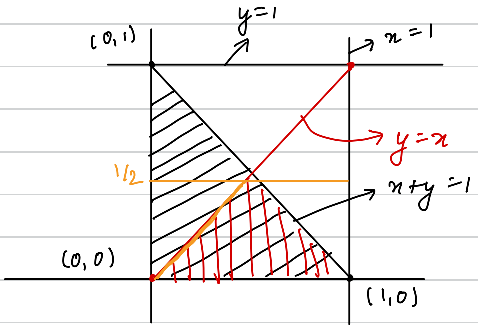
Determine k when $y \in (0,1)$

3. A company produces cans of mixed nuts almonds, cashews, & peanuts. Each can is exactly one pound; But the amt of each kind of nut is random. The Joint PDF of X, (amount of almonds) and Y (amount of cashews).

$$f(x,y) = \begin{cases} 24xy, & 0 \le x \le 1, \ 0 \le y \le 1 \\ & x+y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

i) Show that prob there are more almonds than cashews is 0.5

ii) $Cov(X,Y)$



xy
L → R

i) $P(x>y) = \iint_R f(x,y)\,dx\,dy$

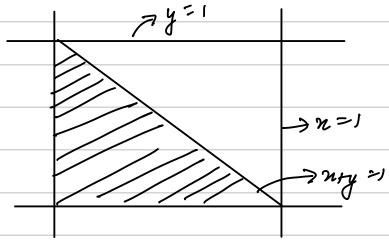
$$P(x>y) = \int_{y=0}^{1/2}\int_{x=y}^{1-y} 24xy\,dx\,dy$$

$$= \int_{y=0}^{1/2} 24y\left(\frac{x^2}{2}\right)_{x=y}^{1-y} dy$$

$$= \int_{y=0}^{1/2} 12y\left((1-y)^2 - y^2\right) dy$$

$$= \int_{y=0}^{1/2} 12y(1-2y)\,dy = \frac{1}{2}$$

ii) $E[x] = \iint x f(x,y)\,dx\,dy$



$$= \int_{y=0}^{1}\int_{x=0}^{1-y} 24x^2y\,dx\,dy$$

$$= \int_{y=0}^{1} \overset{8}{\cancel{24}}y\left(\frac{x^3}{3}\right)_{x=0}^{1-y} dy$$

$$= 8\int_{y=0}^{1} y\left((1-y)^3 - 0\right) dy$$

$E[x] = 0.4$

$$E[y] = \iint y\, f(x,y)\, dx\, dy$$

$$= \int_{y=0}^{1}\int_{x=0}^{1-y} 24xy^2\, dx\, dy$$

$$24\int_{y=0}^{1} y^2 \left(\frac{x^2}{2}\right)_{x=0}^{1-y} dy$$

$$= 12\int_{y=0}^{1} (1-y)^2\, dy$$

$$E[y] = 12\int_{y=0}^{1} (y-y^2)^2\, dy$$

$$E[y] = 0.4$$

iii) $$E[xy] = \iint xy\, f(x,y)\, dx\, dy$$

$$= \int_{y=0}^{1}\int_{x=0}^{1-y} 24x^2y^2\, dx\, dy$$

$$= 24\int_{y=0}^{1} y^2 \left(\frac{x^3}{3}\right)_{x=0}^{1-y} dy$$

$$= 8\int_{y=0}^{1} y^2(1-y)^3\, dy = 0.13$$

$$\text{cov}(x,y) = E[xy] - E[x]E[y]$$

$$= 0.13 - (0.4)(0.4)$$

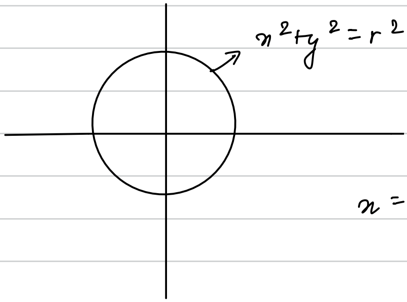
$$= -0.03$$ // $\left\{ \text{neg cov} \Rightarrow \text{almonds} \propto \frac{1}{\text{cashews}} \right\}$

2. An ecologist selects a point inside a circular sampling region according to a uniform distribution. Let X be the x-coordinate of the point selected and Y be the y-coordinate of the point selected. If the circle is centered at (0,0) and has radius r, then the joint pdf of X and Y is $f(x,y)=\begin{cases} c^2 & x^2+y^2\le r^2 \\ 0 & otherwise \end{cases}$

a) Determine the value of c that makes this a valid joint p.d.f.

b) What is the probability that the selected point is within r/2 of the center of the circular region? (*Hint:* Use geometry.)

c) What is the probability that *both* X and Y differ from 0 by at most r/2?



$$x=\pm\sqrt{r^2-y^2}$$

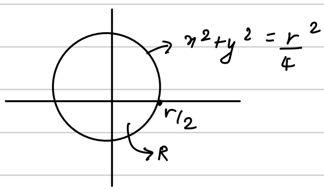
a) $\iint f(x,y)\,dx\,dy = 1$

$$\int_{y=0}^{r}\int_{x=-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} c^2\,dx\,dy = 1$$

do this instead:

$$4c^2\int_{y=0}^{r}\int_{x=0}^{\sqrt{r^2-y^2}} dx\,dy = 1 \quad\Rightarrow\quad c=\frac{1}{r\sqrt{\pi}}$$

b)



$P((x,y)\in R)$

$$=\frac{1}{\pi r^2}\iint_R dx\,dy = \frac{1}{\pi r^2}\,4\int_{y=0}^{r/2}\int_{x=0}^{\sqrt{\frac{r^2}{4}-y^2}} dx\,dy$$

$$= \frac{4}{\pi r^2}\left[\int_{y=0}^{r/2} (x)_{x=0}^{\sqrt{\frac{r^2}{4}-y^2}} dy\right]$$

$$= \frac{4}{\pi r^2}\int_{y=0}^{r/2} \sqrt{\frac{r^2}{4}-y^2}\, dy$$

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

c) same as b.

Stochastic process

Stochastic matrix P is said to be regular stochastic matrix if elements of P^n are positive for same n.

Stochastic process is a set of random variables $\{X(t), t \in T\}$ defined on sample space S with a parameter $t \in T$, index set.

Problems :

6. Check whether the given matrix is regular stochastic matrix and linear. Find unique fixed prob vector, if P is regular

$$P = \begin{bmatrix} 0 & 0 & 0 \\ 1/6 & 1/2 & 1/3 \\ 0 & 2/3 & 1/3 \end{bmatrix}$$

soln; Since $P_{ij} \geq 0$ and $\sum_{j=1}^{3} P_{ij} = 1 \ \forall i$

$\therefore$ Given matrix is stochastic matrix

$$P^2 = \begin{bmatrix} 0 & 1 & 0 \\ 1/6 & 1/2 & 1/3 \\ 0 & 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1/6 & 1/2 & 1/3 \\ 0 & 2/3 & 1/3 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 1/6 & 1/2 & 1/3 \\ 1/12 & 23/36 & 5/12 \\ 1/7 & 5/9 & 1/3 \end{bmatrix}$$

Since all elements of P^2 are +ve

$\therefore$ P is a regular stochastic matrix

$\exists$ unique fixed point vector $v = (v_1, v_2, v_3)$ s.t $vP = v$

Since v is a prob vector

$\Rightarrow v_1 + v_2 + v_3 = 1 \longrightarrow (1)$

Consider $v \cdot P = v$

$$(v_1 \ \ v_2 \ \ v_3) \begin{bmatrix} 0 & 1 & 0 \\ 1/6 & 1/2 & 1/3 \\ 0 & 2/3 & 1/3 \end{bmatrix} = (v_1 \ \ v_2 \ \ v_3)$$

$\frac{v_2}{6} = v_1 \quad \Rightarrow \quad 6v_1 - v_2 = 0 \longrightarrow (2)$

$$v_1 + \frac{v_2}{2} + \frac{2v_3}{3} = v_2 \Rightarrow v_1 - \frac{v_2}{2} + \frac{2v_3}{3} = 0 \rightarrow (3)$$

$$\frac{v_2}{3} + \frac{v_3}{9} = v_3 \Rightarrow \frac{v_2}{3} - \frac{2}{3}v_3 \rightarrow (4)$$

Solving (1), (2), (3) & (4), we get

$v_1 = 0.1$, $v_2 = 0.6$, $v_3 = 0.3$

$\therefore$ $v = (0.1, 0.6, 0.3)$

4. P.T the markov chain when transition prob matrix p is irreducible.

$$P = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

10) Consider a game of "ladder climbing". There are 5 levels in the game, level 1 is the lowest (bottom) and level 5 is the highest (top). A player starts at the bottom. Each time, a fair coin is tossed. If it turns up heads, the player moves up one rung. If tails, the player moves down to the very bottom. Once at the top level, the player moves to the very bottom if a tail turns up, and stays at the top if head turns up.
a) Find the transition probability matrix.
b) Find the two-step transition probability matrix.
c) Find the steady-state distribution of the Markov chain.

i) Status = $\{1, 2, 3, 4, 5\}$ $1 \rightarrow$ Bottom level, $5 \rightarrow$ top level

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{bmatrix} \end{array}$$

Heads $\rightarrow$ level + 1 (1-4)

tails $\rightarrow$ bottom level (1)

ii) Two step transition matrix is given by

$$P^2 = P \times P$$

P.T.O ↓

$$P^2 = \begin{bmatrix} 1/2 & 1/4 & 1/4 & 0 & 0 \\ 1/2 & 1/4 & 0 & 1/4 & 0 \\ 1/2 & 1/4 & 0 & 0 & 1/4 \\ 1/2 & 1/4 & 0 & 0 & 1/4 \\ 1/2 & 1/2 & 0 & 0 & 1/4 \end{bmatrix}$$

Consider $v^{(0)} = (1\ 0\ 0\ 0\ 0)$

2-step prob vector is $v^{(2)} = v^{(0)} P^2$

$v^{(2)} = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0, 0\right)$

3. Steady state distributions:

vector = unique fixed point variable

Find $v = (v_1, v_2, v_3, v_4, v_5)$ s.t $v \cdot p = v$ and $v_1 + v_2 + v_3 + v_4 + v_5 = 1$

Consider $vp = v$

$$(v_1\ v_2\ v_3\ v_4\ v_5) \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{bmatrix} = (v_1, v_2, v_3, v_4, v_5)$$

$v_1 + v_2 + v_3 + v_4 + v_5 = 3v_1$

$\frac{v_1}{2} = v_2$ — ③ $\Rightarrow v_2 = \frac{1}{4}$

$\frac{v_1}{2} = v_3$ — ④ $\Rightarrow v_3 = \frac{1}{8}$

$\frac{v_3}{2} = v_4$ — ⑤ $\Rightarrow v_4 = \frac{1}{16}$

$\frac{v_4}{2} + \frac{v_5}{2} = v_5$ $\rightarrow$ ⑥ $v_5 = \frac{1}{16}$

$\therefore v = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\right)$

15) In a certain city, the weather on a day is reported as sunny, cloudy or rainy. If a day is sunny, the probability that the next day is sunny is 70%, cloudy is 20% and rainy is 10%. If a day is cloudy, the probability that the next day is sunny is 30%, cloudy is 20% and rainy is 50%. If a day is rainy, the probability that the next day is sunny is 30%, cloudy is 30% and rainy is 40%. If a Sunday is sunny, find the probability that the Wednesday is rainy.

$S = \{$ sunny (s), cloudy (c), Rainy (R) $\}$

$$P = \begin{array}{c} \\ S \\ C \\ R \end{array} \begin{array}{c} \begin{array}{ccc} S & C & R \end{array} \\ \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.5 \\ 0.3 & 0.3 & 0.4 \end{bmatrix} \end{array}$$

sunday $\rightarrow$ sunny $= v^{(0)}$

Monday $\rightarrow v^{(1)}$

Tuesday $\rightarrow v^{(2)}$

Wednesday $\rightarrow$ Rainy $\rightarrow v^{(3)}$

Consider

$$P^3 = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.5 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}^3 = \begin{bmatrix} 0.512 & 0.221 & 0.247 \\ 0.468 & 0.233 & 0.299 \\ 0.468 & 0.234 & 0.248 \end{bmatrix}$$

$$v^{(3)} = v^{(3)} \cdot P$$

$$v^{(3)} = (1, 0, 0) \begin{bmatrix} 0.512 & 0.221 & 0.247 \\ 0.468 & 0.233 & 0.299 \\ 0.468 & 0.234 & 0.248 \end{bmatrix}$$

$$v^{(3)} = (0.512 \quad 0.221 \quad 0.247)$$

prob that wednesday is rainy is 0.247 //

16. A housewife buys 3 kinds of cereals A, B, and C. She never buys the same cereal in successive weeks. If she buys cereal A, the next week she buys cereal B. However, if she buys cereal B or C, the next week she is three times as likely to buy cereal A as the other cereal. In the long run how often she buys each of the three cereals?

States $= \{A, B, C\}$

$$P = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 3/4 & 0 & 1/4 \\ 1/4 & 3/4 & 0 \end{bmatrix} \end{array}$$

$3x + x = 1$

$4x = 1$

$x = \frac{1}{4}$ //

$$\begin{bmatrix} 0 & 1 & 0 \\ 3x & 0 & x \\ 3x & x & 0 \end{bmatrix} \rightarrow \quad 3x + x = 1, \quad x = 1/4$$

Find the unique prob vector $v = (v_1, v_2, v_3)$ s.t $v_1 + v_2 + v_3 = 1$ and $vp = v \longrightarrow (1)$

$$(v_1 \; v_2 \; v_3) \begin{bmatrix} 0 & 1 & 0 \\ 3/4 & 0 & 1/4 \\ 3/4 & 1/4 & 0 \end{bmatrix} = (v_1 \; v_2 \; v_3)$$

$$\frac{3v_2}{4} + \frac{3v_3}{4} = v_1 \longrightarrow (2)$$

$$v_1 + \frac{v_3}{4} = v_2 \longrightarrow (3)$$

$$v_2 = \frac{1}{4} \longrightarrow (4)$$

Solving (1), (2), (3) & (4), we get

$v_1 = \frac{3}{7}$ $\quad v_2 = \frac{14}{35}$ $\quad v_3 = \frac{4}{35}$

Statistical Inference

- Statistics
  - Mathematical
  - Applied
    - Descriptive
      - Numerical
      - Graphical data
    - Inferential
      - Estimation
        - Single pt estimation
        - Interval estimation
      - Hypothetical Testing

Testing of Hypothesis

1. Read the question and assume r.v 'X', should be well-defined
2. Identify H_0 : (Null hypothesis) $\mu = k$
3. H_1 : (Alternate hypothesis)

   i) $\mu \neq k \longrightarrow$ (2 tailed test)

   or

   ii) $\mu < k \longrightarrow$ (Left tailed test)

   or

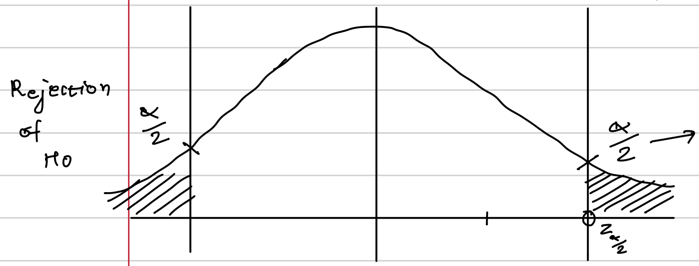
   iii) $\mu > k \longrightarrow$ (Right tailed test)

$H_0 : \mu = k$ ✓

v/s

$H_1 : \mu \neq k$ (2 tailed test)

$\alpha \rightarrow$ level of significance



$\frac{\alpha}{2} = 0.5 - P(0 \leq z \leq z_{\alpha/2})$

$z_{\frac{\alpha}{2}} =$ ______

$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = 0 - \varphi$

2. Mice with an average lifespan of 32 months will live upto 40 months when fed by a certain nutrious food. If 64 mice fed on this diet have an average lifespan of 38 months and standard deviation of 5.8 months, is there any reason to believe that average lifespan is less than 40 months.

soln; $x \rightarrow$ life span of the mice

$H_0 : \mu = 40$

v/s

$H_1 : \mu < 40 \longrightarrow$ (left tailed test)

$\alpha = 5\% = 0.05 \longrightarrow z_\alpha =$

$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{38 - 40}{5.8 / \sqrt{64}} = \underline{-2.758}$

$0.05 = 0.5 - P(0 \le z \le z\alpha)$

$P(0 \le z \le \alpha) = 0.45$

$z\alpha = -1.65$

$\Rightarrow z > z\alpha \Rightarrow$ Reject's H_0

3. A machine runs on an average of 125 hours/year. A random sample of 49 machines has an annual average use of 126.9 hours with standard deviation 8.4 hours. Does this suggest to believe that machines are used on the average more than 125 hours annually at 0.05 level of significance?

Given ;

$x \rightarrow$ No of hrs machine runs manually

$H_0 : \mu = 125$

v/s

$H_1 : \mu \geq 125$ (Right tailed test)

$z = \frac{\bar{x} - 125}{8.4 \sqrt{49}} = \frac{126.9 - 125}{8.4\sqrt{49}} = 1.5833$

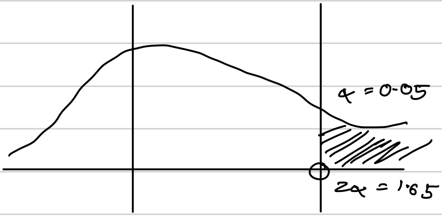
$\alpha = 0.05 \longrightarrow z\alpha$

$\downarrow$

$0.05 = 0.5 - P(0 \le z \le z\alpha)$

$z\alpha = 1.65$

$\Rightarrow z < z\alpha \Rightarrow$ Accept's H_0



Q 5. It has previously been recorded that the average depth of ocean at a particular region is 67.4 fathoms. Is there reason to believe this at 0.01 L.O.S. if the readings at 40 random locations in that particular region showed a mean of 69.3 with S.D of 5.4 fathoms?

Given ;

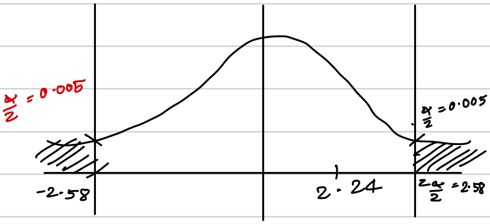
$X \rightarrow$ Depth of the ocean

$H_0 : \mu = 67.4$

v/s

$H_1 : \mu \neq 67.4$ (2 tailed test)

$\rightarrow z = \frac{69.3 - 67.4}{5.4\sqrt{40}} = 2.225$



$$0.005 = 0.5 - P(0 \le z \le z_\alpha)$$

$$z_\alpha = 2.58$$

$|z| < z_\alpha \Rightarrow$ Accepting H_0

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

{ Test of significance for difference b/w two means }

To test the effects a new pesticide on rice production, a farm land was divided into 60 units of equal areas, all portions having identical qualities as to soil, exposure to sunlight etc. The new pesticide is applied to 30 units while old pesticide to the remaining 30. Is there reason to belive that the new pesticide is better than the old pesticide if the mean number of kgs of rice harvested/units using new pesticide (N.P.) is 496.31 with S.D. of 17.18 kgs. Test at a level of significance (a) 0.05 (b) 0.01?

↓ while for the old pesticide it is 485.41 and its S.D = 14.73

A	B
n_A	o_B
30	30

Given;

$X_A \rightarrow$ Rice harvested using new pesticide

$X_B \rightarrow$ " " " old "

$H_0 : \mu_A = \mu_B \Rightarrow \mu_A - \mu_B = 0 = \delta$

$H_1 : \mu_A > \mu_B \rightarrow$ (Right tailed test)

$$\rightarrow Z = \frac{(496.31 - 485.41) - 0}{\sqrt{\frac{(17.18)^2}{30} + \frac{(14.73)^2}{30}}} = 2.638$$

$\alpha = 0.05$

$0.05 = 0.5 - P(z_\alpha)$

$P(z_\alpha) = 0.45$

$z_\alpha = 1.65$

$z > z_\alpha$

$\Rightarrow$ Reject H_0

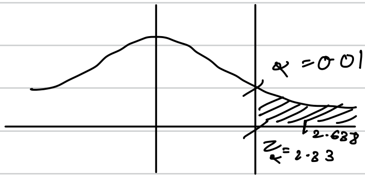
$\alpha = 0.01$

$0.01 = 0.5 - P(z_\alpha)$

$z_\alpha = 2.33$

$z > z_\alpha$

Reject H_0



3. A random sample of 40 'geysers' produced by company A have a mean lifetime (mlt) of 647 hours of continuous use with a S.D. of 27 hours, while a sample 40 produced by another company B have mlt of 638 hours with S.D. 31 hours. Does this substantiate the claim of company A that their 'geyers' are superior to those produced by company B at (a) 0.05 (b) 0.01 L.O.S.

Given;

$A = 40$

$B = 40$

$X_A \rightarrow$ Product by company

III Small sample test concerning single mean (t-distribution)

1) We use this distribution when the sample size is small (n<30)

2) Degree of freedom (dof)

Number of independent variable free to vary is called dof.

$\nu = n-1$ = dof for the sample size n.

Test statistic	Null Hypothesis	Alternative Hypothesis	Accept H_o
$t = \dfrac{\bar{x}-\mu_O}{S/\sqrt{n}}$ here $S^2 = \dfrac{\sum(x_i-\bar{x})^2}{n-1}$ **OR** $t = \dfrac{\bar{x}-\mu_O}{s/\sqrt{n-1}}$ here $s^2 = \dfrac{\sum(x_i-\bar{x})^2}{n}$	$H_0 : \mu = \mu_O$	$H_1 : \mu \neq \mu_O$	$\lvert t\rvert < t_{\alpha/2, n-1}$
		$H_1 : \mu < \mu_O$	$t > t_\alpha$
		$H_1 : \mu > \mu_O$	$t < t_\alpha$

Q1. An ambulance service company claims that on an average it takes 20 minutes between a call for an ambulance and the patient's arrival at the hospital. If in 6 calls the time taken (between a call and arrival at hospital) are 27, 18, 26, 15, 20 and 32. Can the company's claim be accepted?

Given,

$n=6$, R.V X is time taken to reach the hospital

$$\bar{x} = \frac{27+18+26+15+20+32}{6} = 23$$

S.D $$s = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}}$$

$$s = \sqrt{\frac{(27-23)^2+(18-23)^2 - \cdots + (32-23)^2}{6-1}}$$

$$s = 6.3974$$

Null hypothesis : $H_0 : \mu = 20$

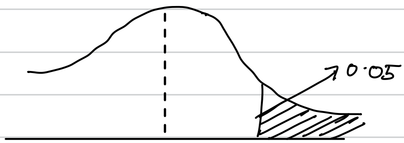
Alternate hypothesis : $H_1 : \mu > 20$ (right tailed test)

P.T.O ↓

Test Statistic :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{23 - 20}{6.38/\sqrt{6}}$$

$$t = 1.1504$$



Critical Value t_α

Let us take 5% Level of significance (LOS)

i.e $\alpha = 0.05$

dof $= \nu = n-1 = 5$

From the t-table, we get

$t_\alpha = 2.015 \Rightarrow P(t_\alpha < t < \infty) = 0.05$

Since $t = 1.1504 < 2.015 = t_\alpha$ (since we are using right tailed test)

$\therefore$ Accept N.H H_0

$\therefore$ An ambulance will reach within 20 mins.

Q 5. In a random sample of 10 bolts produced by a machine the mean length of bolt is 0.53 mm and S.D 0.03mm. Can we claim from this that the machine is in proper working order if in the past it produced bolts of length 0.50 mm? Use (a) 0.05 (b) 0.01 L.O.S

Given $n = 10$, $\bar{x} = 0.53$

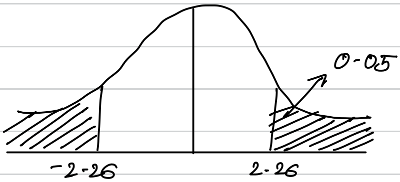
$s = 0.03$, $\mu_0 = 0.50$

Test Statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.53 - 0.50}{0.03/\sqrt{10}}$$

$$t = 3.16$$

Critical value



case 1 : LOS is $\alpha = 0.05$

dof : $\nu = n-1 = 9$

Null hypothesis (H_0) : $\mu = \mu_0$

Alternate hypothesis (H_1) : $\mu \neq \mu_0$ (both tailed)

if $\mu > \mu_0$ $\rightarrow$ right tailed

if $\mu < \mu_0$ $\rightarrow$ left tailed

From the distribution table

$P(-\infty < t < -t_{\alpha/2}) + P(t_{\alpha/2} < t < \infty)$

$= 0.05$

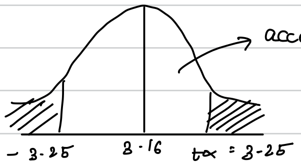
$t_{\frac{\alpha}{2}} = 2.262$

Since $|t| = 3.16 > t_{\frac{\alpha}{2}} = 2.262$

$\therefore$ Reject N.H H_0

$\therefore$ Under 5% LOS, The mean of bolts not equal to 50 mm

case 2 : LOS $\alpha = 0.01$ and dof $\nu = 9$



From the distribution-table :

$t_{\frac{\alpha}{2}} = 3.25$

Since $|t| = 3.16 < 3.25 = t_{\frac{\alpha}{2}}$

$\therefore$ Accept N.H H_0

Q 4. If 5 pieces of certain ribbon selected at random have mean breaking strength of 169.5 pounds with S.D. of 5.7, do they confirm to the specification mean breaking strength of 180 pounds.

Given, $\bar{x} = 169.5$, $s = 5.7$, $n = 5$

$\mu_0 = 180$

Test statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{169.5 - 180}{5.7/\sqrt{5}}$$

$t = -4.115$

Critical Value :

Let LOS be $\alpha = 0.05$ and dof $\nu = n-1 = 4$

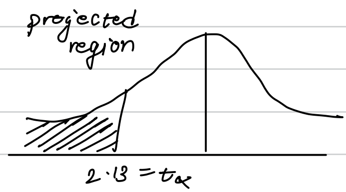
N·H (Ho) : $\mu = \mu_0 = 180$

A·H (H1) : $\mu < 180$

Since in A·H we have $\mu < 180$ use left tailed test.

Since $t = 4.119 < 2.13 = t_\alpha$

∴ Reject N·H Ho



Test Statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ → use this if raw data is given

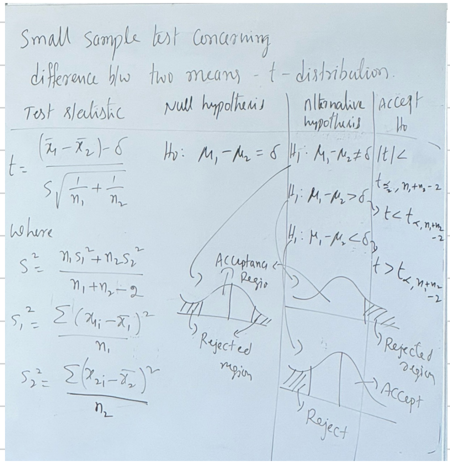
OR

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n-1}}$$ → use this if S·D is given or raw data is given

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

IV. Small Sample Test Concerning Difference Between Two Means: t-Distribution

Test statistic	Null Hypothesis	Alternative Hypothesis	Accept H_o
$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ (i) $S^2 = \frac{\sum (x_{1i} - \bar{x}_1)^2 + \sum (x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2}$ OR (ii) $S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$ if $s_1^2 = \frac{\sum (x_{1i} - \bar{x}_1)^2}{n_1}$ and $s_2^2 = \frac{\sum (x_{2i} - \bar{x}_2)^2}{n_2}$ OR (iii) $S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ if $S_1^2 = \frac{\sum (x_{1i} - \bar{x}_1)^2}{n_1 - 1}$ and $S_2^2 = \frac{\sum (x_{2i} - \bar{x}_2)^2}{n_2 - 1}$	$H_0 : \mu_1 - \mu_2 = \delta$	$H_1 : \mu_1 - \mu_2 \neq \delta$	$\lvert t \rvert < t_{\alpha/2, n_1+n_2-2}$
		$H_1 : \mu_1 - \mu_2 > \delta$	$t < t_\alpha$
		$H_1 : \mu_1 - \mu_2 < \delta$	$t > t_\alpha$



1. In a mathematics examination 9 students of class A and 6 students of class B obtained the following marks. Test at 1% L.O.S. whether the performance in mathematics is same or not for the two classes A and B. assume that the samples are drawn from normal populations having same variance.

A:	44	71	63	59	68	46	69	54	48
B:	52	70	41	62	36	50	-	-	-

Given;

$n_1 = 9$ $\qquad$ $n_2 = 6$

$\bar{x}_1 = \frac{\sum x_{1i}}{n_1}$ $\qquad$ $\bar{x}_2 = \frac{\sum x_{2i}}{n_2}$

$\bar{x}_1 = 58$ $\qquad$ $\bar{x}_2 = 51.83$

$$s_1^2 = \frac{\sum (x_{1i} - \bar{x})^2}{n_1} = \frac{(44-58)^2 + (71-58)^2 + \cdots (48-58)^2}{9}$$

$$s_1^2 = 96.88$$

Similarly;

$$s_2^2 = \frac{\sum (x_{2i} - \bar{x})^2}{n_2} = \frac{(52-51.83)^2 + (70-51.8)^2 + \cdots + (50-51.83)^2}{6}$$

$$\therefore S_2^2 = 134.13$$

$$\therefore s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$$s^2 = \frac{9 \times 96.88 + 6 \times 134.13}{13}$$

$$s^2 = 128.97$$

$$s = 11.356$$

⇒ Test Statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{s\sqrt{1/n_1 + 1/n_2}} = \frac{(58 - 51.83) - 0}{11.356\sqrt{1/9 + 1/6}}$$

$$t = 0.973$$

Null hypothesis $H_0 : \mu_1 - \mu_2 = 0 = \delta$

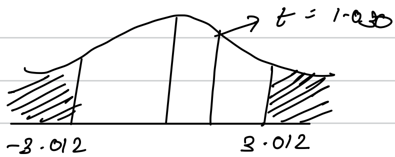
$A.H\ (H_1) = \mu_1 - \mu_2 \neq 0 = \delta$ (Two tailed test)

Critical Value :

Find the $t_{\frac{\alpha}{2}}$ using LOS $\alpha = 0.01$
and dof : $v = n_1 + n_2 - 2 = 13$

By t-distribution table

The CV $t_{\frac{\alpha}{2}}$ S.T



$P(-\infty < t < -t_{\alpha}) + P(t_{\alpha} < t < \infty) = 0.01$

is $t_{\frac{\alpha}{2}} = 3.012$

Since $|t| = 1.030 < t_{\frac{\alpha}{2}} = 3.012$

Accept N.H H_0

performance in maths is same for the two classes A and B

2. Out of random sample of 9 mice, suffering with a disease, 5 mice were treated with a new serum while the remaining were not treated. From the time commencement of experiment, the following are the survival times:

Treatment	2.1	5.3	1.4	4.6	0.9
NoTreatment	1.9	0.5	2.8	3.1	-

Test whether the serum treatment is effective in curing the disease at 5% L.O.S., assuming that the two distributions are normally distributed with equal variances.

Given,

$n_1 = 5 \quad n_2 = 4$

$\bar{x}_1 = \frac{\sum x_{1i}}{n_1} = 2.86$

$\bar{x}_2 = \frac{\sum x_{2i}}{n_2} = 2.075$

$s_1^2 = \frac{\sum (x_{1i} - \bar{x}_1)^2}{n_1} = 3.1064$

$s_2^2 = \frac{\sum (x_{2i} - \bar{x}_2)^2}{n_2} = 1.3625$

$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

$S^2 = \frac{5 \times 3.1066 + 4 \times 1.3625}{7}$

$S^2 = 2.8027$

Test statistic

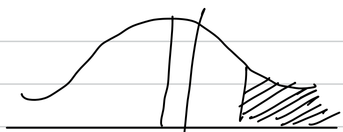
$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{S\sqrt{1/n_1 + 1/n_2}}$

P.T.O ↓

$$t = \frac{(2.86 - 2.075) - 0}{1.674\sqrt{1/5 + 1/4}} = 0.699$$

Null hypothesis $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$ (right tailed test)



0·696 $t_\alpha = 1.895$

Find CV t_α using $\alpha = 0.05$ LOS

and dof $v = n_1 + n_2 - 2 = 7$

By the t-table, we get

$t_\alpha = 1.895$, $\boxed{t < t_\alpha}$

$\therefore$ Accept N·H H_0.

V. **Paired Sample t-Test**

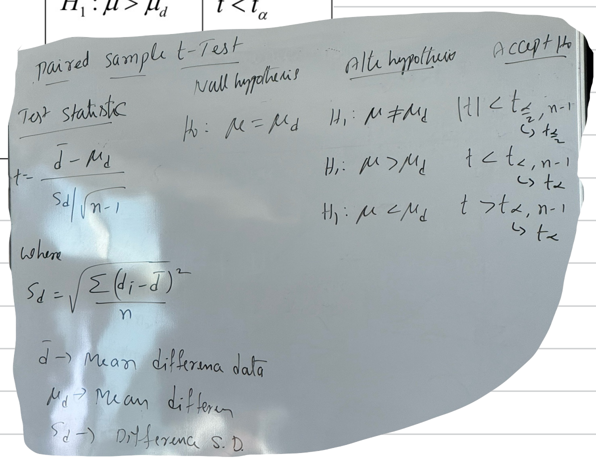
Test statistic	Null Hypothesis	Alternative Hypothesis	Accept H_o
$t = \dfrac{\bar{d} - \mu_d}{S_d/\sqrt{n}}$ here $S_d^2 = \dfrac{\sum (d_i - \bar{d})^2}{n-1}$	$H_0 : \mu = \mu_d$	$H_1 : \mu \neq \mu_d$	$\lvert t \rvert < t_{\alpha/2, n-1}$
OR		$H_1 : \mu < \mu_d$	$t > t_\alpha$
$t = \dfrac{\bar{d} - \mu_d}{S_d/\sqrt{n-1}}$ here $S_d^2 = \dfrac{\sum (d_i - \bar{d})^2}{n}$		$H_1 : \mu > \mu_d$	$t < t_\alpha$

where, $S_d = \sqrt{\dfrac{\sum (d_i - \bar{d})^2}{n}}$

$\bar{d} \rightarrow$ Mean difference data

$\mu_d \rightarrow$ Mean difference

$S_d \rightarrow$ Difference S.D



2. The average weekly losses of man-hours due to strikes in an institute before and after a disciplinary program was implemented are as follows: Is there reason to believe that the disciplinary program is effective at 5% L.O.S.?

Before	45	73	46	124	33	57	83	34	26	17
After	36	60	44	119	35	51	77	29	24	11

Before	45	73	46	124	33	57	83	34	26	17
After	36	60	44	119	35	51	77	29	24	11
$d = x_1 - x_2$	9	7	2	5	-2	6	6	5	2	6

$n = 10$ (sample)

$$\bar{d} = \frac{\sum d_i}{n} = 5.2$$

$$S_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n}}$$

$$= \sqrt{\frac{(9-5.2)^2 + (13-5.2)^2 + \dots + (6-5.2)^2}{10}}$$

$$S_d = 3.867$$

⇒ Test Statistic

$$t = \frac{\bar{d} - \mu_d}{S_d / \sqrt{n-1}}$$

$$t = \frac{5.2 - 0}{3.867/\sqrt{9}} = 4.034$$

Null hypothesis $H_0 : \mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0 = \mu_d$)

Alternative hypothesis $H_1 : \mu_1 > \mu_2$ (Right tailed test)

Note : If $d = x_1 - x_2$ values are mostly +ve : $\mu_1 > \mu_2$

" " −ve : $\mu_1 < \mu_2$

$d = x_1 - x_2$ values are mixed i.e

same amt of +ve and −ve : $\mu_1 \neq \mu_2$

The critical value t_α with LOS $\boxed{\alpha = 0.05}$ and dof $\boxed{v = n-1 = 9}$

By t-table

$t_\alpha = 1.833$

Since $t = 4.034 > 1.833 = t_\alpha$

Reject N.H H_0

Disciplinary program is not effective

3. The pulsality index (P.I.) of 11 patients before and after contracting a disease are given below. Test at 0.05 L.O.S. whether there is a significant increase of the mean of P.I. values.

Before	0.4	0.45	0.44	0.54	0.48	0.62	0.48	0.60	0.45	0.46	0.35
After	0.5	0.60	0.57	0.65	0.63	0.78	0.63	0.80	0.69	0.62	0.68

Before	0.4	0.45	0.44	0.54	0.48	0.62	0.48	0.60	0.45	0.46	0.35
After	0.5	0.60	0.57	0.65	0.63	0.78	0.63	0.80	0.69	0.62	0.68
$d = x_1 - x_2$	−0.1	−0.15	−0.13	−0.11	−0.15	−0.16	−0.15	−0.2	−0.24	−0.16	−0.33

$n = 11$

$\bar{d} = \frac{\Sigma d_i}{n} = -0.1709$

$s_d = \sqrt{\frac{\Sigma (d_i - \bar{d})^2}{n}}$

$s_d = 0.062$

Test statistic

$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n-1}}$

$= \frac{-0.1709 - 0}{0.062 / \sqrt{10}}$

$t = -8.66$

$\mu_1 \geq \mu_2$

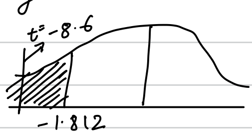
N.H $H_0 : \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0 = \mu_d$

A H $H_1 : \mu_1 < \mu_2$ (left tailed test)

Find the C.V t_α using $\boxed{\text{LOS } \alpha = 0.05}$

and dof $\boxed{v = n-1 = 10}$

By t-table, $t_\alpha = -1.812$ (since left tailed, hence take −ve)



Since $t = -8.66 < -1.812 = t_\alpha$

∴ Reject Null hypothesis.