Graph Theory & Probability/ Graph Theory

DMTH501/DMTH601

GRAPH THEORY AND PROBABILITY/ GRAPH THEORY

SYLLABUS

DMTH501 Graph Theory and Probability

Objectives: To learn the fundamental concept in graph theory and probabilities, with a sense of some of its modern application. Also to learn, understand and create mathematical proof, including an appreciation of why this is important. After the completion of this course prospective students will be able to tackle mathematical issues to applicable in data structure related course.

Sr. No.	Description
1	**Graph Theory:** Graph and sub graphs, Isomorphic, homomorphism graphs, Paths, Hamiltonian Circuits, Eulerian Graph, Connectivity
2	The Bridges of konigsberg,Transversal, Multi graphs, Labelled graph, Complete, regular and bipartite graphs, planar graphs
3	Graph colourings, Chromatic number, Connectivity, Directed graphs, basic definitions, Tree graphs, Binary trees, rooted trees
4	minimal spanning tree, Prim's algorithm, shortest path.
5	Propositions and compound propositions, Basic logical operations, Truth tables, Tautologies and Contradictions, Logical equivalence
6	Logic gates, Logic circuits, and switching function, Partially ordered set
7	Lattice, Boolean algebra,Lattice as Boolean algebra, Application of Boolean algebra to on –off switching theory.
8	Sample space, events, and probability functions, Examples using counting methods, sampling with or without replacement, Algebra of events
9	Conditional probability, partitions of sample space. Theorem of total probability, Baye's theorem, independence, Random variables
10	Probability mass functions,Discrete distributions: - Binomial, Poisson, geometric. Expectation: - mean and variance

DMTH601 Graph Theory

1	Graph Theory: Graph and sub graphs, Isomorphic, homomorphism graphs,
2	Paths, Hamiltonian circuits, Eulerian graph, connectivity
3	The bridges of Konigsberg, Transversal, Multi graphs, Labeled graph
4	Complete, regular and bipartite graphs, planar graphs
5	Graph colorings, Chromatic number, Connectivity, Directed graphs
6	Basic definitions , Tree graphs , Binary trees, rooted trees
7	Minimal spanning tree, Prim's algorithm, shortest path
8	Propositions and compound propositions,
9	Basic logical operations, Truth tables, Tautologies and Contradictions, Logical equivalence
10	Logic gates, Logic circuits, and switching function, Partially ordered set

CONTENTS

Unit 1: Introduction to Graph Theory

Objectives

After studying this unit, you will be able to:

- Define graph
- Evaluate the terms of graphs
- Describe the different types of path
- Explain homomorphism graphs
- Describe the paths and connectivity
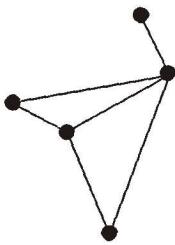
Introduction

The word *graph* refers to a specific mathematical structure usually represented as a diagram consisting of points joined by lines. In applications, the points may, for instance, correspond to chemical atoms, towns, electrical terminals or anything that can be connected in pairs. The lines may be chemical bonds, roads, wires or other connections. Applications of graph theory are found in communications, structures and mechanisms, electrical networks, transport systems, social networks and computer science.

1.1 Related Terms

A graph is a mathematical structure comprising a set of vertices, V, and a set of edges, E, which connect the vertices. It is usually represented as a diagram consisting of points, representing the **vertices** (or **nodes**), joined by lines, representing the **edges** (Figure 1.1). It is also formally denoted by G (V, E).

Figure 1.1

Vertices are incident with the edges which joins them and an edge is incident with the vertices it joins.

The **degree** of a vertex v is the number of edges incident with v, Loops count as 2.

The **degree sequence** of a graph G is the sequence obtained by listing, in ascending order with repeats, the degrees of the vertices of G (e.g. in Figure 1.2 the degree sequence of G is (1, 2, 2, 3, 4).

The **Handshaking Lemma** states that the sum of the degrees of the vertices of a graph is equal to that twice the no. of edge this follow reality from the fact that each edge join two vertices necessarily distinct) and so contributes 1 to the degree of each of those vertices.
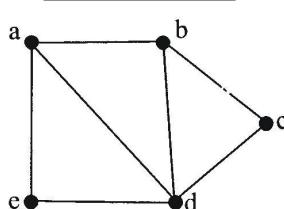
A walk of length *k* in a graph is a succession of *k* edges joining two vertices. NB Edges can occur more than once in a walk.

A trail is walk in which all the edges (but not necessarily all the vertices) are distinct.

A path is a walk in which all the edges and all the vertices are distinct.

So, in Figure 1.2, *abdcbde* is a walk of length 6 between *a* and *e*. It is not a trail (because edge *bd* is traversed twice). The walk *adcbde* is a trail length 5 between *a* and *e*. It is not a path (because vertex *d* is visited twice). The walk *abcde* is a path of length 4 between *a* and *e*.

Figure 1.2

Did u know? **What is closed walk?**

A closed walk or trail is a walk or trail starting and ending at the same vertex.
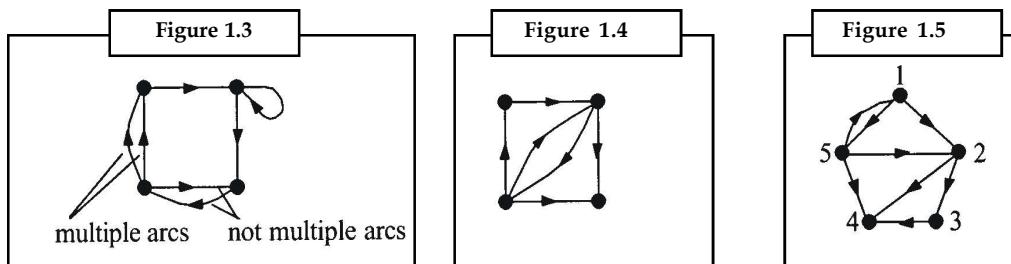
Digraphs

Digraphs are similar to graphs except that the edges have a direction. To distinguish them from undirected graphs the edges of a digraph are called arcs. Much of what follows is exactly the same as for undirected graphs with 'arc' substituted for 'edge'. We will attempt to highlight the differences.

A **digraph** consists of a set of elements, V, called **vertices**, and a set of elements, A, called **arcs**. Each arc joins two vertices in a specified direction.

Two or more arcs joining the same pair of vertices in the same direction are called **multiple arcs** (see Figure 1.3). NB two arcs joining the same vertices in opposite directions are not multiple arcs.

An arc joining a vertex to itself is a loop (see Figure 1.3).

Figure 1.3   Figure 1.4   Figure 1.5

A digraph with no multiple arcs and no loops is a **simple digraph** (e.g. Figure 1.4 and Figure 1.5).

Two vertices joined by an arc in either direction are **adjacent**.

Vertices are **incident** to and form the arc which joins them and an arc is incident to and from the vertices its joins.

Two digraphs G and H are **isomorphic** if H can be obtained by relabelling the vertices of G, i.e. there is a one-one correspondence between the vertices of G and those of H such that the number and direction of the arcs joining each pair of vertices in G is equal to the number and direction of the arc joining the corresponding pair of vertices in H.

A **subdigraph** of G is a digraph all of whose vertices and arcs are vertices and arcs of G.

The underlying graph of a digraph is the graph obtained by replacing of all the arcs of the digraph by undirected edges.

The **out-degree** of a vertex v is the number of arcs incident from v and the **in-degree** of a vertex v is the number of arcs incident to v. Loops count as one of each.

The **out-degree sequence** and **in-degree sequence** of a digraph G are the sequences obtained by listing, in ascending order with repeats, the out-degrees and in-degrees of the of the vertices of G.

The **Handshaking Lemma** states that the sum of the out-degrees and of the in-degrees of the vertices of a graph are both equal to the number of arcs. This is pretty obvious since every arc contributes one to the out-degree of the vertex from which it is incident and one to the in-degree of the vertex to which it is incident.

A **walk of length** k in a digraph is a succession of k arcs joining two vertices.

A **trail** is a walk in which all the arcs (but not necessarily all the vertices) are distinct.

A **path** is a walk in which all the arcs and all the vertices are distinct.

A connected digraph is one whose underlying graph is a connected graph. A **disconnected digraph** is a digraph which is not connected. A digraph is **strongly connected** if there is a path between every pair of vertices. Notice here we have a difference between graphs and digraphs. The underlying graph can be connected (a path of edges exists between every pair of vertices) whilst the digraph is not because of the directions of the arcs (see Figure 1.5 for a graph which is connected but not strongly connected).

A **closed walk/trail** is a walk/tail starting and ending at the same vertex.

A **cycle** is a closed path, i.e. a path starting and ending at the some vertex.

An **Eulerian digraph** is a connected digraph which contains a closed trail which includes every arc. The trail is called an **Eulerian trail**.

A **Hamiltonian digraph** is a connected digraph which contains a cycle which includes every vertex. The cycle is called an **Hamiltonian cycle**.

A connected digraph is Eulerian iff for each vertex the out-degree equals the in-degree. The proof of this is similar to the proof for undirected graphs.

An Eulerian digraph can be split into cycles no two of which have an arc in common. The proof of this is similar to the proof for undirected graphs.

Task   Analyse the difference between path and trail.

1.2 Multigraph

In mathematics, a **multigraph** or **pseudograph** is a graph which is permitted to have multiple edges, (also called "parallel edges"), that is, edges that have the same end nodes. Thus, two vertices may be connected by more than one edge. Formally, a multigraph G is an ordered pair $G:=(V, E)$ with

1. V a set of *vertices* or *nodes*,

2. E a multiset of unordered pairs of vertices, called *edges* or *lines*.

Multigraphs might be used to model the possible flight connections offered by an airline. In this case the multigraph would be a directed graph with pairs of directed parallel edges connecting cities to show that it is possible to fly both *to* and *from* these locations.

Some authors also allow multigraphs to have loops, that is, an edge that connects a vertex to itself, while others call these **pseudographs**, reserving the term multigraph for the case with no loops.

A **multidigraph** is a directed graph which is permitted to have *multiple arcs,* i.e., arcs with the same source and target nodes. A multidigraph G is an ordered pair $G:=(V,A)$ with

1. V a set of *vertices* or *nodes*,

2. A a multiset of ordered pairs of vertices called *directed edges*, *arcs* or *arrows*.
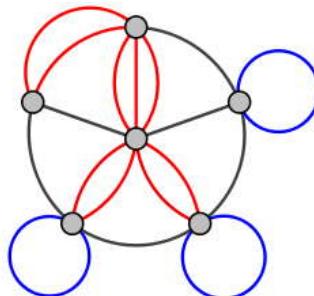
In category theory a small category can be defined as a multidigraph equipped with an associative composition law and a distinguished self-loop at each vertex serving as the left and right identity

for composition. For this reason, in category theory the term *graph* is standardly taken to mean "multidigraph", and the underlying multidigraph of a category is called its **underlying digraph**.

A **mixed multigraph** $G:=(V,E, A)$ may be defined in the same way as a mixed graph.

Figure 1.6
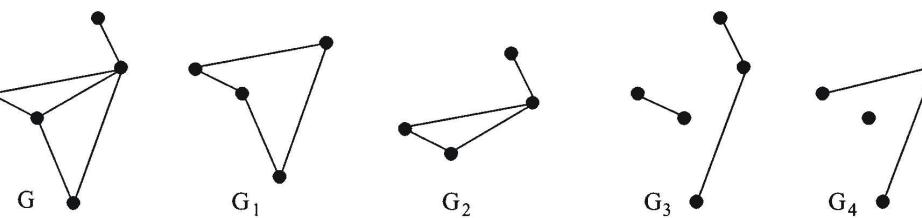
Notes   The Multigraphs and multidigraphs also support the notion of graph labeling, in a similar way. However, there is no unity in terminology in this case.

1.3 Subgraph

A **subgraph** of G is a graph all of whose vertices and edges are vertices and edges of G (Figure 1.7 shows a series of subgraph of G).

Figure 1.7
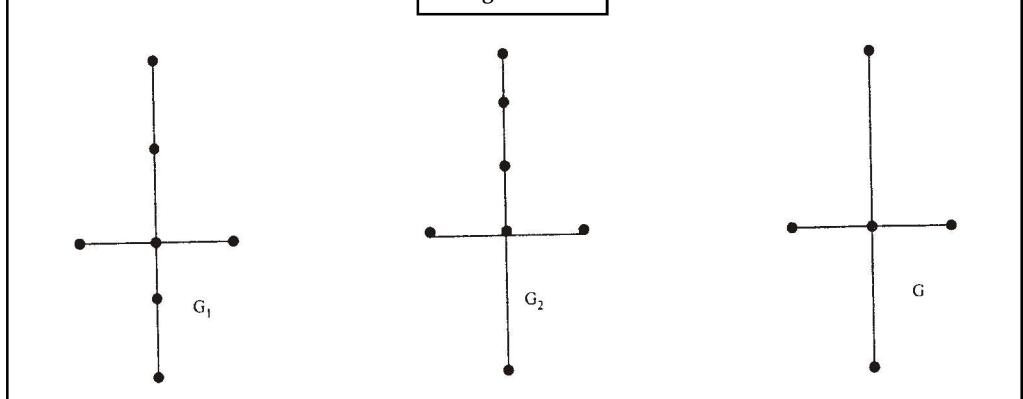
Did u know? **What is degree?**

The degree of a vertex is the number of edges incident.

1.4 Homomorphism Graphs

Let G_1 be a given graph. Another graph G_2 can be obtained from this graph by dividing an edge of G, with additional vertices.

Two graphs G_1 and G_2 are said to be Homeomorphic if these can be obtained from the same graph or isomorphic graphs by this method. Graphs G_1 and G_2, though not isomorphic, are Homeomorphic because each of them can be obtained from graph G by adding appropriate vertices.
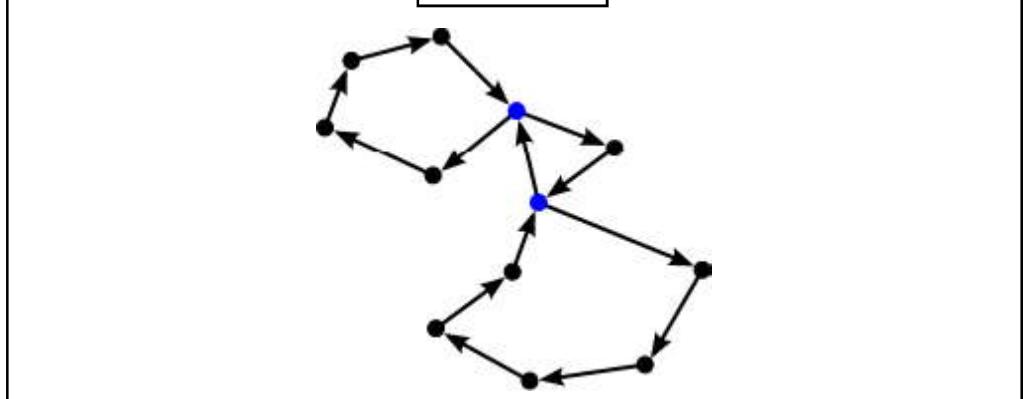
Figure 1.8

> *Notes*   They are Homeomorphic because each of them can be obtained from graph G by appropriate.

1.5 Path

In graph theory, a path in a graph is a sequence of vertices such that from each of its vertices there is an edge to the next vertex in the sequence. A path may be infinite, but a finite path always has a first vertex, called its start vertex, and a last vertex, called its end vertex. Both of them are called end or terminal vertices of the path. The other vertices in the path are internal vertices. A cycle is a path such that the start vertex and end vertex are the same. Note that the choice of the start vertex in a cycle is arbitrary.

Paths and cycles are fundamental concepts of graph theory, described in the introductory sections of most graph theory texts. See e.g. Bondy and Murty (1976), Gibbons (1985), or Diestel (2005). Korte et al. (1990) cover more advanced algorithmic topics concerning paths in graphs.

Figure 1.9

Different Types of Path

The same concepts apply both to undirected graphs and directed graphs, with the edges being directed from each vertex to the following one. Often the terms directed path and directed cycle are used in the directed case.

A path with no repeated vertices is called a simple path, and a cycle with no repeated vertices or edges aside from the necessary repetition of the start and end vertex is a simple cycle. In modern graph theory, most often "simple" is implied; i.e., "cycle" means "simple cycle" and "path" means "simple path", but this convention is not always observed, especially in applied graph theory. Some authors (e.g. Bondy and Murty 1976) use the term "walk" for a path in which vertices or edges may be repeated, and reserve the term "path" for what is here called a simple path.

A path such that no graph edges connect two nonconsecutive path vertices is called an induced path.

A simple cycle that includes every vertex, without repetition, of the graph is known as a Hamiltonian cycle.

A cycle with just one edge removed in the corresponding spanning tree of the original graph is known as a Fundamental cycle.

Two paths are independent (alternatively, internally vertex-disjoint) if they do not have any internal vertex in common.

The length of a path is the number of edges that the path uses, counting multiple edges multiple times. The length can be zero for the case of a single vertex.

Did u know? **What is weighted graph?**

A weighted graph associates a value (weight) with every edge in the graph. The weight of a path in a weighted graph is the sum of the weights of the traversed edges. Sometimes the words cost or length are used instead of weight.

1.6 Connectivity

In mathematics and computer science, **connectivity** is one of the basic concepts of graph theory. It is closely related to the theory of network flow problems. The connectivity of a graph is an important measure of its robustness as a network.

1.6.1 Definitions of Connectivity

In an undirected graph G, two vertices u and v are called connected if G contains a path from u to v. Otherwise, they are called disconnected. If the two vertices are additionally connected by a path of length 1, i.e. by a single edge, the vertices are called adjacent. A graph is said to be connected if every pair of vertices in the graph are connected.

A connected component is a maximal connected subgraph of G. Each vertex belongs to exactly one connected component, as does each edge.

A directed graph is called weakly connected if replacing all of its directed edges with undirected edges produces a connected (undirected) graph. It is connected if it contains a directed path from u to v or a directed path from v to u for every pair of vertices u, v. It is strongly connected or strong if it contains a directed path from u to v and a directed path from v to u for every pair of vertices u, v. The strong components are the maximal strongly connected subgraphs.

A cut, vertex cut, or separating set of a connected graph G is a set of vertices whose removal renders G disconnected. The connectivity or vertex connectivity κ (G) is the size of a smallest vertex cut. A graph is called k-connected or k-vertex-connected if its vertex connectivity is k or greater. This means a graph G is said to be k-connected if there does not exist a set of k-1 vertices

whose removal disconnects the graph. A complete graph with n vertices has no vertex cuts at all, but by convention its connectivity is n-1. A vertex cut for two vertices u and v is a set of vertices whose removal from the graph disconnects u and v. The local connectivity $\kappa(u, v)$ is the size of a smallest vertex cut separating u and v. Local connectivity is symmetric for undirected graphs; that is, $\kappa(u, v) = \kappa(v, u)$. Moreover, except for complete graphs, $\kappa(G)$ equals the minimum of $\kappa(u, v)$ over all nonadjacent pairs of vertices u, v.

2-connectivity is also called biconnectivity and 3-connectivity is also called triconnectivity.

Analogous concepts can be defined for edges. In the simple case in which cutting a single, specific edge would disconnect the graph, that edge is called a bridge. More generally, the edge cut of G is a group of edges whose total removal renders the graph disconnected. The edge-connectivity $\lambda(G)$ is the size of a smallest edge cut, and the local edge-connectivity $\lambda(u, v)$ of two vertices u, v is the size of a smallest edge cut disconnecting u from v. Again, local edge-connectivity is symmetric.

⚠

Caution A graph is called k-edge-connected if its edge connectivity is k or greater.

1.6.2 Menger's Theorem

One of the most important facts about connectivity in graphs is Menger's theorem, which characterizes the connectivity and edge-connectivity of a graph in terms of the number of independent paths between vertices.

If u and v are vertices of a graph G, then a collection of paths between u and v is called independent if no two of them share a vertex (other than u and v themselves). Similarly, the collection is edge-independent if no two paths in it share an edge. The greatest number of independent paths between u and v is written as $\kappa^2(u,v)$, and the greatest number of edge-independent paths between u and v is written as $\lambda^2(u,v)$.

Menger's theorem asserts that the local connectivity $\kappa(u,v)$ equals $\kappa^2(u,v)$ and the local edge-connectivity $\lambda(u,v)$ equals $\lambda^2(u,v)$ for every pair of vertices u and v. This fact is actually a special case of the max-flow min-cut theorem.

1.6.3 Computational Aspects

The problem of determining whether two vertices in a graph are connected can be solved efficiently using a search algorithm, such as breadth-first search. More generally, it is easy to determine computationally whether a graph is connected (for example, by using a disjoint-set data structure), or to count the number of connected components. A simple algorithm might be written in pseudo-code as follows:

1. Begin at any arbitrary node of the graph, G.

2. Proceed from that node using either depth-first or breadth-first search, counting all nodes reached.

3. Once the graph has been entirely traversed, if the number of nodes counted is equal to the number of nodes of G, the graph is connected; otherwise it is disconnected.

By Menger's theorem, for any two vertices u and v in a connected graph G, the numbers $\kappa(u,v)$ and $\lambda(u,v)$ can be determined efficiently using the max-flow min-cut algorithm. The connectivity and edge-connectivity of G can then be computed as the minimum values of $\kappa(u,v)$ and $\lambda(u,v)$, respectively.

In computational complexity theory, SL is the class of problems log-space reducible to the problem of determining whether two vertices in a graph are connected, which was proved to be equal to L by Omer Reingold in 2004. Hence, undirected graph connectivity may be solved in O(logn) space.

The problem of computing the probability that a Bernoulli random graph is connected is called Network reliability and the problem of computing whether two given vertices are connected the ST-reliability problem. Both of these are #P-hard.

Examples:
1. The vertex and edge-connectivities of a disconnected graph are both 0.
2. 1-connectedness is synonymous with connectedness.
3. The complete graph on n vertices has edge-connectivity equal to n–1. Every other simple graph on n vertices has strictly smaller edge-connectivity.
4. In a tree, the local edge-connectivity between every pair of vertices is 1.

1.6.4 Bounds on Connectivity

1. The vertex-connectivity of a graph is less than or equal to its edge-connectivity. That is, $\kappa(G) \leq \lambda(G)$. Both are less than or equal to the minimum degree of the graph, since deleting all neighbors of a vertex of minimum degree will disconnect that vertex from the rest of the graph.

2. For a vertex-transitive graph of degree d, we have: $2(d+1)/3 \leq \kappa(G) \leq \lambda(G) = d$.[4]

3. For a vertex-transitive graph of degree $d \leq 4$, or for any (undirected) minimal Cayley graph of degree d, or for any symmetric graph of degree d, both kinds of connectivity are equal: $\kappa(G) = \lambda(G) = d$.

1.6.5 Other Properties

1. Connectedness is preserved by graph homomorphisms.

2. If G is connected then its line graph L(G) is also connected.

3. If a graph G is k-connected, then for every set of vertices U of cardinality k, there exists a cycle in G containing U. The converse is true when k = 2.

4. A graph G is 2-edge-connected if and only if it has an orientation that is strongly connected.

5. Balinski's theorem states that the polytopal graph (1-skeleton) of a k-dimensional convex polytope is a k-vertex-connected graph. As a partial converse, Steinitz showed that any 3-vertex-connected planar graph is a polytopal graph (Steinitz theorem).

1.7 Summary

● A graph is a mathematical structure comprising a set of vertices, V, and a set of edges, E, which connect the vertices.

● An edge in a connected graph is a **bridge** if its removal leaves a disconnected graph.

● Vertices are **incident** with the edges which joins them and an edge is **incident** with the vertices it joins.

● A **path graph** is a tree consisting of a single path through all its vertices.

1.8 Keywords
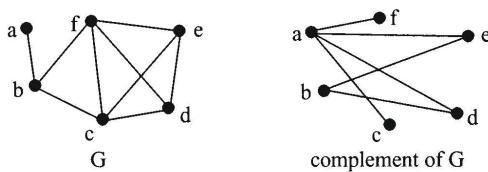
Incident: Vertices are incident with the edges which join them and an edge is incident with the vertices it joins.

Tree: A tree is a connected graph with no cycles.
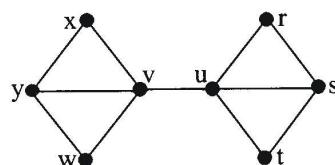
1.9 Self Assessment

1. Draw the graphs whose vertices and edges are as follows. In each case say if the graph is a simple graph.

    (a) V = {u, v, w, x}, E = {uv, vw, wx, vx}

    (b) V = {1, 2, 3 4, 5, 6, 7, 8}, E = {12, 22, 23, 34, 35, 67, 68, 78}

    (c) V = {n, p, q, r, s, t}, E = {np, nq, nt, rs, rt, st, pq}

2. Use the Handshaking Lemma to prove that the number of vertices of odd degree in any graph must be even.

3. This is a more challenging question than the first sixteen.

    The complement of a simple graph G is the graph obtained by taking the vertices of G (without the edges) and joining every pair of vertices which are not joined in G. For instance
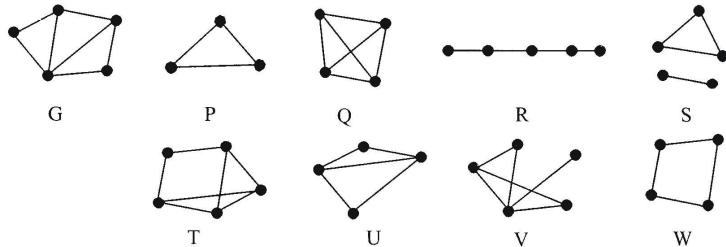
    G   complement of G

    (a) Verify that the complement of the path graph P_4 is P_4.

    (b) What are the complement of K_4, $K_{3,3}$, C_5 ?

    (c) What is the relationship between the degree sequence of a graph and that of its complement?

    (d) Show that if a simple graph G is isomorphic to its complement then the number of vertices of G has the form $4k$ or $4k + 1$ for some integer k.

    (e) Find all the simple graphs with 4 or 5 vertices which are isomorphic to their complements.

    (f) Construct a graph with eight vertices which is isomorphic to its complement.

4. Complete the following statements with walks/trail/path:
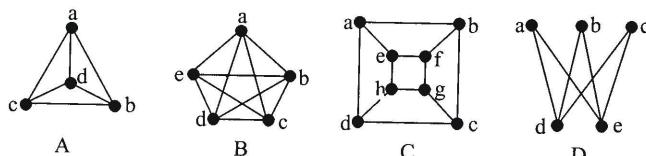
    (a) *wvyxvurs* is a .... of length .... between *w* and *s*

    (b) *vxvur* is a .... of length .... between *v* and *r*

(c)   *uvyxvw* is a .... of length .... between *u* and *w*

(d)   *ruvwy* is a .... of length .... between *r* and *y*

5.   Which of the graphs P, Q, ... W are subgraphs of G?
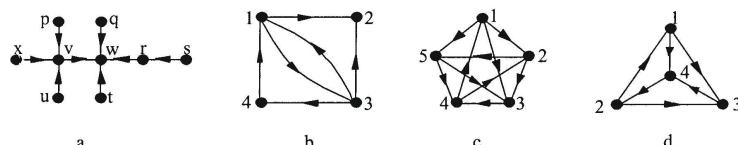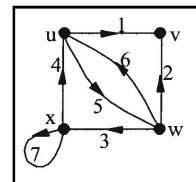
6.   Which of the following graphs are Eulerian and/or Hamiltonian. Give an Eulerian trail or a Hamiltonian cycle where possible.
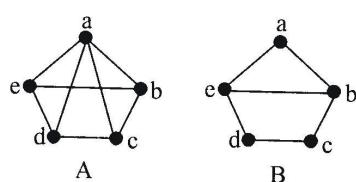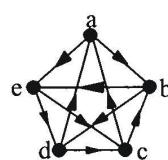
1.10 Review Questions

1.   Digraphs G_1 and G_2 have the same in-degree and out-degree sequence — are they necessarily isomorphic? If your answer is no, give a counter example.

2..   Digraphs G_1 and G_2 are isomorphic. Do they necessarily have the same in-degree and out-degree sequences? If your answer is no, give a counter example.

3.   In the digraph shown give (if possible)

    (a)   A walk of length 7 from *u* to *w*,

    (b)   Cycles of length 1, 2, 3 and 4, and

    (c)   A path of maximum length

4.   Which of the following connected digraphs are strongly connected?
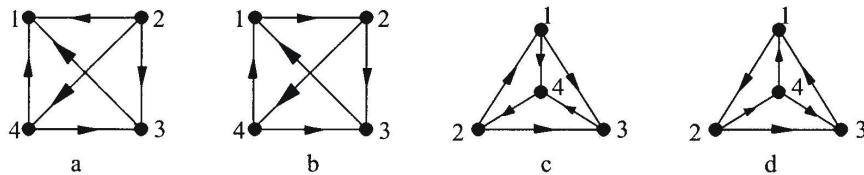
5.   In the digraph shown, find

    (a)   all the cycles of length 3, 4, and 5,

    (b)   an Eulerian trail, and

    (c)   a Hamiltonian cycle.
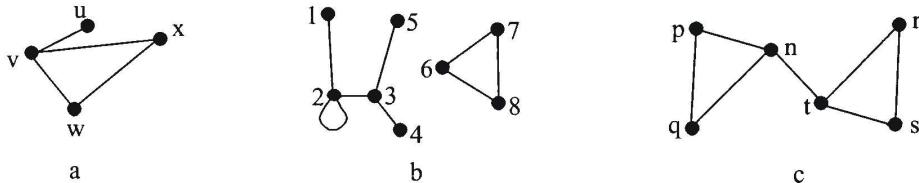
6.   Check whether the conditions of Ore's theorem hold for these Hamiltonian graphs.

7. Draw the digraphs whose vertices and arcs are as follows. In each case say if the digraph is a simple digraph.

    (a)   V = {u, v, w, x}, A = {vw, wu, wv, wx, xu}

    (b)   V = {1, 2, 3, 4, 5, 6, 7, 8}, A = {12, 22, 23, 24, 34, 35, 67, 68, 78}

    (c)   V = {n, p, q, r, s, t}, A = {np, nq, nt, rs, rt, st, pq, tp}

8. Which two of the following digraphs are identical, which one is isomorphic to the two identical ones and which is unrelated? Write down the in-degree sequence and the out-degree sequence for each digraph.

Answers: Self Assessment

1.

2. The Handshaking Lemma states that the sum of the degrees of all the vertices is twice the number of edges. Hence the degree sum is even. Since the sum of any number of even numbers is and the sum of an even number of odd numbers is even, whilst the sum of an odd number of odd numbers is odd, the degree sum must be the sum of any number of even numbers and an even number of odd numbers. So the number of vertices of odd degree is even.

3. (a)   It can readily be seen that the complement of P_4 and P_4.
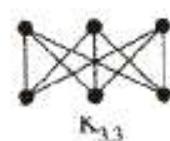
P_4 complement of P_4
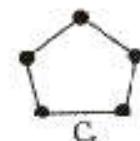
    (b)   It can readily be seen that complement of K_4 is N_4, the complement of $K_{3,3}$ is a disjoint graph comprising a pair of C_3 s and the complement of C_5 is C_5.

K_4 $K_{3,3}$ C_5

complement of K_4 complement of $K_{3,3}$ complement of C_5
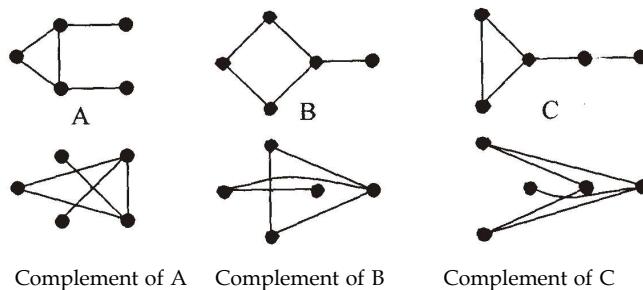
(c) Each vertex in the complement of G(V, E) is connected to every vertex in V to which it is not connected in G. Hence the sum of degree of a given vertex in G and its degree in the complement of G is one less than the number of vertices of G. So, if the graph has n vertices and degree sequence $(d_1, d_2, ... d_n)$ then the degree sequence of the complement is $(n - 1 - d_n, n - 1 - d_{n-1}, ... n - 1 - d_1)$.

(d) The edges of a graph G and of its complement make a complete graph so the sum of the vertex degrees of G and its complement is $n(n - 1)$. If a graph is isomorphic to its complement then the degree sequences of G and its complement are identical so the degree sums of G and its complement are equal. Hence the number of edges in G must be $n(n - 1)/4$. This must be an integer, k say, so we must have $n/4 = k$ or $(n - 1)/4 = k$, i.e. $n = 4k$ or $4k + 1$.

(e) For a graph with 4 vertices and a degree sequence (p, q, r, s) the degree sequences of the complement is $(3\text{-}s, 3\text{-}r, 3\text{-}q, 3\text{-}p)$. If the graph is isomorphic to its complement then the degree sequences are identical so we have $p = 3\text{-}s$ and $q = 3\text{-}r$. The degree sequences are in ascending order so the only possible sequences are $(0, 0, 3, 3)$, $(0, 1, 2, 3)$, $(1, 1, 2, 2)$. For $(0, 0, 3, 3)$ only 2 vertices have incident edges so there must be multiple edges or loops and the graph is not simple. For $(0, 1, 2, 3)$ only 3 vertices have incident edges so any vertex of degree greater than 2 is incident to multiple edges or a loop and so the graph is not simple. The only graph with degree sequence $(1, 1, 2, 3)$ is the path graph P_4 and we have already shown that P_4 is isomorphic to its complement (part (a)).
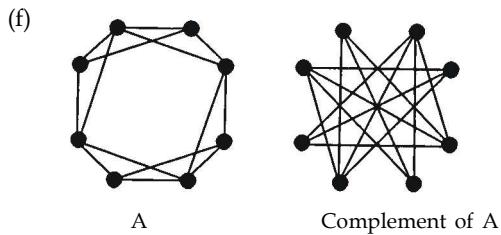
For a graph with 5 vertices and a degree sequence (p, q, r, s, t) the degree sequence of the complement is $(4\text{-}t, 4\text{-}s, 4\text{-}r, 4\text{-}q, 4\text{-}p)$. If the graph is isomorphic to its complement then the degree sequences are identical so we have $p = 4\text{-}t$, $q = 4\text{-}s$ and $r = 4\text{-}r$, that is $p + t = 4$, $q + s = 4$ and $r = 2$. The degree sequences are in ascending order so the only possible sequences are $(0, 0, 2, 4, 4)$, $(0, 1, 2, 3, 4)$, $(0, 2, 2, 2, 4)$, $(1, 1, 2, 3, 3)$, $(1, 2, 2, 2, 3)$, $(2, 2, 2, 2, 2)$. For $(0, 0, 2, 4, 4)$ only 3 vertices have incident edges so any vertex of degree greater than 2 incident to multiple edges or a loop and so the graph is not simple. For $(0, 1, 2, 3, 4)$ and $(0, 2, 2, 2, 4)$ only 4 vertices have incident edges so any vertex of degree greater than 3 is incident to multiple edges or a loop and so the graph is not simple. The only graph with degree sequence $(2, 2, 2, 2, 2)$ is the cycle graph C_5 and we have already shown that C_5 is isomorphic to its complement (part (b)). The only simple graph with degree sequence $(1, 1, 2, 3, 3)$ is A shown below and is isomorphic to its complement. The only simple graphs with degree sequence $(1, 2, 2, 2, 3)$ are B and C shown below and the complement of B is C and, of course, vice versa.

So the only simple graph with 4 vertices which is isomorphic to its complement is P_4 and the only simple graphs with 5 vertices which are isomorphic to their complements are C_5 and the graph A below.

A   B   C

Complement of A   Complement of B   Complement of C

(f)

A   Complement of A

4.   (a)   *wvyxvurs* is a trail of length 7 between *w* and *s*.

     (b)   *vxvur* is a walk of length 4 between *v* and *r*.

     (c)   *uvyxvw* is a trail of length 5 between *u* and *w*.

     (d)   *ruvwy* is a path of length 4 between *r* and *y*.

5.   P, R, S, U, W are subgraphs of G.

6.   (a)   This graph is 3-regular with 4 vertices. It is not Eulerian because its vertices all have odd degree. It has a Hamiltonian cycle *abcda* and so is Hamiltonian.

     (b)   This graph is 4-regular. All its vertices are of even degree so it is Eulerian. An Eulerian trail is *abecadbcdea*. It has a Hamiltonian cycle *abcdea* and so is Hamiltonian.

     (c)   This graph is 3-regular. All its vertices are of odd degree so it is not Eulerian. It has a Hamiltonian cycle *abcdhgfea* and so is Hamiltonian.

     (d)   This graph has 2 vertices of odd degree so it is not Eulerian. It is the complete bipartite graph $K_{2,3}$ so to construct a Hamiltonian cycle we necessarily have to visit vertices from the set {*a*, *b*, *c*} and the set {*d*, *e*} alternately. Start at any vertex in {*a*, *b*, *c*}, go to a vertex in {*d*, *e*} then to a different vertex in {*a*, *b*, *c*} then to the other vertex in {*d*, *e*} then to the only unvisited vertex in {*a*, *b*, *c*}. Now in order to get back to the starting vertex we must visit another vertex in {*d*, *e*}. But we have visited both of those already so we cannot return to the start without revisiting a vertex which is already in the walk. Thus no Hamiltonian cycle exists and the graph is not Hamiltonian.

1.11 Further Readings

Books   Béla Bollobás, *Modern Graph Theory*, Springer

            Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph-theory, Combinatorics, and Algorithms*, Birkhäuser

Online links   http://en.wikipedia.org/wiki/Multigraph

            http://en.wikipedia.org/wiki/Diagraph

Unit 2: Types of Graphs

Objectives

After studying this unit, you will be able to:

- Describe the different types of graphs

- Calculate the matrix representation

Introduction

In mathematics and computer science, **graph theory** is the study of *graphs*, mathematical structures used to model pair-wise relations between objects from a certain collection. A "graph" in this context refers to a collection of vertices or 'nodes' and a collection of *edges* that connect pairs of vertices. A graph may be *undirected*, meaning that there is no distinction between the two vertices associated with each edge, or its edges may be *directed* from one vertex to another; see graph (mathematics) for more detailed definitions and for other variations in the types of graphs that are commonly considered. The graphs studied in graph theory should not be confused with graphs of functions or other kinds of graphs.

2.1 The Bridge of Konigsberg

Dr. Martin mentioned, in his introduction, the classic Kōnigsberg Bridges problem. Kōnigsberg lies at the confluence of two rivers. In the middle of the town lies an island, the Kneiphof. Spanning the rivers are seven bridges as shown in Figure 2.1.

Figure 2.1

Figure 2.2

The citizens of Kōnigsberg used to amuse themselves, during their Sunday afternoon strolls, trying to devise a route which crossed once only the returned to its starting point. Leonhard Euler finally proved (during the 1730s) that the task was impossible. The graph in Figure 2.2 where the vertices represent the different areas of land and the edges the bridges. What the citizens of Kōnigsberg were seeking was an Eulerian trail through this graph. The degree sequence of the graph in Figure 2.2 is (3, 3, 3, 5) so it is not Eulerian and no Eulerian trail exists.

Notes   If the citizens of Kōnigsberg relaxed the requirement to finish up where they started (in other words, to seek a semi-Eulerian trail) they would still fail since theorem tells us that a graph a semi-Eulerian if every vertex bar two has even degree. Figure 2.2 fails this test also.

2.2 Transversal Multigraphs

Graph traversal refers to the problem of visiting all the nodes in a graph in a particular manner. Tree traversal is a special case of graph traversal. In contrast to tree traversal, in general graph traversal, each node may have to be visited more than once, and a root-like node that connects to all other nodes might not exist.

Depth-first Search (DFS)

A Depth-first Search is a technique for traversing a finite undirected graph. DFS visits the child nodes before visiting the sibling nodes, that is, it traverses the depth of the tree before the breadth.

Breadth-first Search (BFS)

A Breadth-first Search is another technique for traversing a finite undirected graph. BFS visits the sibling nodes before visiting the child nodes.
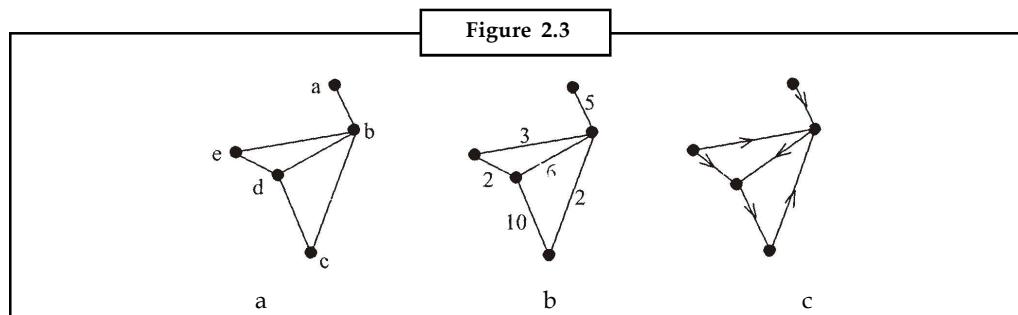
2.3 Different Types of Graphs
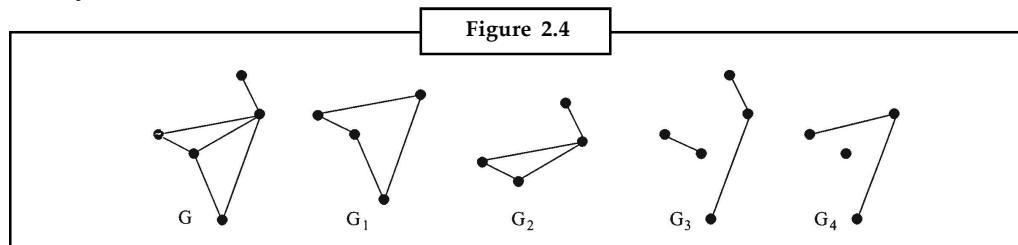
In a **labelled graph** the vertices have labels or names [Figure 2.3 (a)].

In a **weighted graph** each edge has a **weight** associated with it [Figure 2.3 (b)].
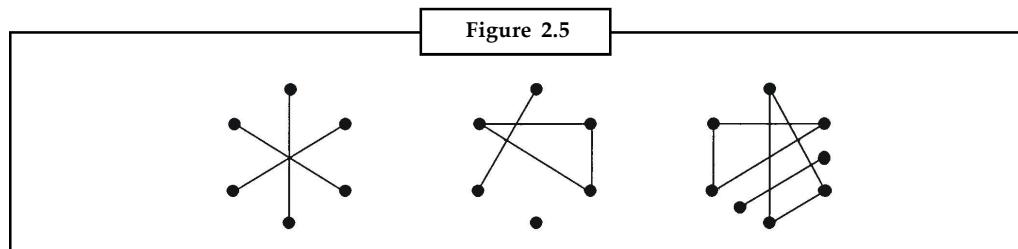
A **digraph** (**directed graph**) is a diagram consisting of points, called vertices, joined by directed lines, called **arcs** [Figure 2.3 (c)].

Figure 2.3

a   b   c

A **connected graph** has a path between every pair of vertices. A **disconnected graph** is a graph which is not connected. e.g., Figure 2.4, G and the subgraphs G_1 and G_2 are connected whilst G_3 and G_4 are disconnected.

Figure 2.4

G G_1 G_2 G_3 G_4

Every disconnected graph can be split into a number of connected subgraphs called its components. It may not be immediately obvious that a graph is disconnected. For instance Figure 2.5 shows 3 graphs, each disconnected and comprising 3 components.

Figure 2.5

An edge in a connected graph is a **bridge** if its removal leaves a disconnected graph.

A **closed walk** or **closed trail** is a walk or trail starting and ending at the same vertex.

A **cycle** is a closed path, i.e. a path starting and ending at the same vertex.

Walks/trails/paths which are not closed are **open**.
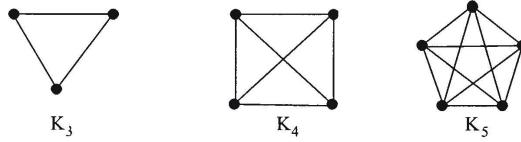
In a **regular graph** all vertices have the same degree. If the degree is *r* the graph is ***r*-regular**.

If G is *r*-regular with *m* vertices it must have 1/2 *mr* edges (from the Handshaking Theorem).

A **complete graph** is a graph in which every vertex is joined to every other vertex by exactly one edge. The complete graph with *m* vertices is denoted by K_m. K_m is (*m* – 1)-regular and so has $1/2m$ (*m* – 1) edges.

Figure 2.6

K_3 K_4 K_5
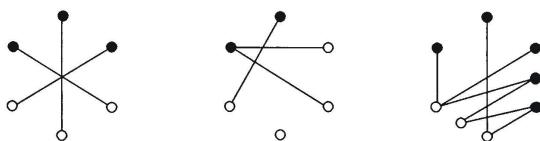
A null graph is a graph with no edges. The null graph with in vertices is denoted Nm is o-regular.

A cycle graph consines of a single cycle of vertices a edges. The cycle graph with m vertices is denoted cm
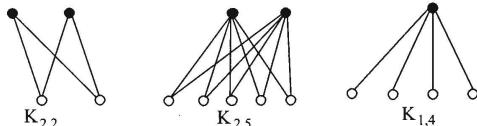
A **bipartite graph** is a graph whose vertices can be split into two subsets A and B in such a way that every edge of G joins a vertex in A with one in B. Figure 2.7 shows some bipartite graphs. Notice that the allocation of the nodes to the sets A and B can sometimes be done in several ways.

Figure 2.7
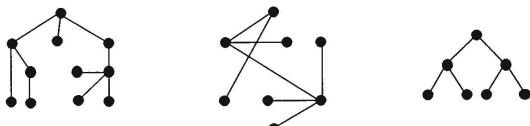
A **complete bipartite graph** is a bipartite graph in which every vertex in A is joined to every vertex in B by exactly one edge. The complete bipartite graph with r vertices in A and s vertices in B is denoted $K_{r,s}$. Figure 2.8 shows some complete bipartite graphs.

Figure 2.8

$K_{2,2}$ $K_{2,5}$ $K_{1,4}$
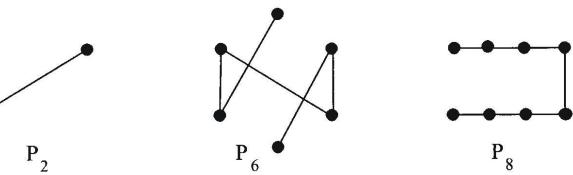
A **tree** is a connected graph with no cycles. In a tree there is just one path between each pair of vertices. Figure 2.9 shows some trees. Every tree is a bipartite graph. Start at any node, assign each node connected to that node to the other set, then repeat the process with those nodes!

Figure 2.9

A **path graph** is a tree consisting of a single path through all its vertices. The path graph within vertices is denoted P_m. Figure 2.10 shows some path graphs.

Figure 2.10

P_2 P_6 P_8
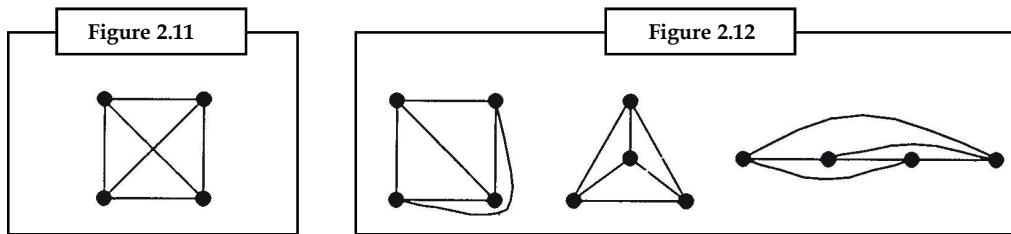
Did u know? **Is a null graph is a graph with no edges?**

Yes a null graph is a graph with no edges.

2.4 Planar Graphs

A graph G is **planar** if it can be drawn on a plane in such a way that is no two edges meet except at a vertex with which they both incident. Any such drawing is a **plane drawing** of G. A graph is non-planar if no plane drawing of it exists. Figure 2.11 shows a common representation of the graph K_4 and Figure 2.12 shows three possible plane drawings of K_4.

Figure 2.11

Figure 2.12

The complete bipartite graph $K_{3,3}$ and the complete graph K_5 are important simple examples of non-planar graphs.

Any plane drawing of a planar graph G divides the set of points of the plane not lying on G into regions called **faces**. The region outside the graph is of infinite extent and is called the **infinite face**.

The **degree** of a face f of a connected planar graph G, denoted deg (f), is the number of edges encountered in a walk around the boundary of f. If all faces have the same degree, g, then G is **face-regular of degree** g.

In Figure 2.12 each plane drawing of the graph K_4 has 4 faces (including the infinite face) each face being of degree 3 so K_4 is face-regular of degree 3.

In any plane drawing of a planar graph, the sum of all the face degrees is equal to twice the number of edges.

In Figure 2.12 the plane drawing of the graph K_4 has 4 face (including the infinite face) each face being of degree 3 so K_4 is face-regular of degree 3. The sum of the face degrees is therefore 12 whilst the number of edges is 6.

Euler's formula for planar graphs. If n, m and f denote respectively the number of vertices, edges and faces in a plane drawing of a connected plane graph G then $n - m + f = 2$.

In Figure 2.12 we have $n = 4$, $m = 6$ and $f = 4$ satisfying Euler's formula.

Proof of Euler's formula : A plane drawing of any connected planar graph G can be constructed by taking a spanning tree of G and adding edges to it, one at a time, until a plane drawing of G is obtained.

We prove Euler's formula by showing that:

1.   for any spanning tree G, $n - m + f = 2$, and

2.   adding an edge to the spanning tree does not change the value of $n - m + f$.

Let T be any spanning tree of G. We may draw T in a plane without crossings. T has *n* vertices *n* – 1 edges and 1 face (the infinite face). Thus $n - m + f = n - (n - 1) + 1 = 2$ so we have shown (*a*).

Now, if we add an edge to T either it joins two different vertices, or it joins a vertex to itself (it is a loop). In either case it divides some face into two faces, so adding one face. Hence we have increased *m*, the number of edges, and *f*, the number of faces, by one each. The value of the expression $n - m + f$ is unchanged. We add more edges, one at a time, and at each addition the value of $n - m + f$ is unchanged. Hence we have shown (b) and so proved Euler's theorem.

⚠️

Caution It is useful to be able to test a graph for planarity. There are a variety of algorithms for determining planarity, mostly quite complex. Here we will describe a simple test, the **cycle method**, which determines the planarity of a Hamiltonian graph.

First we identify a Hamiltonian cycle, C, in the graph. Now list the remaining edges of the graph, and then try to divide those edges into two disjoint sets A and B such that
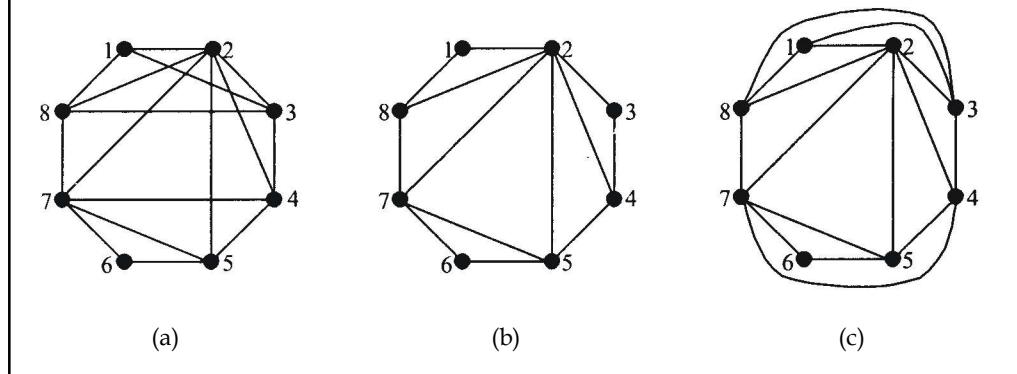
A is a set of edges which can be drawn inside C without crossings, and

B is a set of edges which can be drawn outside C without crossings.

📝

Example: Determine whether the graph in Figure 2.13 (a) is planar.

Figure 2.13

(a)   (b)   (c)

First find a Hamiltonian cycle. 123456781 will do. Draw a plane drawing of the cycle. The remaining edges are {24, 25, 27, 28, 31, 38, 47, 57}. Take the first edge, 24, and put it in set A. Take the second edge, 25, and put it in set A if compatible—it is so put it in set A. Consider the next edge, 27—it is comparable with set A so add it to set A. At this point we have A = {24, 25, 27}, B = { } . The next edge is 28—it is compatible with set A so add it to set A (A = {24, 25, 27, 28}). The next edge is 31 which is not compatible with set A so put it in set B(B = {31}). The next edge is 38 which is not compatible with set A so put it in set B(B = {31, 38}). The next edge is 47 which is not compatible with set A so put it in set B (B = {31, 38, 47}). The next edge 57 which is compatible with set A so put it in set A(A = 24, 25, 27, 28, 57}). Figure 2.13 (b) shows the Hamiltonian cycle 123456781 and the edges in set A drawn inside the cycle. Now if we can add the edges from set B, all outside the cycle, without crossings then we have a plane drawing of the graph and it will be planar. Figure 2.13 (c) shows that the edges in set B can be drawn in that way so the graph is planar and Figure 2.13 (c) is a plane drawing.

2.5 Matrix Representation

The adjacency matrix A(G) of a graph G with n vertices is an n x n matrix with a_{ij} being the number of edges joining of edges joining the vertices i and j.

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{bmatrix}$$

The adjacency matrix A(D) of a digraph D with n vertices is an n x n matrix with a_{ij} being the number of arcs from vertex i to vertex j.

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

What is the number of walks of length 2 from vertex i to vertex j? There are, for instance, two walks of length 2 from 1 to 4. And so on. The matrix of these is

$$\begin{bmatrix} 0 & 1 & 1 & 2 \\ 0 & 1 & 2 & 2 \\ 0 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

and we see that this is just A(D)^2. In fact this generalises.

Theorem 1. The number of walks of length k from vertex i to vertex j in a digraph D with n vertices is given by the ijth element of the matrix A^k where A is the adjacency matrix of the digraph.

Proof. We prove this result by mathematical induction.

Assume that the result is true for $k \leq K - 1$. We will show that it is then also true for K. Consider any walk from vertex i to vertex j of length K. Such a walk consists of a walk of length K – 1 from vertex i to a vertex p which i adjacent to vertex j followed by a walk of length 1 from vertex p to vertex j. The number of such walks is $[A^{K-1}]_{ip} \times A_{pj}$. The total number of walks of length k from vertex i to vertex j will then be the sum of the walks through any p, *i.e.* $\sum_{p=1}^{n} [A^{K-1}]_{ip} A_{pj}$ but this is just the expression for the ij'th element of the matrix $A^{K-1}A = A^K$ so the result is true for $k = K$. But the result is certainly true for walks of length 1, i.e. $k = 1$, because that is the definition of the adjacency matrix A. Hence the theorem is true for all k.

Now we can create a method of automating the determination of whether a given digraph is strongly connected or not. For the graph to be strongly connected there must be paths (of any length) from every vertex to every other vertex. The length of these paths cannot exceed $n - 1$ where n is the number of vertices in the graph (otherwise a path would be visiting at least one vertex at least twice). So the number of paths from a vertex i to a vertex j of any length from 1 to $n - 1$ is the sum of the ijth elements of the matrices A, A^2, A^3, ... A^{n-1}. So we introduce the matrix B = A + A^2 + A^3 + ... A^{n-1} whose element B$_{ij}$ represent the number of paths between all the vertices.

If any off-diagonal element of B is zero then there are no paths from some vertex i to some vertex j. The digraph is strongly connected provided all the off-diagonal elements are non-zero!

Theorem 2. If A is the adjacency matrix of a digraph D with n vertices and B is the matrix $B = A + A^2 + A^3 + ... A^{n-1}$ then D is strongly connected iff each nondiagonal element of B is greater then 0.

Proof. To prove this theorem we must show both "if each non diagonal element of B is greater then 0 then D is strongly connected" and "if D is strongly connected then each non diagonal element of B is greater than 0".

Firstly, let D be a digraph and suppose that each non-diagonal element of the matrix B > 0, i.e. B_{ij} > 0 for all $i \neq j$. Then $[A^k]_{ij}$ > 0 for some $k \in [1, n-1]$, i.e. there is a walk of some length k between 1 and $n-1$ from every vertex i to every vertex j. So the digraph is strongly connected.

Secondly, suppose the digraph is strongly connected. Then, by definition, there is a path from every vertex i to every j. Since the digraph has n vertices the path is of length no more than $n-1$.

Hence, for all $i \neq j$, $[A^k]_{ij}$ > 0 for some $k \leq n-1$. Hence, for all $i \neq j$, B_{ij} > 0.

Returning to the example of the digraph D we have

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, A^2 = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 0 & 1 & 2 & 2 \\ 0 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 \end{bmatrix}, A^3 = \begin{bmatrix} 0 & 2 & 2 & 2 \\ 0 & 3 & 2 & 2 \\ 0 & 1 & 2 & 2 \\ 0 & 1 & 0 & 2 \end{bmatrix},$$

$$B = A + A^2 + A^3 = \begin{bmatrix} 0 & 4 & 3 & 5 \\ 0 & 5 & 4 & 6 \\ 0 & 3 & 2 & 4 \\ 0 & 2 & 1 & 2 \end{bmatrix}$$

so the graph is not strongly connected because we cannot get from vertex 2, vertex 3 or vertex 4 to vertex 1. Inspecting the digraph that is intuitively obviously! But, of course, this method is valid for large and complex digraphs which are less amenable to ad hoc analysis.

If we are only interested in whether there is at least one path from vertex i to vertex j (rather than wanting to know how many paths), then all of this can also be done using Boolean matrices. In this case the Boolean matrix (in which $a_{ij} = 1$ if there is at least one arc from vertex i to vertex j and 0 otherwise) of the graph is

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and the calculation of A^2, A^3 etc. is done using Boolean arithmetic (so × is replaced by $\wedge$ and + by $\vee$) so

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, A^2 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}, A^3 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$R = A \cup A^2 \cup A^3 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

A^2 is a matrix in which $(A^2)_{ij} = 1$ if there is at least one walk of length 2 from vertex i to vertex j and 0 otherwise and R is a matrix in which $R_{ij} = 1$ if there is at least one walk of length less than n from vertex i to vertex j and 0 otherwise. R is called the **reachability matrix**. In general R is easier and quicker to compute than B because it uses Boolean arithmetic. But, better, there is an even faster method called Warhsall's algorithm.

Let D = (V, A) where V = $\{v_1, v_2, ... v_n\}$ and there are no multiple arcs. Warshall's algorithm computes a sequence of $n + 1$ matrices, $M_0, M_1, ... M_n$. For each $k \in [0 ... n]$, $[M_k]_{ij} = 1$ iff there is a path in G from v_i to v_j whose interior nodes come only from $\{v_1, v_2, ... v_k\}$. Warshall's algorithm is

procedure: Warshall (var M:n x n matrix);

{initially M = A, the adjacency matrix of G}

    begin

        for $k := 1$ **to** n **do**

            for $i := 1$ **to** n **do**

                for $j := 1$ **to** n **do**

                    $M[i, j] := M[i, j] \lor (M[i, k] \land M[k, j])$;

    end;

Example: Find the reachability matrix of the digraph G using Warhall's algorithm.

Lets look at this in detail. The following table shows the derivation of the elements of M_1 from those of M_0. Notice that we use the updated elements of M as soon as they are available!

k	i	j		
1	1	1	$M_1[1,1] := M_0[1,1] \lor (M_0[1,1] \land M_0[1,1])$	$M_1[1,1] := 0 \lor (0 \land 0) = 0$
1	1	2	$M_1[1,2] := M_0[1,2] \lor (M_1[1,1] \land M_0[1,2])$	$M_1[1,2] := 1 \lor (0 \land 1) = 1$
1	1	3	$M_1[1,3] := M_0[1,3] \lor (M_1[1,1] \land M_0[1,3])$	$M_1[1,3] := 0 \lor (0 \land 0) = 0$
1	1	4	$M_1[1,4] := M_0[1,4] \lor (M_1[1,1] \land M_0[1,4])$	$M_1[1,4] := 1 \lor (0 \land 1) = 1$
1	2	1	$M_1[2,1] := M_0[2,1] \lor (M_0[2,1] \land M_1[1,1])$	$M_1[2,1] := 0 \lor (0 \land 0) = 0$
1	2	2	$M_1[2,2] := M_0[2,2] \lor (M_1[2,1] \land M_1[1,2])$	$M_1[2,2] := 1 \lor (0 \land 1) = 1$
1	2	3	$M_1[2,3] := M_0[2,3] \lor (M_1[2,1] \land M_1[1,3])$	$M_1[2,3] := 0 \lor (0 \land 0) = 0$
1	2	4	$M_1[2,4] := M_0[2,4] \lor (M_1[2,1] \land M_1[1,4])$	$M_1[2,4] := 1 \lor (0 \land 1) = 1$
1	3	1	$M_1[3,1] := M_0[3,1] \lor (M_0[3,1] \land M_1[1,1])$	$M_1[3,1] := 0 \lor (0 \land 0) = 0$
1	3	2	$M_1[3,2] := M_0[3,2] \lor (M_1[3,1] \land M_1[1,2])$	$M_1[3,2] := 1 \lor (0 \land 1) = 1$
1	3	3	$M_1[3,3] := M_0[3,3] \lor (M_1[3,1] \land M_1[1,3])$	$M_1[3,3] := 0 \lor (0 \land 0) = 0$
1	3	4	$M_1[3,4] := M_0[3,4] \lor (M_1[3,1] \land M_1[1,4])$	$M_1[3,4] := 0 \lor (0 \land 1) = 0$
1	4	1	$M_1[4,1] := M_0[4,1] \lor (M_0[4,1] \land M_1[1,1])$	$M_1[4,1] := 0 \lor (0 \land 0) = 0$
1	4	2	$M_1[4,2] := M_0[4,2] \lor (M_1[4,1] \land M_1[1,2])$	$M_1[4,2] := 0 \lor (0 \land 1) = 0$
1	4	3	$M_1[4,3] := M_0[4,3] \lor (M_1[4,1] \land M_1[1,3])$	$M_1[4,3] := 1 \lor (0 \land 0) = 1$
1	4	4	$M_1[4,4] := M_0[4,4] \lor (M_1[4,1] \land M_1[1,4])$	$M_1[4,4] := 0 \lor (0 \land 1) = 0$

$$M_0 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, M_1 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, M_2 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

$$M_3 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}, M_4 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

The major advantage of Warshall's algorithm is that it is computationally more efficient. The number of operations for a digraph with n vertices is $O(n^3)$ whilst the number of operations required to compute R from $R = A \cup A^2 \cup A^3 \cup ... \cup A^{n-1}$ is $O(n^4)$.

To see this proceed thus:

Let an *and* take m and an or s msec. For the power method, to compute each element of A^2 takes $n*m + (n - 1)*s$. A has n^2 elements to compute the whole of A^2 takes $n^2*(n*m + (n - 1)*s)$. To compute A^3 we can multiply A times A^2 and this just takes the same time as computing A^2. For a graph with n vertices there are $n - 2$ matrix multiplications and then an *or* of $n - 1$ matrices, so the total time is $(n - 1)*(n^2*(n*m + (n - 1)*s)) + n^2*(n - 2)*s = (n^4 - 2*n^3)*(m + s) = O(n^4)$.

For Warshall's algorithm, each basic operation takes one *and* and one *or*. The triple loop means there are n^3 basic operations so the total time taken is $n^3*(m + s) = O(n^3)$.

Overall therefore, for a small graph we can compute the reachability matrix by hand using the power method relatively quickly. But if we are looking at a larger graph (think of 50 vertices and then consider the situation for 500 or 5000 vertices), we need computational help and Warshall's algorithm will take $1/n$ of the compute time taken by the power method.

Case Study

Draw the digraphs corresponding to the adjacency matrices.

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$
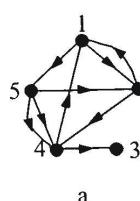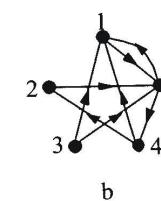
(a)   (b)   (c)

Ans.

a   b   c
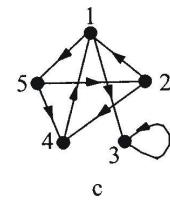
2.6 Summary

- The Königsberg bridge problem asks if the seven bridges of the city of Königsberg

- **Graph traversal** refers to the problem of visiting all the nodes in a graph in a particular manner.

- An **edge labeling** is a function mapping edges of *G* to a set of "labels". In this case, *G* is called an **edge-labeled graph**.

- When the edge labels are members of an ordered set (e.g., the real numbers), it may be called a **weighted graph**.

- A **complete graph** is a simple graph in which every pair of distinct vertices is connected by a unique edge.

2.7 Keywords
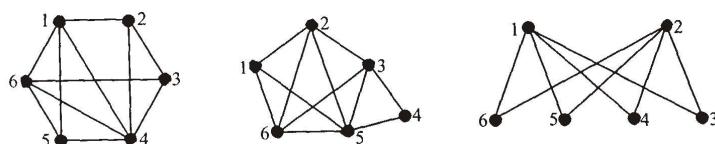
Bipartite Graph: A **bipartite graph** (or **bigraph**) is a graph whose vertices can be divided into two disjoint sets *U* and *V* such that every edge connects a vertex in *U* to one in *V*; that is, *U* and *V* are independent sets.

Planar Graph: A **planar graph** is a graph that can be embedded in the plane, i.e., it can be drawn on the plane in such a way that its edges intersect only at their endpoints.

Strongly Regular Graph: A strongly regular graph is a regular graph where every adjacent pair of vertices has the same number *l* of neighbors in common, and every non-adjacent pair of vertices has the same number *n* of neighbors in common.

2.8 Self Assessment

1. Find a planar drawing of each of the following graphs.
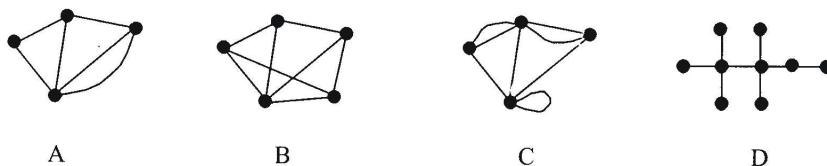
2. If G is a simple, connected, planar graph with $n(\geq 3)$ vertices and m edges, and if g is the length of the shortest cycle in G, show that

$m \leq g(n - 2)/(g - 2)$

Hint : The edges around a face in a plane drawing of a planar graph are a cycle. Find a lower bound on the face degree sum of G then use the Handshaking Lemma and Euler's formula.

3. Write down the degree sequence of each of the following graphs:

A   B   C   D
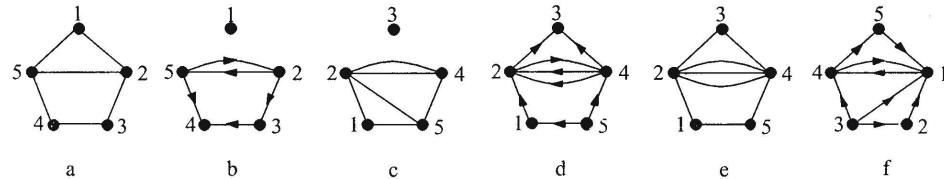
4. Draw an *r*-regular graph with 6 vertices for *r* = 3 and *r* = 4.

5. Why are there no 3-regular graphs with 5 vertices?

Fill in the blanks:

6. A Breadth-first Search (BFS) is another technique for ……………… a finite undirected graph.

7. In a ……………….. graph the vertices have labels or names.

8. A path is a walk I which all the edges and all the ……………… are distinct.

9. Every tree is a ……………… graph.

10. A graph G is planner if it can be drawn on a plane in such a way that is no two edge meet except at a ……………. with they both incident.

11. …………………… refers to the problem of visiting all the nodes in a graph in a particular manner.

12. A graph may be ……………… , meaning that there is no distinction between the two vertices associated with each edge.

13. A ………………… is a technique for traversing a finite undirected graph.

14. A "graph" in this context refers to a collection of vertices or 'nodes' and a collection of …………… that connect pairs of vertices.

2.9 Review Questions

1. Use Warshall's algorithm to determine if the digraph defined by the adjacency matrix in question 4 is strongly connected.

2. Write down the adjacency matrices of the following graphs and digraphs.

a   b   c   d   e   f

3. Draw the graphs corresponding to the following adjacency matrices.

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 2 & 0 & 1 & 1 \\ 2 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 2 \\ 1 & 1 & 0 & 2 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$
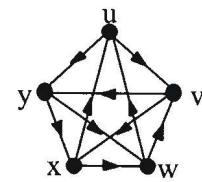
4. Determine the in-degree and the out-degree of each vertex in the following digraph and hence determine if it is Eulerian. Draw the digraph and determine an Eulerian trail.
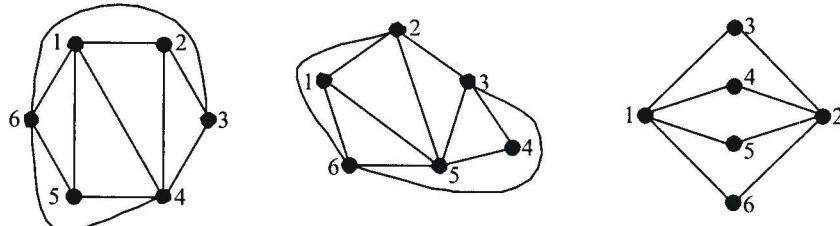
$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

5. Write down the adjacency matrices of the digraph shown, calculate the matrices A^2, A^3 and A^4, and hence find the number of walks of length 1, 2, 3 and 4 from w to u. Is there a walk of length 1, 2, 3 or 4 from u to w? Find the matrix B (= $A + A^2 + A^3 + A^4$) for the digraph and hence say whether it is strongly connected. Write down the Boolean matrix for the digraph and use Warshall's algorithm to find the reachability matrix.

6. Under what conditions on r and s is the complete bipartite graph $K_{r,s}$ a regular graph?

7. Show that, in a bipartite graph, every cycle has an even number of edges.

8. Draw the complete bipartite graphs $K_{2,3}$, $K_{3,5}$, $K_{4,4}$. How many edges and vertices does each graph have? How many edges and vertices would you expect in the complete bipartite graphs $K_{r,s}$.

Answers: Self Assessment
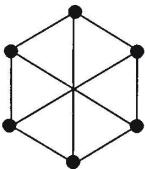
1. Possible planar drawing are:
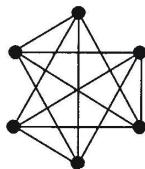
2. In a plane drawing of a planar graph the edges surrounding a face are a cycle. Thus, if g is the length of the shortest cycle in a planar graph, the degree of each face in the plane drawing is $\geq g$. Therefore, the sum of the face degrees is $\geq gf$. The Handshaking Lemma tells us that the sum of the face degrees is twice the number of edges = $2m$, so $2m \geq gf$. Now Euler's formula tells us that $n - m + f = 2$ so $f = m + 2 - n$. Hence $2m \geq gf = g(m + 2 - n)$. Hence, we have $2m \geq g(m + 2 - n)$. Hence $g(n - 2) \geq (g - 2)m$, i.e. $g(n - 2)/(g - 2) > m$.

3. Degree sequence of A is (2, 3, 3, 4).

   Degree sequence of B is (3, 3, 3, 3, 4).

   Degree sequence of C is (3, 3, 5, 5).

   Degree sequence of D is (1, 1, 1, 1, 1, 1, 2, 4, 4).

Notes

4.

3-regular graph   4-regular graph

5.   A 3-regular graph with 5 vertices would have a sum of vertex degree of 15, an odd number. But the sum of vertex degree is twice the number of edges and so is an even number. Hence there can be no 3-regular graph with 5 vertices.

6.   Traversing

7.   Labelled

8.   Vertices

9.   Bipartite

10.   Vertex

11.   Graph Traversal

12.   Undirected

13.   Depth-first Search (DFS)

14.   Edges

2.10 Further Readings

Books   Béla Bollobás, *Modern Graph Theory*, Springer

Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph Theory, Combinatorics, and Algorithms*, Birkhäuser

Online links   http://en.wikipedia.org/wiki/Bipartite_graph

http://en.wikipedia.org/wiki/Planar_graph

Unit 3: Eulerian and Hamiltonian Graphs

CONTENTS

Objectives

After studying this unit, you will be able to:
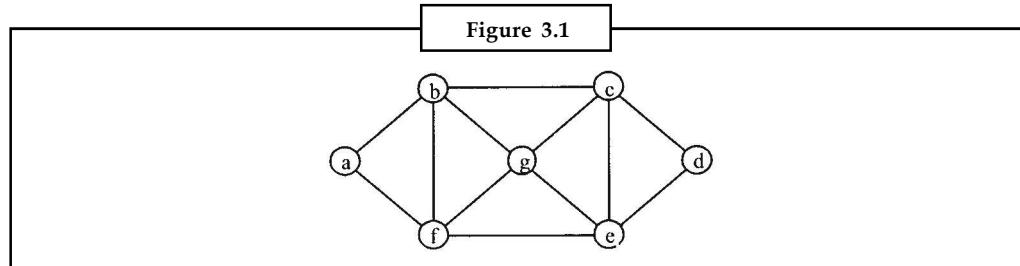
- Discuss Eulerian Graph
- Describe Hamiltonian Graph
- Explain Isomorphic Graph

Introduction

As you studied in last unit that graph is an abstract representation of a set of objects where some pairs of the objects are connected by links. The interconnected objects are represented by mathematical abstractions called vertices, and the links that connect some pairs of vertices are called edges. Graph theory began with Euler's study of a particular problem: the Seven Bridges of K¨onigsberg. This unit provides you clear understanding of eulerian and Hamiltonian graphs.

3.1 Eulerian and Hamiltonian Graphs

Consider the following map of 7 towns and the roads connecting them.

Figure 3.1

A highway engineer (E) wishes to inspect all the roads whilst an Egon Ronay inspector (H) wishes to dine in a restaurant in each town. Having studied Engineering Mathematics, each wishes to achieve their objective in as efficient a way as possible. So E states her aim as "I wish, if possible, to traverse every road once and only once and return to my starting point" whilst H says "I wish, if possible, to visit each town once and only once and return to my starting point".

A range of real problems give rise to versions of these two objectives, so graph theory has formalised them in the following way.

An **Eulerian graph** is a connected graph wishes contains a closed trail which includes every edge. The trail is called an **Eulerian trail**.

A **Hamiltonian graph** is a connected graph which contains a cycle which includes every vertex. The cycle is called an **Hamiltonian cycle**.

So E is saying "I want an Eulerian trail" and H is saying "I want a Hamiltonian cycle". Considering the map in Figure 3.1 as a graph, both an Eulerian trail and a Hamiltonian cycle exist, for instance *abcdecgefgbfa* and *abcdegfa* respectively. So the graph is both Eulerian and Hamiltonian.
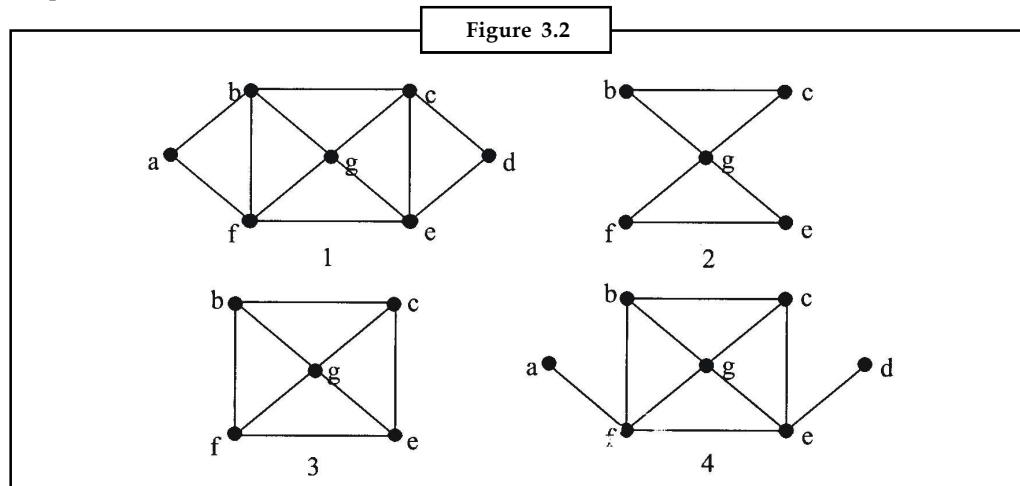
In Figure 3.2 we see some more examples of Eulerian and Hamiltonian graphs.

Graph 1 (the graph of the map in Figure 3.2) is both Eulerian and Hamiltonian.

Graph 2 is Eulerian (*e.g. bcgefgb*) but not Hamiltonian.

Graph 3 is Hamiltonian (*e.g. bcgefb*) but not Eulerian.

Graph 4 is neither Eulerian nor Hamiltonian.

Figure 3.2
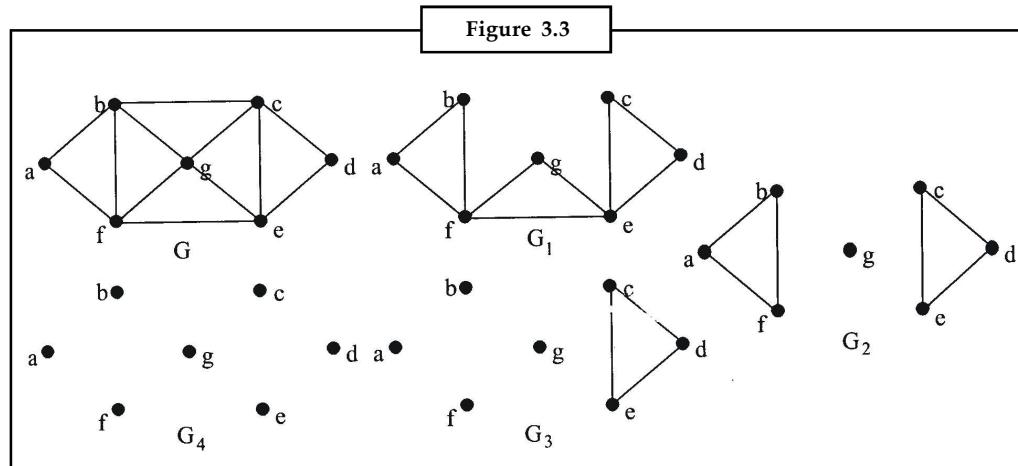
3.2 Eulerian Graphs

Theorem 1. A connected graph is Eulerian iff every vertex has even degree.

To prove this we first need a simpler theorem.

Theorem 2. If G is a graph all of whose vertices have even degree, then G can be split into cycles no two of which have an edge in common.

Proof. Let G be a graph all of whose vertices have even degree. Start at any vertex u and traverse edges in an arbitrary manner, without repeating any edge. Since every vertex has even degree it is always possible to find a different edge by which to leave a vertex. Since there is a finite number of vertices, eventually we must arrive at a vertex, v say, which we have already visited. The edges visited since the previous visit to v constitute a closed cycle, C_1 say. Remove all the edges in C_1 from the graph, leaving a subgraph G_1 say. Since we have removed a closed cycle of edges the vertices of G_1 will either have the same degree as the vertices of G or degrees 2 less than the equivalent vertices of G—either way G_1 is a graph all of whose vertices have been degree. We repeat the process with G_1, finding a cycle C_2, removing the edges in this cycle from G_1 and leaving G_2. Continue in this way until there are no edges left. Then we have a set of cycles, $C_1, C_2, C_3, \ldots$ which together include all edges of G and no two of which have an edge in common.

Figure 3.3

For instance, traverse *abcgb*—found cycle C_1 = *bcgb*—form G_1—traverse *degfe*—found cycle C_2 = *egfe*—form G_2—traverse *bafb*—found cycle C_3 = *bafb*—from G_3—traverse *edce*—found cycle C_4 = *edce*—from G_4—no edges left. The graph G can be split into cycles {*bcgb*, *egfe*, *bafb*, *edce*}. The decomposition is, of course, not unique. For instance we could equally decompose G into {*bcefb*, *gcdeg*, *abgfa*}.

Now we can prove **Theorem 1.** A connected graph is Eulerian iff every vertex has even degree.

Proof:

First we prove "If a graph G is Eulerian then each vertex of G has even degree." Since G is Eulerian there exists an Eulerian trail. Every time the trail passes through a vertex it traverses two different edges incident with the vertex and so adds two to the degree of the vertex. Since the trail is Eulerian it traverses every edge, so the total degree of each vertex is 2 or 4 or 6 or ...., i.e. of the form $2k$, $k = 1, 2, 3, \ldots$. Hence every has even degree.

Now we prove "If each vertex of a connected graph G has even degree then G is Eulerian." Since all the vertices of G have even degree, by theorem 2 G can be decomposed into a set of cycles no two of which have an edge in common. We will fit these cycles together to create an Eulerian trail. Start at any vertex of a cycle C_1. Travel round C_1 until we find a vertex which belongs also

to another cycle, C_2 say. Travel round C_2 and then continue along C_1 until we reach the starting point. We have closed trail C_{12} which includes all the edges of C_1 and C_2. If this includes all the edges of G we have the required Eulerian trail, otherwise repeat the process starting at any point of C_{12} and travelling around it until we come to a vertex which is a member of another cycle, C_3 say. Travel round C_3 and then continue along C_{12} thus creating a closed trail C_{123}. Continue the process until we have a trail which includes all the edges of G and that will be an Eulerian trail in G.

We have proved both "If a graph G is Eulerian then each vertex of G has even degree" and "If each vertex of a connected graph G has even degree then G is Eulerian" and so we have "A connected graph is Eulerian iff every vertex has even degree".

A **semi-Eulerian graph** is a connected graph which contains an open trail which includes every edge. The trail is called a semi-Eulerian trail.

Theorem 3: A connected graph is semi-Eulerian iff exactly two vertices have odd degree.

Proof: (*a*) If G is a semi-Eulerian graph then there is an open trail which includes every edge. Let u and v be the vertices at the start and end of this trail. Add the edge uv to the graph. The graph is now Eulerian and so every vertex has even degree by theorem 1. If the added edge is now removed the degrees of the vertices u and v edge are reduced by one and so are odd, the degrees of all other vertices are unaltered and are even. So if G is semi-Eulerian it has exactly two vertices of odd degree.

(*b*) Suppose G is a connected graph with exactly two vertices of odd degree. Let those two vertices be u and v. Add an edge uv to the graph. Now every vertex of G has even degree and so G is Eulerian. The Eulerian trail in G includes every edge and so includes the edge uv. Now remove the edge uv, then there is a trail starting at vertex u and ending at vertex v (or vice versa) which includes every edge. Hence if G is a connected graph with exactly two vertices of odd degree then G is semi Eulerian.

Hence we see that a connected graph is semi-Eulerian iff exactly two vertices have odd degree.

3.3 Hamiltonian Graphs

No simple necessary and sufficient condition for a graph to be Hamiltonian is known—this is an open area of research in graph theory.

But we can identify some classes of graph which are Hamiltonian. Obviously the cycle graph C_n is Hamiltonian for all n. The complete graph K_n is Hamiltonian for all $n \geq 3$.—obvious because, if the vertices are denoted $\{v_1, v_2, \dots v_n\}$ then the path $v_1v_2v_3 \dots v_nv_1$ is a Hamiltonian path.

If we add an edge to a Hamiltonian graph then the resulting graph is Hamiltonian—the Hamiltonian cycle in the original graph is also a Hamiltonian cycle in the enhanced graph. Adding edges may make a non-Hamiltonian graph into a Hamiltonian graph but cannot convert a Hamiltonian graph into a non-Hamiltonian one so graphs with high vertex degrees are more likely to be Hamiltonian then graphs with small vertex degrees. Ore's theorem is one possible more precise statement relating Hamiltonian graphs and their vertex degrees.

Ore's Theorem (stated without proof) : If G is a simple connected graph with n vertices ($n \geq 3$) then G is Hamiltonian if deg (v) + deg (w) $\geq n$ for every non-adjacent pair of vertices v and w.

If G is a simple connected graph with n vertices ($n \geq 3$) then G is Hamiltonian if deg (v) $\geq n/2$ for every vertex v. This follows from Ore's theorem. From this we can determine that all the complete bipartite graphs $K_{p,p}$ are Hamiltonian (the degree of every vertex is p, the graph has $2p$ vertices, hence deg (v) $\geq n/2$ ($=p$) for every vertex).

A **semi-Hamiltonian graph** is a connected graph which contains a path, but not a cycle, which includes every vertex. The path is called a **semi-Hamiltonian path**.
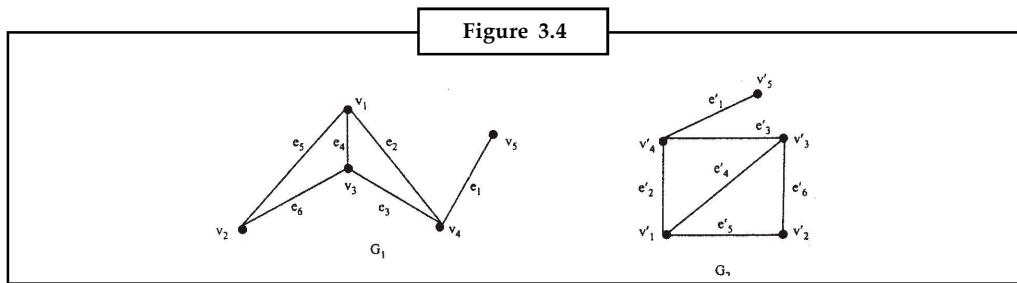
3.4 Isomorphism

Two graphs G_1 and G_2 are said to be isomorphic to each other if there is a one to one correspondence between their vertices and between their edges so that the incidence relationship is maintained.

It means that if in graph G_1 an edge e_k is incident with vertices v_1 and v_j then in graph G_2 its corresponding edge e'_k must be incident with the vertices v'_i and v'_j that correspondent to the vertices v_i and v_j respectively.
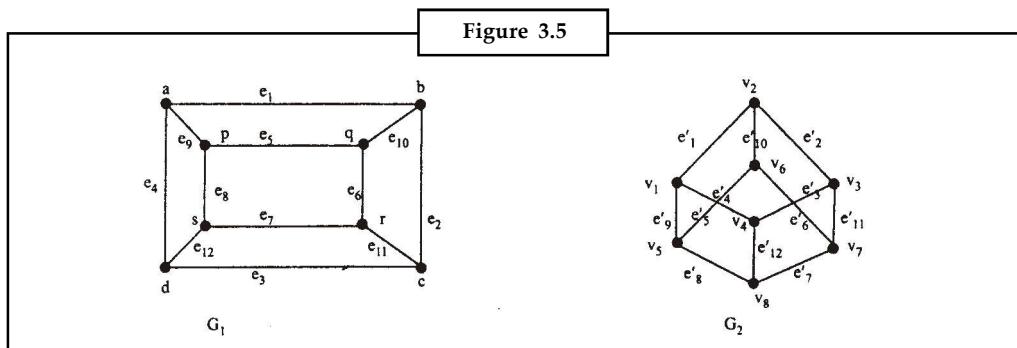
The following two graphs G_1 and G_2 are isomorphic graphs.

Figure 3.4

Vertices v_1, v_2, v_3, v_4 and v_5 in G_1 corresponds to v'_1 v'_2, v'_3, v'_4 and v'_5 respectively in G_2. Edges e_1, e_2, e_3, e_4, e_5 and e_6 in G_1 corresponds to e'_1, e'_2, e'_3, e'_4, e'_5 and e'_6 respectively in G_2.

Here we can see that if any edge is incident with two vertices in G_1 then its corresponding edge shall be incident with the corresponding vertices in G_2 e.g. edges e_1, e_2 and e_3 are incident on vertex v_4, then the corresponding edges e'_1, e'_2 and e'_3 shall be incident on the corresponding vertex v'_4. In the way the incidence relationship shall be preserved.

In fact isomorphic graphs are the same graphs drawn differently. The difference is in the names or labels of their vertices and edges. The following two graphs are also isomorphic graphs in which vertices a, b, c, d, p, q, r, and s in G_1 corresponds to vertices v_1, v_2, v_3, v_4, v_5, v_6, v_7 and v_8 respectively in G_2 and edges e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9, e_{10}, e_{11} and e_{12} in G_1 corresponds to edges e'_1, e'_2, e'_3, e'_4, e'_5, e^x_6, e'_7, e'_8, e'_9, e'_{10}, e'_{11} and e'_{12} in G_2 to preserve the incidence relationship.
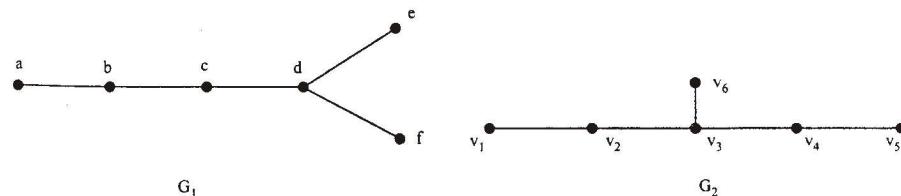
Figure 3.5

The incidence relationship between vertices and edges in between corresponding vertices and edges in G_2.

The following two graphs are not isomorphic

Figure 3.6

Vertex d in G_1 corresponds to vertex v_3 in G_2 as these are the only two vertices of degree 3.

In G_1, there are two pendant vertices adjacent to vertex d, while in G_2 there is only one pendant vertex adjacent to the corresponding vertex v_3. Thus the relationship of adjacency and incidence is not preserved and the two graphs are not isomorphic.

There is no simple and efficient criterion to identify isomorphic graphs.
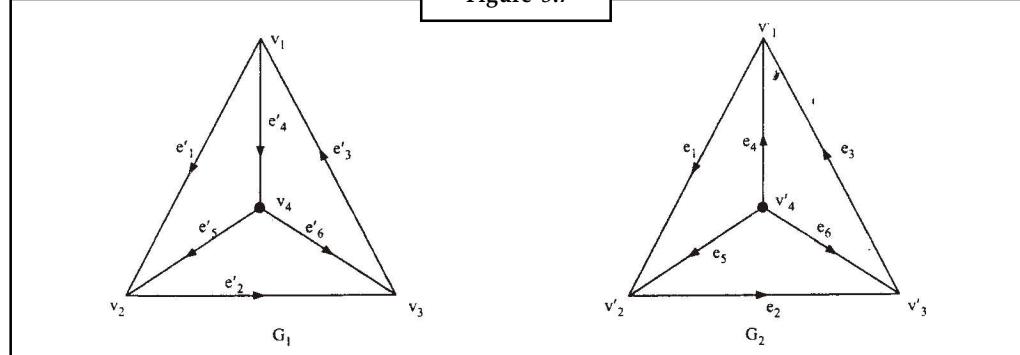
3.5 Isomorphic Digraphs

Two digraphs are said to be isomorphic if,

1.   Their corresponding undirected graphs are isomorphic.

2.   Directions of the corresponding edges also agree.

The following two digraphs are not isomorphic because the directions of the two corresponding edges e_4 and e'_4 do not agree (although their corresponding undirected graphs are isomorphic).

Figure 3.7

Isomorphism may also be defined as follows:

Isomorphism from a graph $G_1 = (V_1, E_1)$ to $G_2 = (V_2, E_2)$ is defined as mapping $f : V_1 \rightarrow V_2$ such that

(a) f is one–one and onto

(b) edge $v_i\, v_j \in E_1$ if and

only if $f(v_1) \cdot f(v_j) \in E_2$

where $f(v_1)$ and $f(v_j)$ are the images of v_i and v_j respectively in graph G_2

3.5.1 Some Properties of Isomorphic Graphs
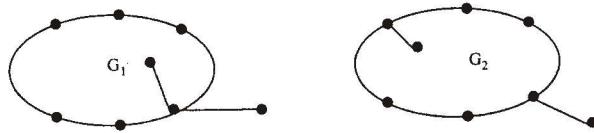
1.   Number of vertices in isomorphic graphs is the same.

2.   Number of edges in isomorphic graphs is also the same.

3.   Each one of the isomorphic graphs has an equal number of vertices with a given degree.

This property is utilized in identifying two non-isomorphic graphs by writing down the degree sequence of their respective vertices.

Example: The degree sequence of graph G_1 is 4, 2, 2, 2, 2, 2, 1, 1 and that of G_2 is 3, 3, 2, 2, 2, 2, 1, 1 which are not the same. Therefore G_1 and G_2 are non-isomorphic.

3.5.2 Some Observation on Isomorphism

Let G_1 be a graph with vertices v_0, v_1, v_2, ... v_n and f is an isomorphism from G_1 to G_2, then:

1. G_2 will be connected only if G, is connected, because a path from v_i to v_j in G, induces a path from f (v_i) to f (v_j) in G_2.

2.   If a Hamiltonian circuit exists in G_1, then a similar circuit shall exist in G_2, because if v_0, v_1, v_2, ..., v_n, v_0 is is a Hamiltonian circuit in G, then f (v_0), f(v_1), f(v_2), ..., , f(v_n), f(v_0) must be a Hamiltonian circuit in G_2 (as it is a circuit and f (v_i) ≠ f(v_j for 0 <= i < j <= n).

3.   Both G_1 and G_2 shall have the same number of isolated vertices, because if any vertex v_i is an isolated vertex in G_1 then its corresponding vertex f(v_i) shall also be isolated in G_2.

4.   If G_1 is bipartite, than G_2 shall also be bipartite.

5.   If G, has an Euler circuit, than G_2 shall also have an Euler circuit,

6.   Both G_1 and G_2 shall have the same crossing number because they can be drawn the same way in the plane.

Task   Analyse the properties which makes the two graphs isomorphic.

3.6 Summary

●   A range of real problems give rise to versions of these two objectives, so graph theory has formalised them in the following way.

   ❖   An **Eulerian graph** is a connected graph wishes contains a closed trail which includes every edge. The trail is called an **Eulerian trail**.

   ❖   A **Hamiltonian graph** is a connected graph which contains a cycle which includes every vertex. The cycle is called an **Hamiltonian cycle**.

●   A **closed walk/trail** is a walk/tail starting and ending at the same vertex.

●   A **cycle** is a closed path, i.e. a path starting and ending at the some vertex.

3.7 Keywords
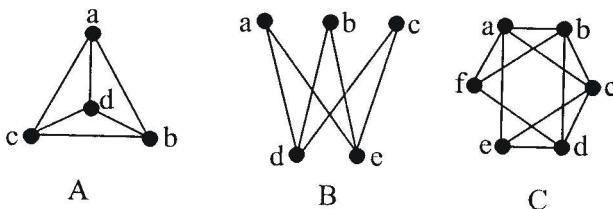
Eulerian graph: It is a connected graph wishes contains a closed trial which includes every edge.

Eulerian trial: Eulerian graph wishes contains a closed trial i.e., called eulerian trial.
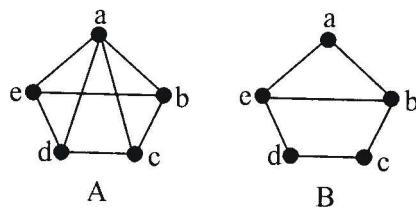
Hamiltonian graph: It is a connected graph which contains a cycle which includes every vertex.

3.8 Self Assessment

1. Which of the graphs K_8, $K_{4,4}$, C_6, $K_{2,5}$ are Eulerian graphs (use theorem 1 to decide). For those which are Eulerian, find an Eulerian trail.

2. By finding a Hamiltonian cycle show that the complete bipartite graph $K_{3,3}$ is Hamiltonian. Show that the complete bipartite graph $K_{2,4}$ is not Hamiltonian. What condition on *r* and *s* is necessary for the complete bipartite graph $K_{r,s}$ to be Hamiltonian?

3. Which of the following graphs are semi-Hamiltonian? Give a semi-Hamiltonian path where possible.

A   B   C
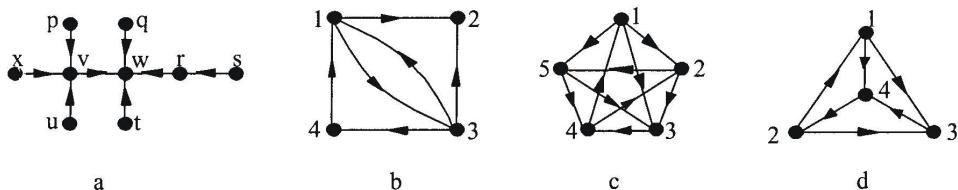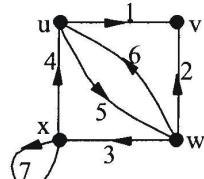
4. Check whether the conditions of Ore's theorem hold for these Hamiltonian graphs.

A   B
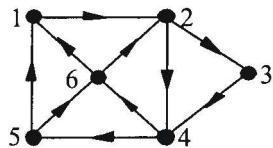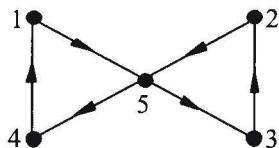
3.9 Review Questions
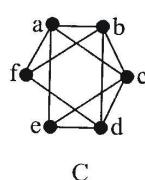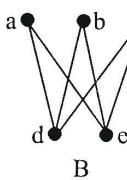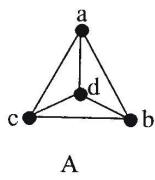
1. Discuss Eulerian graph with example.

2. In the Eulerian graph shown give (if possible)

   (*a*) a walk of length 7 from *u* to *w*,

   (*b*) cycles of length 1, 3 and 4, and

   (*c*) a path of maximum length

3. Which of the following connected Homaltonian graph are strongly connected?

a   b   c   d

4. Are the following digraphs Eulerian/semi-Eulerian and/or Hamiltonian/semi-Hamiltonian? Give the Eulerian/semi-Eulerian trails and the Hamiltonian cycles/semi-Hamiltonian paths where they exist.

5. Which of the following graphs are semi-Hamiltonian? Give a semi-Hamiltonian path where possible.

A   B   C
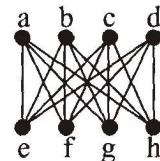
Answers: Self Assessment

1. K_8 is 7-regular, so all its vertices are of odd degree and it is not Eulerian.

    $K_{4,4}$ is 4-regular, so all its vertices are of even degree and it is Eulerian. An Eulerian trail, referred to the diagram below, is *aebfcgdhcedfagbha*.
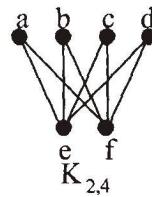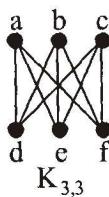
    C_6 is 2-regular, so all its vertices are of even degree and it is Eulerian. Since it's a cycle graph the whole graph constitutes an Eulerian cycle.

    $K_{2,5}$ has 5 vertices of degree 2 and 2 vertices of degree 5. Not all its vertices are of even degree so it is not Eulerian.

2. A Hamiltonian cycle in $K_{3,3}$ is *adbecfa* (see figure below) so $K_{3,3}$ is Hamiltonian.

    To construct a Hamiltonian cycle is $K_{2,4}$ we need to visit vertices from the set B = {a, b, c, d} and the set A = {e, f} alternately. Start at any vertex in B, go to a vertex in A then to a different vertex in B then to the other vertex in A then to another vertex in B. Now in order to visit the remaining vertices in B and to get back to the starting vertex we must visit another vertex in A. But we have visited both of those already so we cannot return to the start without re-visit a vertex which is already in the walk. Thus no Hamiltonian cycle exists and the graph is not Hamiltonian.

    A generalisation of this argument demonstrates that the only complete bipartite graphs, $K_{r,s}$, which are Hamiltonian are those with $r = s$.

3. (A) This graph is, in fact, Hamiltonian so it is not semi-Hamiltonian.

    (B) This graph is not Hamiltonian but *adbec* is a semi-Hamiltonian path so it is semi-Hamiltonian.

(C) This graph is, in fact, Hamiltonian (*fabcedf* is a Hamiltonian cycle) so it is not semi-Hamiltonian.

4. (A) The non-adjacent vertex pairs are (*c*, *e*) and (*b*, *d*). We have *n* = 5, deg (*c*) + deg (*e*) = 6 and deg (*b*) + deg (*d*) = 6, so the conditions of Ore's theorem hold.

    (B) The non-adjacent vertex pairs are (*a*, *d*), (*a*, *c*), (*c*, *e*) and (*b*, *d*). We have *n* = 5, deg (*a*) + deg (*d*) = 4 = deg (*a*) + deg (*c*) and deg (*c*) + deg (*e*) = 5 = deg (*b*) + deg (*d*), so the conditions of Ore's theorem do not hold.

3.10 Further Readings

Books   Béla Bollobás, *Modern Graph Theory*, Springer

Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph-theory, Combinatorics, and Algorithms*, Birkhäuser

Online links   http://en.wikipedia.org/wiki/Multigraph

http://en.wikipedia.org/wiki/Diagraph

Unit 4: Graphs Colouring

Objectives

After studying this unit, you will be able to:

- Define vertex colouring

- Explain the properties of colouring of a graph

- Describe the applications of colouring

Introduction

In graph theory, **graph colouring** is a special case of graph labeling; it is an assignment of labels traditionally called "colours" to elements of a graph subject to certain constraints. In its simplest form, it is a way of colouring the vertices of a graph such that no two adjacent vertices share the same colour; this is called a **vertex colouring.** Similarly, an **edge colouring** assigns a colour to each edge so that no two adjacent edges share the same colour, and a **face colouring** of a planar graph assigns a colour to each face or region so that no two faces that share a boundary have the same colour.
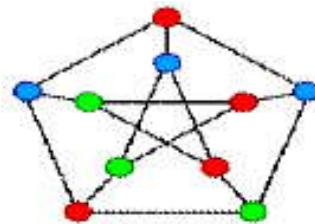
Vertex colouring is the starting point of the subject, and other colouring problems can be transformed into a vertex version. For example, an edge colouring of a graph is just a vertex colouring of its line graph, and a face colouring of a planar graph is just a vertex colouring of its planar dual. However, non-vertex colouring problems are often stated and studied *as is*. That is partly for perspective, and partly because some problems are best studied in non-vertex form, as for instance is edge colouring.

The convention of using colours originates from colouring the countries of a map, where each face is literally coloured. This was generalized to colouring the faces of a graph embedded in the

plane. By planar duality it became colouring the vertices, and in this form it generalizes to all graphs. In mathematical and computer representations it is typical to use the first few positive or nonnegative integers as the "colours". In general one can use any finite set as the "colour set". The nature of the colouring problem depends on the number of colours but not on what they are.

Figure 4.1: A proper vetex colouring of the petersen graph with 3 colours, the minimum possible
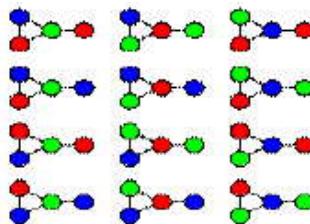
Graph colouring enjoys many practical applications as well as theoretical challenges. Beside the classical types of problems, different limitations can also be set on the graph, or on the way a colour is assigned, or even on the colour itself.

Notes   It has even reached popularity with the general public in the form of the popular number puzzle Sudoku. Graph colouring is still a very active field of research.

Definition and Terminology

Figure 4.2: This graph can be 3-coloured in 12 different ways

4.1 Vertex Colouring

When used without any qualification, a **colouring** of a graph is almost always a proper vertex colouring, namely a labelling of the graph's vertices with colours such that no two vertices sharing the same edge have the same colour. Since a vertex with a loop could never be properly coloured, it is understood that graphs in this context are loopless.

The terminology of using *colours* for vertex labels goes back to map colouring. Labels like *red* and *blue* are only used when the number of colours is small, and normally it is understood that the labels are drawn from the integers {1, 2, 3, ...}

A colouring using at most k colours is called a (proper) **k-colouring.** The smallest number of colours needed to colour a graph G is called its **chromatic number,** χ(G). A graph that can be assigned a (proper) k-colouring is **k-colourable,** and it is **k-chromatic** if its chromatic number is exactly *k*. A subset of vertices assigned to the same colour is called a *colour class*, every such class forms an independent set.
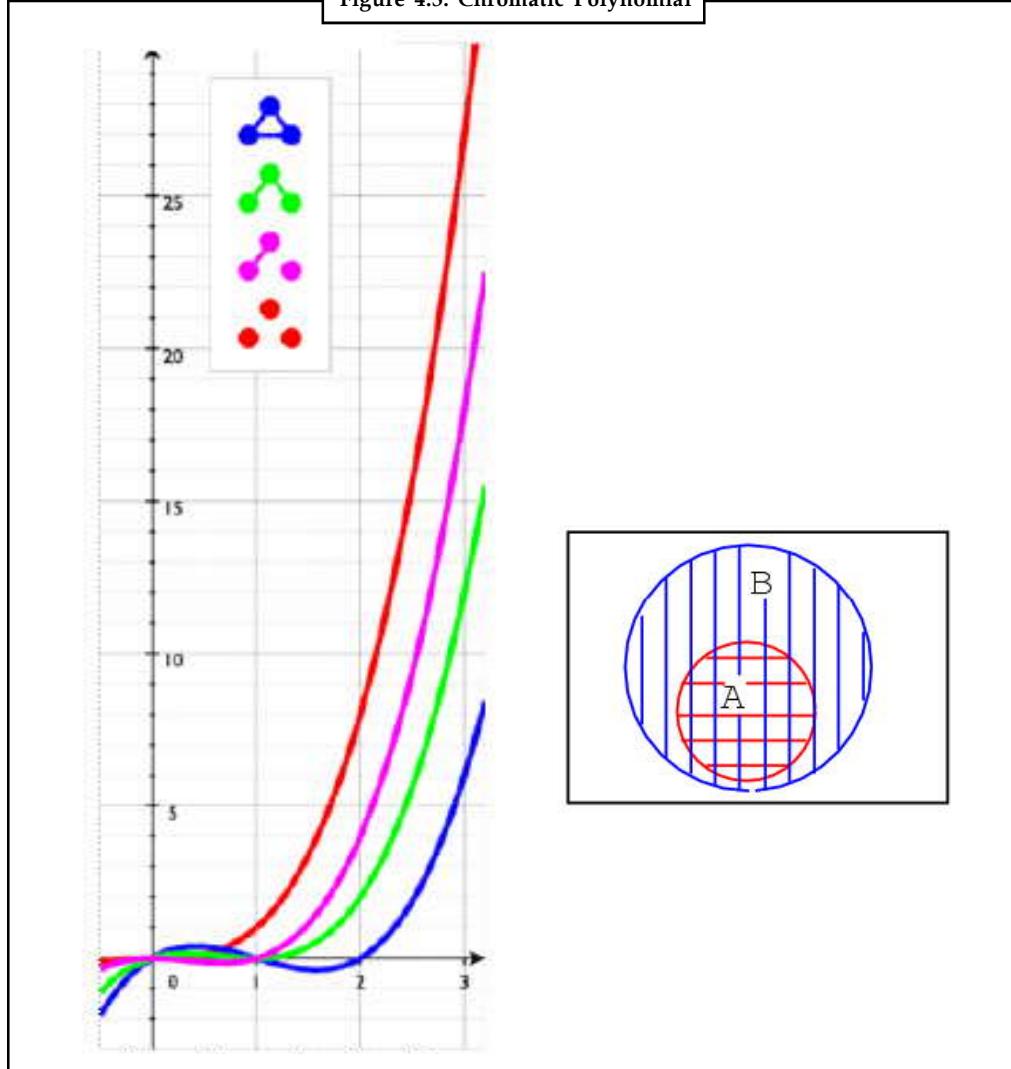
Did u know? **The k-partite and k-colourable have the same meaning?**

A *k-colouring* is the same as a partition of the vertex set into *k* independent sets, and the terms *k-partite* and *k-colourable* have the same meaning.

4.2 Chromatic Number

In the below given figure all non-isomorphic graphs on 3 vertices and their chromatic polynomials. The empty graph E_3 (red) admits a 1-colouring, the others admit no such colourings. The green graph admits 12 colourings with 3 colours.

Figure 4.3: Chromatic Polynomial

The **chromatic polynomial** counts the number of ways a graph can be coloured using no more than a given number of colours. For example, using three colours, the graph in the image to the right can be coloured in 12 ways. With only two colours, it cannot be coloured at all. With four colours, it can be coloured in 24 + 4Å"12 = 72 ways: using all four colours, there are 4! = 24 valid colourings (*every* assignment of four colours to *any* 4-vertex graph is a proper colouring); and for

every choice of three of the four colours, there are 12 valid 3-colourings. So, for the graph in the example, a table of the number of valid colourings would start like this:

Available colours	1	2	3	4	…
Number of colourings	0	0	12	72	…

The chromatic polynomial is a function P(G, t) that counts the number of *t*-colourings of G. As the name indicates, for a given G the function is indeed a polynomial in *t*. For the example graph, $P(G, t) = t(t-1)^2(t-2)$, and indeed P(G, 4) = 72.

The chromatic polynomial includes at least as much information about the colourability of G as does the chromatic number. Indeed, χ is the smallest positive integer that is not a root of the chromatic polynomial

$$\chi(G) \;=\; \min\{k : P(G,k) > 0\}.$$

Chromatic polynomials for certain graphs

Triangle K_3 $t(t-1)(t-2)$

Complete graph K_n

Tree with *n* vertices $t(t-1)^{n-1}$

Cycle C_n $(t-1)^n + (-1)^n(t-1)$

Petersen graph $t(t-1)(t-2)(t^7 - 12t^6 + 67t^5 - 230t^4 + 529t^3 - 814t^2 + 775t - 352)$

Edge colouring

An **edge colouring** of a graph, is a proper colouring of the *edges*, meaning an assignment of colours to edges so that no vertex is incident to two edges of the same colour. An edge colouring with *k* colours is called a *k*-edge-colouring and is equivalent to the problem of partitioning the edge set into *k* matchings. The smallest number of colours needed for an edge colouring of a graph *G* is the **chromatic index,** or **edge chromatic number,** χ^2 (G). A **Tait colouring** is a 3-edge colouring of a cubic graph. The four colour theorem is equivalent to the assertion that every planar cubic bridgeless graph admits a Tait colouring.

Task   Analyse what does the chromatic number do?

4.3 Properties

Bounds on the Chromatic Number

Assigning distinct colours to distinct vertices always yields a proper colouring, so

$$1 \le \chi(G) \le n.$$

The only graphs that can be 1-coloured are edgeless graphs, and the complete graph K_n of *n* vertices requires $\chi(K_n) = n$ colours. In an optimal colouring there must be at least one of the graph's *m* edges between every pair of colour classes, so

$$\chi(G)\left(\chi(G)-1\right) \le 2m.$$

If *G* contains a clique of size *k*, then at least *k* colours are needed to colour that clique; in other words, the chromatic number is at least the clique number:

$$\chi(G) \ge \omega(G).$$

For interval graphs this bound is tight.

The 2-colourable graphs are exactly the bipartite graphs, including trees and forests. By the four colour theorem, every planar graph can be 4-coloured.

A greedy colouring shows that every graph can be coloured with one more colour than the maximum vertex degree,

$$\chi(G) \leq \Delta(G) + 1.$$

Complete graphs have $\div(G) = n$ and $\ddot{A}(G) = n$ " 1, and odd cycles have $\div(G) = 3$ and $\ddot{A}(G) = 2$, so for these graphs this bound is best possible. In all other cases, the bound can be slightly improved; Brooks' theorem states that

Brooks' theorem: $\chi(G) \leq \Delta(G)$ for a connected, simple graph G, unless G is a complete graph or an odd cycle.

Graphs with High Chromatic Number

Graphs with large cliques have high chromatic number, but the opposite is not true. The Grötzsch graph is an example of a 4-chromatic graph without a triangle, and the example can be generalised to the Mycielskians.

Mycielski's Theorem: There exist triangle-free graphs with arbitrarily high chromatic number.

From Brooks's theorem, graphs with high chromatic number must have high maximum degree. Another local property that leads to high chromatic number is the presence of a large clique. But colourability is not an entirely local phenomenon: A graph with high girth looks locally like a tree, because all cycles are long, but its chromatic number need not be 2:

Theorem (Erdõs): There exist graphs of arbitrarily high girth and chromatic number.

Bounds on the Chromatic Index

An edge colouring of G is a vertex colouring of its line graph $L(G)$, and vice-versa. Thus,

$$\chi(G) - \chi'\big(L(G)\big).$$

There is a strong relationship between edge colourability and the graph's maximum degree $\ddot{A}(G)$. Since all edges incident to the same vertex need their own colour, we have

$$\chi'(G) \geq \Delta(G)$$

Moreover,

König's theorem: $\chi'(G) = \Delta(G)$ if G is bipartite.

In general, the relationship is even stronger than what Brooks's theorem gives for vertex colouring:

Vizing's Theorem: A graph of maximal degree Δ has edge-chromatic number Δ or $\Delta + 1$.

Other Properties

For planar graphs, vertex colourings are essentially dual to nowhere-zero flows.

About infinite graphs, much less is known. The following is one of the few results about infinite graph colouring:

If all finite subgraphs of an infinite graph G are k-colourable, then so is G, under the assumption of the axiom of choice.

Open Problems

The chromatic number of the plane, where two points are adjacent if they have unit distance, is unknown, although it is one of 4, 5, 6, or 7. Other open problems concerning the chromatic number of graphs include the Hadwiger conjecture stating that every graph with chromatic number k has a complete graph on k vertices as a minor, the Erdõs–Faber–Lovász conjecture bounding the chromatic number of unions of complete graphs that have at exactly one vertex in common to each pair, and the Albertson conjecture that among k-chromatic graphs the complete graphs are the ones with smallest crossing number.

When Birkhoff and Lewis introduced the chromatic polynomial in their attack on the four-colour theorem, they conjectured that for planar graphs G, the polymomial $P(G,t)$ has no zeros in the region [4, ∞]. Although it is known that such a chromatic polynomial has no zeros in the region [5, ∞] and that P(G, 4) ≠ 0, their conjecture is still unresolved.

Caution It also remains an unsolved problem to characterize graphs which have the same chromatic polynomial and to determine which polynomials are chromatic.

4.4 Algorithms

Efficient Algorithms

Determining if a graph can be coloured with 2 colours is equivalent to determining whether or not the graph is bipartite, and thus computable in linear time using breadth-first search. More generally, the chromatic number and a corresponding colouring of perfect graphs can be computed in polynomial time using semidefinite programming. Closed formulas for chromatic polynomial are known for many classes of graphs, such as forest, chordal graphs, cycles, wheels, and ladders, so these can be evaluated in polynomial time.

Brute-force Search

Brute-force search for a k-colouring considers every of the k^n assignments of k colours to n vertices and checks for each if it is legal. To compute the chromatic number and the chromatic polynomial, this procedure is used for every $k = 1, ..., n – 1$, impractical for all but the smallest input graphs.

Contraction

The contraction G/uv of graph G is the graph obtained by identifying the vertices u and v, removing any edges between them, and replacing them with a single vertex w where any edges that were incident on u or v are redirected to w. This operation plays a major role in the analysis of graph colouring.

The chromatic number satisfies the recurrence relation:

$$\chi(G) \;=\; \min\{\chi(G + uv), \chi(G / uv)\}$$

due to Zykov, where u and v are nonadjacent vertices, $G + uv$ is the graph with the edge uv added. Several algorithms are based on evaluating this recurrence, the resulting computation tree is sometimes called a Zykov tree. The running time is based on the heuristic for choosing the vertices u and v.

The chromatic polynomial satisfies following recurrence relation

$$P(G - uv,k) = P(G / uv,k) + P(G,k)$$

where *u* and *v* are adjacent vertices and *G − uv* is the graph with the edge *uv* removed. *P(G − uv,k)* represents the number of possible proper colourings of the graph, when the vertices may have same or different colours. The number of proper colourings therefore come from the sum of two graphs. If the vertices *u* and *v* have different colours, then we can as well consider a graph, where *u* and *v* are adjacent.

If *u* and *v* have the same colours, we may as well consider a graph, where *u* and *v* are contracted. Tutte's curiosity about which other graph properties satisfied this recurrence led him to discover a bivariate generalization of the chromatic polynomial, the Tutte polynomial.

The expressions give rise to a recursive procedure, called the *deletion–contraction algorithm*, which forms the basis of many algorithms for graph colouring. The running time satisfies the same recurrence relation as the Fibonacci numbers, so in the worst case, the algorithm runs in time

within a polynomial factor of $\left((1+\sqrt{5})/2\right)^{n+m} = 0(1.6180)^{mm}$. The analysis can be improved to

within a polynomial factor of the number *t(G)* of spanning trees of the input graph.
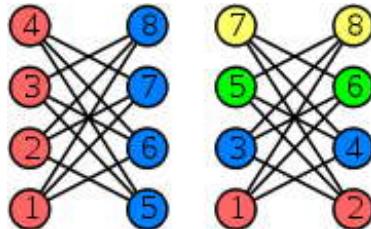
> *Task*   In practice, branch and bound strategies and graph isomorphism rejection are employed to avoid some recursive calls, the running time depends on the heuristic used to pick the vertex pair. Comment

Greedy Colouring

Figure 4.4

In the above figure two greedy colourings of the same graph using different vertex orders. The right example generalises to 2-colourable graphs with *n* vertices, where the greedy algorithm expends *n / 2* colours.

The greedy algorithm considers the vertices in a specific order $v_1,…,v_n$ and assigns to v_i the smallest available colour not used by v_i's neighbours among $v_1,…,v_{i-1}$, adding a fresh colour if needed. The quality of the resulting colouring depends on the chosen ordering. There exists an ordering that leads to a greedy colouring with the optimal number of ÷(G) colours. On the other hand, greedy colourings can be arbitrarily bad; for example, the crown graph on *n* vertices can be 2-coloured, but has an ordering that leads to a greedy colouring with *n / 2* colours.

If the vertices are ordered according to their degrees, the resulting greedy colouring uses at most $\max_i \min\{d(x_i) + 1,i\}$ colours, at most one more than the graph's maximum degree. This heuristic is sometimes called the Welsh–Powell algorithm. Another heuristic establishes the ordering dynamically while the algorithm proceeds, choosing next the vertex adjacent to the largest number of different colours.

Did u know? **What is sequential colouring?**

Many other graph colouring heuristics are similarly based on greedy colouring for a specific static or dynamic strategy of ordering the vertices, these algorithms are sometimes called **sequential colouring** algorithms.

Computational Complexity

Graph colouring is computationally hard. It is NP-complete to decide if a given graph admits a *k*-colouring for a given *k* except for the cases *k* = 1 and *k* = 2. Especially, it is NP-hard to compute the chromatic number. Graph colouring remains NP-complete even on planar graphs of degree 4.

The best known approximation algorithm computes a colouring of size at most within a factor $O(n(\log n)''^3(\log \log n)^2)$ of the chromatic number. For all $\varepsilon > 0$, approximating the chromatic number within $n^{1''\varepsilon}$ is NP-hard.

It is also NP-hard to colour a 3-colourable graph with 4 colours and a *k*-colourable graph with $k^{(\log k)} / 25$ colours for sufficiently large constant *k*.

Computing the coefficients of the chromatic polynomial is #P-hard. In fact, even computing the value of $\chi(G,k)$ is #P-hard at any rational point *k* except for *k* = 1 and *k* = 2. There is no FPRAS for evaluating the chromatic polynomial at any rational point $k \geq 1.5$ except for *k* = 2 unless NP = RP.

For edge colouring, the proof of Vizing's result gives an algorithm that uses at most $\Delta+1$ colours. However, deciding between the two candidate values for the edge chromatic number is NP-complete. In terms of approximation algorithms, Vizing's algorithm shows that the edge chromatic number can be approximated within 4/3, and the hardness result shows that no (4/3-ε)-algorithm exists for any $\varepsilon > 0$ unless P = NP. These are among the oldest results in the literature of approximation algorithms, even though neither paper makes explicit use of that notion.

Parallel and Distributed Algorithms

In the field of distributed algorithms, graph colouring is closely related to the problem of *symmetry breaking*. In a symmetric graph, a deterministic distributed algorithm cannot find a proper vertex colouring. Some auxiliary information is needed in order to break symmetry. A standard assumption is that initially each node has a *unique identifier*, for example, from the set {1, 2, ..., *n*} where *n* is the number of nodes in the graph. Put otherwise, we assume that we are given an *n*-colouring. The challenge is to *reduce* the number of colours from *n* to, e.g., $\Delta + 1$.

A straightforward distributed version of the greedy algorithm for $(\Delta + 1)$-colouring requires $\theta(n)$ communication rounds in the worst case – information may need to be propagated from one side of the network to another side. However, much faster algorithms exist, at least if the maximum degree Δ is small.

The simplest interesting case is an *n*-cycle. Richard Cole and Uzi Vishkin show that there is a distributed algorithm that reduces the number of colours from *n* to $O(\log n)$ in one synchronous communication step. By iterating the same procedure, it is possible to obtain a 3-colouring of an *n*-cycle in $O(\log^* n)$ communication steps (assuming that we have unique node identifiers).

The function log*, iterated logarithm, is an extremely slowly growing function, "almost constant". Hence the result by Cole and Vishkin raised the question of whether there is a *constant-time* distribute algorithm for 3-colouring an *n*-cycle. Linial showed that this is not possible: any

deterministic distributed algorithm requires $\Omega(\log^* n)$ communication steps to reduce an *n*-colouring to a 3-colouring in an *n*-cycle.

The technique by Cole and Vishkin can be applied in arbitrary bounded-degree graphs as well; the running time is poly(Δ) + O(log* *n*). The current fastest known algorithm for (Δ + 1)-colouring is due to Leonid Barenboim and Michael Elkin, which runs in time $O(\ddot{A}) + \log^*(n)/2$, which is optimal in terms of *n* since the constant factor 1/2 cannot be improved due to Linial's lower bound.

⚠️

Caution The problem of edge colouring has also been studied in the distributed model. We can achieve a (2Δ – 1)-colouring in O($\ddot{A}$+log* *n*) time in this model. Linial's lower bound for distributed vertex colouring applies to the distributed edge colouring problem as well.

4.5 Applications

Scheduling

Vertex colouring models to a number of scheduling problems. In the cleanest form, a given set of jobs need to be assigned to time slots, each job requires one such slot. Jobs can be scheduled in any order, but pairs of jobs may be in *conflict* in the sense that they may not be assigned to the same time slot, for example because they both rely on a shared resource. The corresponding graph contains a vertex for every job and an edge for every conflicting pair of jobs. The chromatic number of the graph is exactly the minimum *makespan*, the optimal time to finish all jobs without conflicts.

Details of the scheduling problem define the structure of the graph. For example, when assigning aircrafts to flights, the resulting conflict graph is an interval graph, so the colouring problem can be solved efficiently. In bandwidth allocation to radio stations, the resulting conflict graph is a unit disk graph, so the colouring problem is 3-approximable.
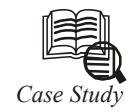
Register Allocation

A compiler is a computer program that translates one computer language into another. To improve the execution time of the resulting code, one of the techniques of compiler optimization is register allocation, where the most frequently used values of the compiled program are kept in the fast processor registers. Ideally, values are assigned to registers so that they can all reside in the registers when they are used.

The textbook approach to this problem is to model it as a graph colouring problem. The compiler constructs an *interference graph*, where vertices are symbolic registers and an edge connects two nodes if they are needed at the same time. If the graph can be coloured with *k* colours then the variables can be stored in *k* registers.

Other Applications

The problem of colouring a graph has found a number of applications, including pattern matching.

The recreational puzzle Sudoku can be seen as completing a 9-colouring on given specific graph with 81 vertices.

Case Study

M any of you are surely familiar with the scene. It is the end of the quarter, and there is a day-long meeting to review the performance and forecast the future. Preparations have an air similar to that of schoolboys brushing up before exam, except this time it is an open book exam – and the issue is whether one can face all the unexpected further questions that are sparked off by the answers to the regular ones.

Almost inevitably, the presentation starts with explanations, however weak, contrived or just plain irrelevant, for failure to reach the sales and profit numbers. If difficulties in the market place, or tougher competition for some of the company's brands, are trotted out as the reason, some one mutters: "well that's what we have you there for!" or even more pointedly, "if all we had to sell were products that had a ready and willing customer, we wouldn't need so many highly-paid sales people or a whole department, would we?".

This time the barb goes home. The failure has been personalised, to be soon converted into a failure of personality and character or simply incompetence. And the blame has been laid at the door of the senior-most marketing manager or the person in the room most directly responsible for sales results.

Board rage is the term I suggest (on the analogy of road rage) for the flare-up of tempers across the board room table, which reduces some meetings to monologues and tirades and others to a mutual trading of charges or a blame game.

Sometimes, particularly if the top management at the Director level is not represented in the room, someone makes a weak attempt at collegial humour, and tries to laugh the whole thing off. If not, there are strained silences, and shifting glances, as the audience gets ready for a proper dressing down. On the rare occasion when a junior person who is set upon by a few decides to defend himself, the conversation, if it can be called that, becomes more strident. Believe me; I have personal experience, in my salad days, of telling someone several levels above me in the organisation that he was being unfair in his judgement of my behaviour. "Well, if you don't like it, you can lump it!" was the answer given in a rising and angry tone that actually said – how dare you even open your mouth!

Enough said, I think. One does not want to rake up painful memories; almost every reader who has endured the rigours of organisational life, would have personal evidence of such embarrassing scenes. And yet, in the same companies no doubt there are people who are most concerned about the health of the people who arrange classes at company expense on stress management, anger management, meditation, team-building off-site days, training in EQ and what have you. Life is both rough and smooth, would be their explanations.

Yet, in my view, this two-pronged and paradoxical action is avoidable. Much of the stress comes from such board rage, which in turn comes from an excessive emphasis on pinning down accountability and an equal anxiety to shake it off and preferably pass on the blame to someone else.

The reward-punishment mechanism is still based on a crude behaviourist philosophy of human behaviour, even after all the modernisation of human resources management in theory. Thus, sales blames marketing, both blame the manufacturing plant and all of them roundly curse the supplier who did not supply some part on time – as if that particular shoe-nail was the reason the battle was lost. And so the game goes on like a round robin

with the one who sometimes has a major share (the topmost person who took the vital decisions on pricing or similar issues) going unmentioned — by tacit, and common, consent.

Why does anger play such a part in organisational life? To answer this, one must explore the real force behind anger – usually it is fear in some form: fear of consequences, inability to face the loss of face, fearing the loss of self-image or worse still, how one would appear to the other members of the Board, the CEO, or the shareholders. The only change is in the level of the meeting and the occasion – regular internal review, Board meeting or the Company's General Meeting — but the sentiment is the same. Everyone wants to get away unscathed, reputation more or less intact, and sometimes, in the case of the junior level sales force, or other functionaries, their jobs intact.

One wonders, however, whether such anger does anyone any good. Perhaps the person who lets off steam and is at the top of the heap goes home feeling relieved. Yet, even that is doubtful. There is usually much mute gnashing of teeth, sulking, and extreme defensiveness. All of these tend to be counter-productive and seem to motivate no one. Yet, getting angry, flying into a rage or showing one's displeasure in public is seen almost as a symbol of corporate manliness.

The worship of success and expectation of continued high hit rates and the boost given to the ones who have no negative variances against their name in the review all lead to totally unrealistic hopes. With the inevitable come down when the graph turns southwards, as it sometimes must.

Organisational life must take the damages wrought by board rage into account before it becomes too rooted in the character of the company. Often, the tone is set by the head of the company and every other department head begins even unconsciously, to imitate his manner and ways.

Source: http://www.thehindubusinessline.in/manager/2008/02/11/stories/2008021150321100.htm

4.6 Summary

- When used without any qualification, a **colouring** of a graph is almost always a proper vertex colouring, namely a labelling of the graph's vertices with colours such that no two vertices sharing the same edge have the same colour.

- Graphs with large cliques have high chromatic number, but the opposite is not true.

- The Grötzsch graph is an example of a 4-chromatic graph without a triangle, and the example can be generalised to the Mycielskians.

- Determining if a graph can be coloured with 2 colours is equivalent to determining whether or not the graph is bipartite, and thus computable in linear time using breadth-first search.

4.7 Keywords

Brute-force Search : Brute-force search for a *k*-colouring considers every of the k^n assignments of *k* colours to *n* vertices and checks for each if it is legal.

Compiler: A compiler is a computer program that translates one computer language into another.

Vertex Colouring: Vertex colouring is the starting point of the subject, and other colouring problems can be transformed into a vertex version.

4.8 Self Assessment

Fill in the blanks:

1. The convention of using colours …………… from colouring the countries of a map, where each face is literally coloured.

2. Graph colouring enjoys many ……………. applications as well as theoretical challenges.

3. The nature of the colouring problem depends on the …………… of colours but not on what they are.

4. An ………………. of a graph, is a proper colouring of the *edges*, meaning an assignment of colours to edges so that no vertex is incident to two edges of the same colour.

5. A colouring using at most k colours is called a (proper) …………………..**.**

6. The smallest number of colours needed to colour a graph G is called its……………….. **,** $\chi(G)$.

7. The chromatic polynomial includes at least as much information about the ……………….. of G as does the chromatic number.

8. The **chromatic polynomial** counts the number of ways a graph can be …………. using no more than a given number of colours.

9. A greedy colouring shows that every graph can be coloured with one more colour than the ……………. vertex degree.

10. Graphs with large ………………….. have high chromatic number, but the opposite is not true.

11. In the field of distributed algorithms, graph colouring is closely related to the problem of ………………...

12. In a symmetric graph, a deterministic distributed algorithm cannot find a proper vertex colouring. Some auxiliary information is needed in order to ……………….

13. A standard assumption is that initially each node has a …………………

14. ………………. Vertex colouring models to a number of scheduling problems.

15. The technique by Cole and Vishkin can be applied in ……………… bounded-degree graphs as well.

4.9 Review Questions

1. What do you understand by colouring of graphs?

2. Describe vertex colouring.

3. What do you mean by edge colouring?

4. What are the main properties of colouring of a graph?

5. Discuss applications of colouring of a graph.

6. What is Pólya enumeration theorem?

7. Give the formal statement of Pólya enumeration theorem and its proof.SchedulingOpen Problems.

8. Explain the major advantages and disadvantages of graph colouring.

9.   Explain how graph colouring helps in the preparation of cars.

10.   Nowadays graph colouring is very much booming. Why?

Answers: Self Assessment

1.   Originates   2.   Practical

3.   Number   4.   edge colouring

5.   k-colouring   6.   chromatic number

7.   colourability   8.   coloured

9.   maximum   10.   cliques

11.   symmetry breaking   12.   break symmetry

13.   unique identifier   14.   Scheduling

15.   arbitrary

4.10 Further Readings

Books   Béla Bollobás, *Modern Graph Theory*, Springer

                 Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph Theory, Combinatorics, and Algorithms*, Birkhäuser

Online links   http://en.wikipedia.org/wiki/Graph_colouring

                 http://www.streamtech.nl/problemset/193.html

Unit 5: Tree Graphs

Objectives

After studying this unit, you will be able to:

- Define directed graph and tree graph

- Describe the types of binary trees

- Explain the properties of binary trees

Introduction

In mathematics, more specifically graph theory, a **tree** is an undirected graph in which any two vertices are connected by *exactly one* simple path. In other words, any connected graph without cycles is a tree. A **forest** is a disjoint union of trees.

The various kinds of data structures referred to as trees in computer science are similar to trees in graph theory, except that computer science trees have directed edges. Although they do not meet the definition given here, these graphs are referred to in graph theory as ordered directed trees.
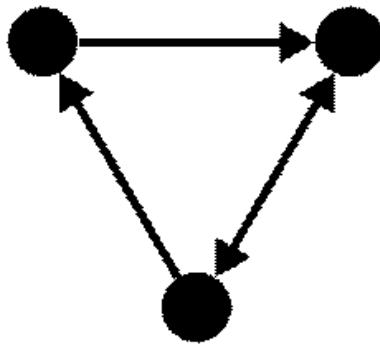
5.1 Directed Graph

A **directed** graph or **digraph** is a pair $G = (V,A)$ (sometimes $G = (V,E)$) of:

1. a set V, whose elements are called vertices or nodes,

2. a set A of ordered pairs of vertices, called **arcs**, **directed edges**, or **arrows** (and sometimes simply **edges** with the corresponding set named *E* instead of *A*).

It differs from an ordinary or undirected graph, in that the latter is defined in terms of unordered pairs of vertices, which are usually called edges.

Sometimes a digraph is called a simple digraph to distinguish it from a directed multigraph, in which the arcs constitute a multiset, rather than a set, of ordered pairs of vertices. Also, in a simple digraph loops are disallowed. (A loop is an arc that pairs a vertex to itself.) On the other hand, some texts allow loops, multiple arcs, or both in a digraph.

An arc $e = (x,y)$ is considered to be directed **from** *x* **to** *y*; *y* is called the **head** and *x* is called the **tail** of the arc; *y* is said to be a **direct successor** of *x*, and *x* is said to be a **direct predecessor** of *y*. If a path made up of one or more successive arcs leads from *x* to *y*, then *y* is said to be a **successor** of *x*, and *x* is said to be a **predecessor** of *y*. The arc (y,x) is called the arc (x,y) **inverted**.

A directed graph *G* is called **symmetric** if, for every arc that belongs to *G*, the corresponding inverted arc also belongs to *G*. A symmetric loopless directed graph is equivalent to an undirected graph with the pairs of inverted arcs replaced with edges; thus the number of edges is equal to the number of arcs halved.

A distinction between a simple directed graph and an oriented graph is that if *x* and *y* are vertices, a simple directed graph allows both (*x*,*y*) and (*y*,*x*) as edges, while only one is permitted in an oriented graph.

A **weighted digraph** is a digraph with weights assigned for its arcs, similarly to the weighted graph.

The adjacency matrix of a digraph (with loops and multiple arcs) is the integer-valued matrix with rows and columns corresponding to the digraph nodes, where a nondiagonal entry a_{ij} is the number of arcs from node *i* to node *j*, and the diagonal entry a_{ii} is the number of loops at node *i*. The adjacency matrix for a digraph is unique up to the permutations of rows and columns.
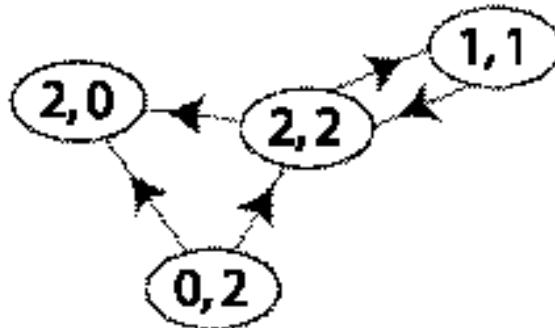
Did u know? **What is oriented graph?**

The **orientation** of a simple undirected graph is obtained by assigning a direction to each edge. Any directed graph constructed this way is called an **oriented graph**.

Indegree and Outdegree

Figure 5.1: A Digraph with Vertices Labeled (Indegree, Outdegree)

For a node, the number of head endpoints adjacent to a node is called the **indegree** of the node and the number of tail endpoints is its **outdegree**.

The indegree is denoted $\deg^-(v)$ and the outdegree as $\deg^+(v)$. A vertex with $\deg^-(v) = 0$ is called a **source**, as it is the origin of each of its incident edges. Similarly, a vertex with $\deg^+(v) = 0$ is called a **sink**.

The **degree sum formula** states that, for a directed graph

$$\sum_{v \in V} \deg^+(v) = \sum_{v \in V} \deg^-(v) = |A|.$$

Notes   If for every node, $v \in V$, $\deg^+(v) = \deg^+(v)$, the graph is called a **balanced digraph**.
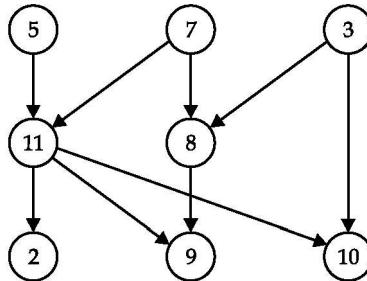
Digraph Connectivity

A digraph G is called weakly connected (or just connected) if the undirected underlying graph obtained by replacing all directed edges of G with undirected edges is a connected graph. A digraph is strongly connected or strong if it contains a directed path from u to v and a directed path from v to u for every pair of vertices u,v. The strong components are the maximal strongly connected subgraphs.

Classes of Digraphs
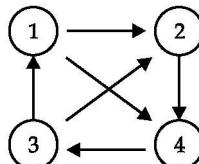
Figure 5.2: A Simple Directed Acyclic Graph

An acyclic digraph (occasionally called a **dag** or **DAG** for "directed acyclic graph", although it is not the same as an orientation of an acyclic graph) is a directed graph with no directed cycles.

A rooted tree naturally defines an acyclic digraph, if all edges of the underlying tree are directed away from the root.

Figure 5.3: A Tournament on 4 Vertices

A tournament is an oriented graph obtained by choosing a direction for each edge in an undirected complete graph.

In the theory of Lie groups, a quiver Q is a directed graph serving as the domain of, and thus characterizing the shape of, a representation V defined as a functor, specifically an object of the functor category FinVctKF(Q) where F(Q) is the free category on Q consisting of paths in Q and FinVctK is the category of finite dimensional vector spaces over a field K. Representations of a quiver label its vertices with vector spaces and its edges (and hence paths) compatibly with linear transformations between them, and transform via natural transformations.

Task   An acyclic digraph is a directed graph with no directed cycles. Comment

---

5.2 Tree Graphs

A tree is a connected graph which has no cycler.

Trees are a relatively simple type of graph but they are also very important. Many applications use trees as a mathematical representation, e.g., decision trees in OR, some utility networks, linguistic analysis, family trees, organisation trees.
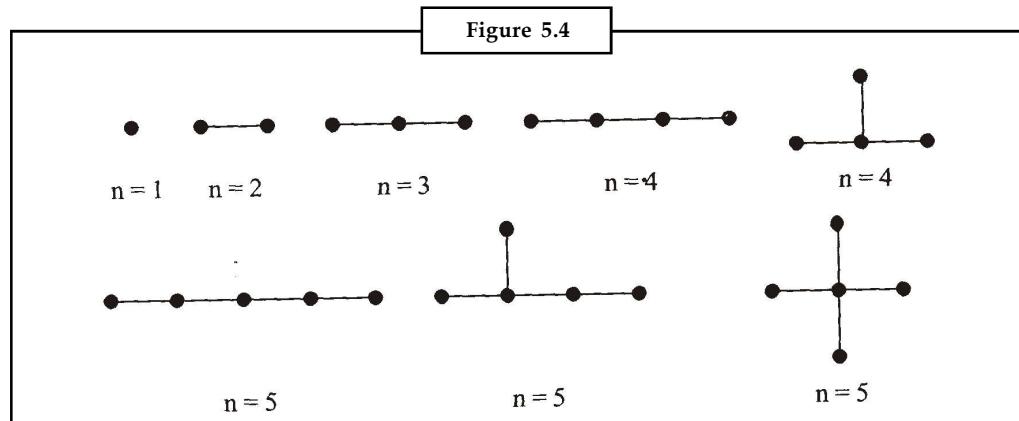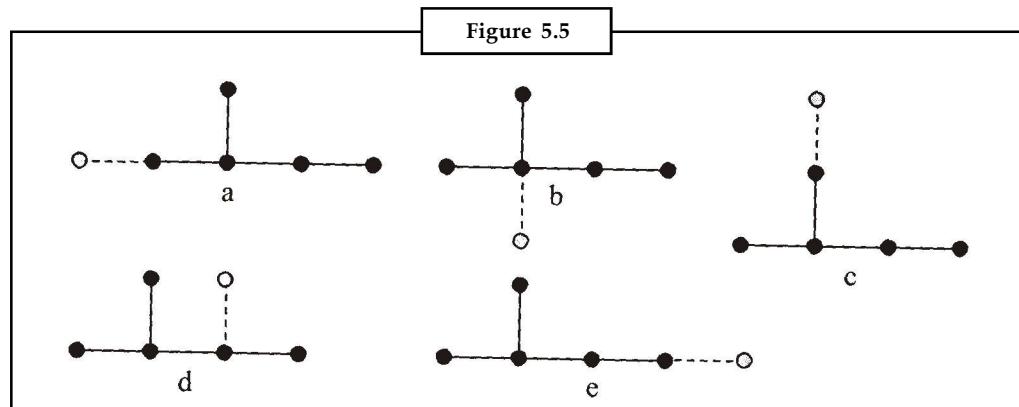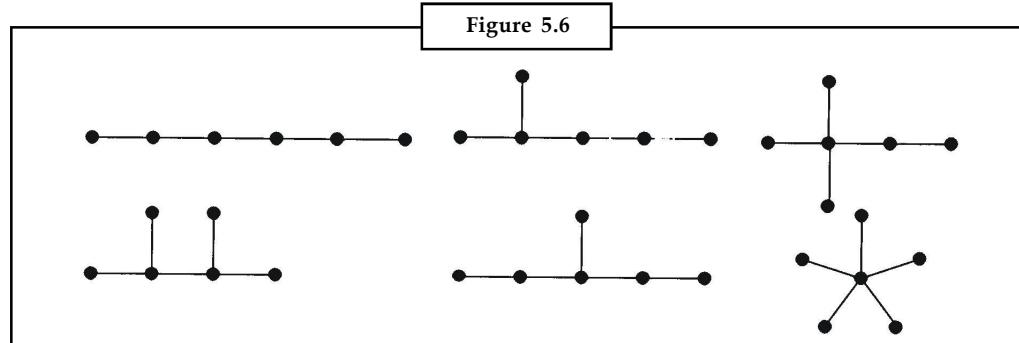
Figure 5.4

Figure 5.4 shows all the possible unlabelled trees with up to five vertices. Every tree with $n + 1$ vertices can be formed by adding a new vertex joined by a new edge to one of the vertices of one of the n vertex trees. For instance if we take the second 5-vertex tree we would obtain the trees shown in Figure 5.5

Figure 5.5

Of course (*a*) and (*c*) are isomorphic. If we complete this process with all three 5-vertex trees and eliminate the isomorphic duplicates we obtain the six trees with 6 vertices shown in Figure 5.6.

Figure 5.6

Trees have a number of special properties as follows:

1. It is obvious, from the constructive process of building all possible trees step by step from the simplest tree (one vertex, no edges) that a tree with *n* vertices has exactly *n* – 1 edges.

2. When a new vertex and edges is added to a tree, no cycle is created (since the new edge joins an existing vertex to a new vertex) and the tree remains connected.

3. There is exactly one path from any vertex in a tree to any other vertex—if there were two or more paths between any two vertices then the two paths would form a cycle and the graph would not be a tree.

4. Because there is exactly one path between any two vertices then there is one (and only one) edge joining any two adjacent vertices. If this edge is removed, the graph is no longer connected (and so is not a tree). So the removal of any edge from a tree disconnects the graph.

5. Since there is a path between any two vertices, if an edge is added to the tree joining two existing vertices then a cycle is formed comprising the existing path between the two vertices together with the new edge.

All these properties can be used to define a tree. If T is a graph with *n* vertices then the following are all equivalent definitions of a tree :

>   T is connected and has no cycles.

>   T has *n* – 1 edges and has no cycles.
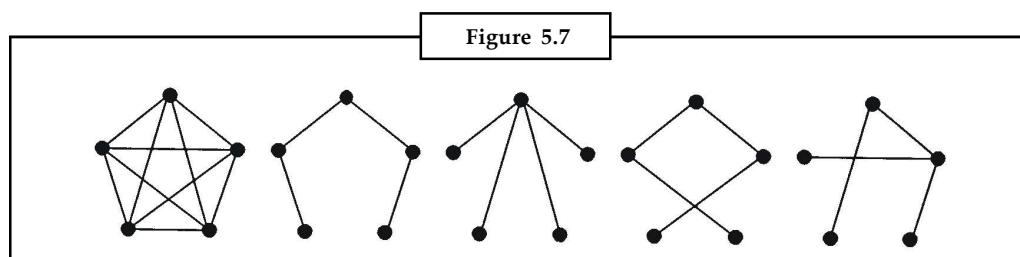
>   T is connected and has *n* – 1 edges.

>   Any two vertices of T are connected by exactly one path.

>   T is connected and the removal of any edge disconnects T.

>   T contains no cycles but the addition of any new edge creates a cycle.

A **spanning tree** in a connected graph G is a subgraph which includes every vertex and a tree. For instance Figure 5.7 below shows the complete graph K_5 and several possible spanning trees. Large and complex graphs may have very many spanning trees.

Figure 5.7
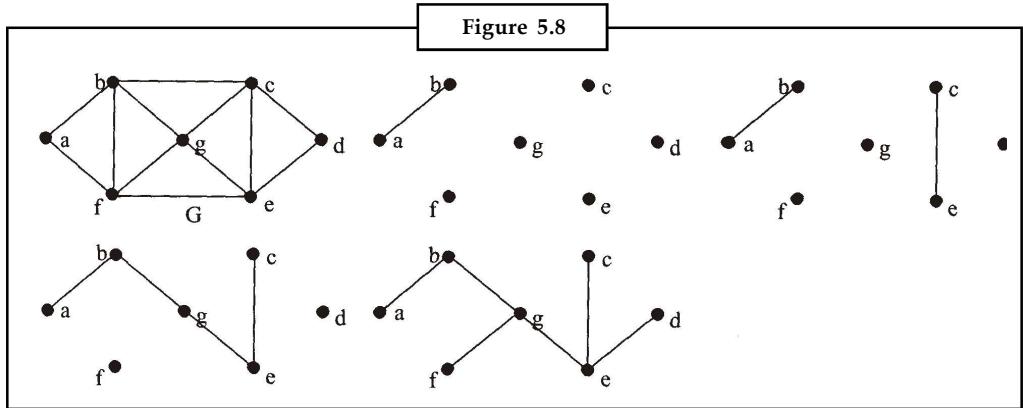
A spanning tree may be found by the building-up method or the cutting-down method. The building-up algorithm is select edges of the graph, one by one, such a way that no cycles are created; repeating until all vertices are included and the cutting-down method is choose any cycle in the graph and remove one of its edges; repeating until no cycles remain.

For instance, in Figure 5.8 below a spanning tree in the graph G is built up by selecting successively edges *ab* (1st diagram), then *ce* (2nd diagram), then *bg*, then *ge* (3rd diagram), then *gf* and finally *de* (final diagram).
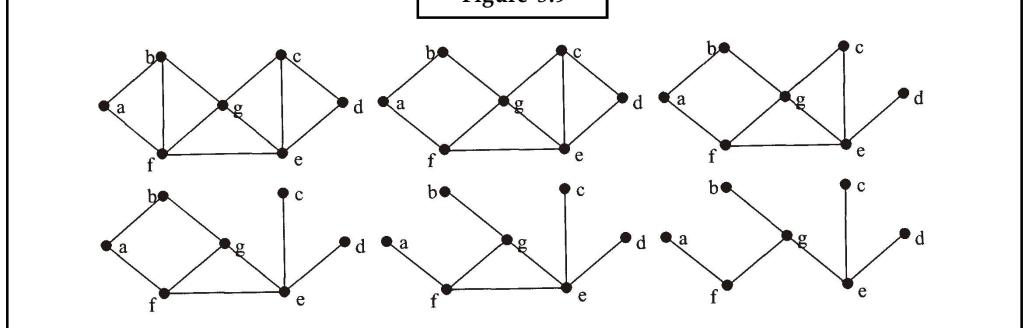
Figure 5.8

In Figure 5.9, a spanning tree in the graph G of Figure 5.8 is derived by the cutting down method, by successively finding cycles and removing edges. First cycle is *abcdefa*—remove *bc* (1st diagram).
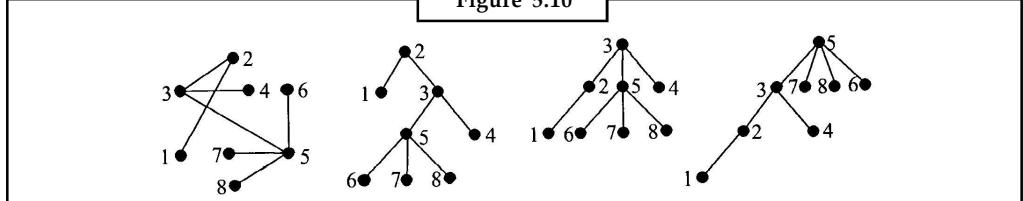
Cycle *bgfb*—remove *fb* (2nd diagram). Cycle *gedcg*—remove *cd*. Cycle *cgec*—remove *gc*. Cycle *gfabg*—removed *ab*. Cycle *gfeg*—remove *fe*. No more cycles so this is a spanning tree.

Figure 5.9
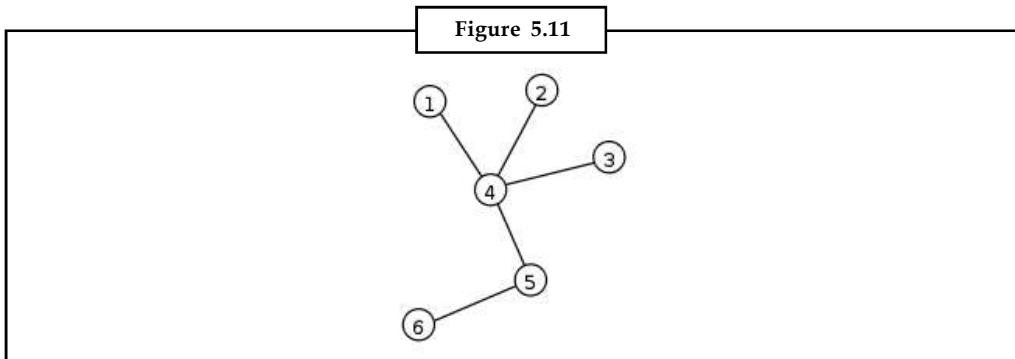
A **rooted tree** is a tree in which one vertex is selected as a root and all other edges branch out from the root vertex. Any given tree can be drawn as a rooted tree in a variety of ways depending on the choice of root vertex (see Figure 5.10).

Figure 5.10

In mathematics, more specifically graph theory, a **tree** is an undirected graph in which any two vertices are connected by *exactly one* simple path. In other words, any connected graph without cycles is a tree. A **forest** is a disjoint union of trees.

The various kinds of data structures referred to as trees in computer science are similar to trees in graph theory, except that computer science trees have directed edges. Although they do not meet the definition given here, these graphs are referred to in graph theory as ordered directed trees.

Figure 5.11

A **tree** is an undirected simple graph *G* that satisfies any of the following equivalent conditions:

1.   G is connected and has no cycles.

2.   G has no cycles, and a simple cycle is formed if any edge is added to G.

3.   G is connected, and it is not connected anymore if any edge is removed from G.

4.   G is connected and the 3-vertex complete graph K$_3$ is not a minor of G.

5.   Any two vertices in *G* can be connected by a unique simple path.

If *G* has finitely many vertices, say *n* of them, then the above statements are also equivalent to any of the following conditions:

1.   *G* is connected and has *n* " 1 edges.

2.   *G* has no simple cycles and has *n* " 1 edges.

Did u know? **You know irreducible tree?**

An **irreducible** (or *series-reduced*) tree is a tree in which there is no vertex of degree 2.

A **forest** is an undirected graph, all of whose connected components are trees; in other words, the graph consists of a disjoint union of trees. Equivalently, a forest is an undirected cycle-free graph. As special cases, an empty graph, a single tree, and the discrete graph on a set of vertices (that is, the graph with these vertices that has no edges), all are examples of forests.

The term **hedge** sometimes refers to an ordered sequence of trees.

A **polytree** or **oriented tree** is a directed graph with at most one undirected path between any two vertices. In other words, a polytree is a directed acyclic graph for which there are no undirected cycles either.

A **directed tree** is a directed graph which would be a tree if the directions on the edges were ignored. Some authors restrict the phrase to the case where the edges are all directed towards a particular vertex, or all directed away from a particular vertex.

A tree is called a **rooted tree** if one vertex has been designated the **root**, in which case the edges have a natural orientation, *towards* or *away* from the root. The *tree-order* is the partial ordering on the vertices of a tree with *u* ≤ *v* if and only if the unique path from the root to *v* passes through *u*. A rooted tree which is a subgraph of some graph *G* is a **normal tree** if the ends of every edge in *G* are comparable in this tree-order whenever those ends are vertices of the tree (Diestel 2005, p. 15). Rooted trees, often with additional structure such as ordering of the neighbors at each vertex, are a key data structure in computer science; see tree data structure.

> *Notes*   In a context where trees are supposed to have a root, a tree without any designated root is called a **free tree**.

In a rooted tree, the **parent** of a vertex is the vertex connected to it on the path to the root; every vertex except the root has a unique parent. A **child** of a vertex v is a vertex of which v is the parent. A **leaf** is a vertex without children.

A **labeled tree** is a tree in which each vertex is given a unique label. The vertices of a labeled tree on n vertices are typically given the labels 1, 2, …, n. A **recursive tree** is a labeled rooted tree where the vertex labels respect the tree order (*i.e.*, if $u < v$ for two vertices u and v, then the label of u is smaller than the label of v).

An **ordered tree** is a rooted tree for which an ordering is specified for the children of each vertex.

An **n-ary tree** is a rooted tree for which each vertex which is not a leaf has at most n children. *2-ary trees* are sometimes called binary trees, while *3-ary trees* are sometimes called ternary trees.

A **terminal vertex** of a tree is a vertex of degree 1. In a rooted tree, the leaves are all terminal vertices; additionally, the root, if not a leaf itself, is a terminal vertex if it has precisely one child.

Example: The example tree shown in the Figure 5.11 has 6 vertices and 6–1 = 5 edges. The unique simple path connecting the vertices 2 and 6 is 2-4-5-6.

Facts

1. Every tree is a bipartite graph and a median graph. Every tree with only countably many vertices is a planar graph.

2. Every connected graph G admits a spanning tree, which is a tree that contains every vertex of G and whose edges are edges of G.

4. Every connected graph with only countably many vertices admits a normal spanning tree.

5. There exist connected graphs with uncountably many vertices which do not admit a normal spanning tree.

6. Every finite tree with n vertices, with n > 1, has at least two terminal vertices. This minimal number of terminal vertices is characteristic of path graphs; the maximal number, n –1, is attained by star graphs.

7. For any three vertices in a tree, the three paths between them have exactly one vertex in common.

5.3 Binary Trees

A *binary tree* is a rooted tree in which each vertex has at most two children, designated as *left child* and *right child.* If a vertex has one child, that child is designated as either a left child or a right child, but not both. A *full binary tree* is a binary tree in which each vertex has exactly two children or none. The following are a few results about binary trees:

1. If T is a full binary tree with i internal vertices, then T has $i + 1$ terminal vertices and $2i + 1$ total vertices.

2. If a binary tree of height h has t terminal vertices, then $t < 2^h$.

More generally we can define a *m-ary tree* as a rooted tree in which every internal vertex has no more than *m* children. The tree is called a *full m-ary tree* if every internal vertex has exactly *m* children. An *ordered rooted tree* is a rooted tree where the children of each internal vertex are ordered. A binary tree is just a particular case of *m-ary* ordered tree (with *m* = 2).
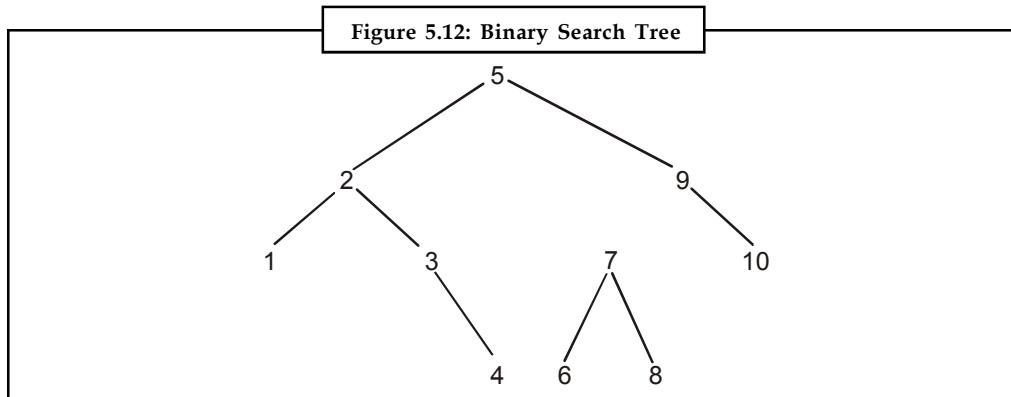
5.3.1 Binary Search Trees
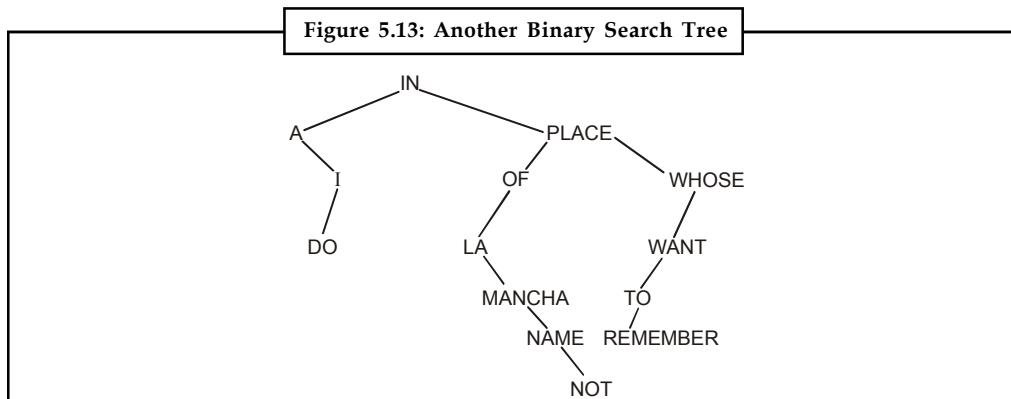
Assume S is a set in which elements (which we will call "data") are ordered; e.g., the elements of S can be numbers in their natural order, or strings of alphabetic characters in lexicographic order. A *binary search tree* associated to S is a binary tree T in which data from S are associate with the vertices of T so that, for each vertex u in T, each data item in the left subtree of v is less than the data item in v, and each data item in the right subtree of v is greater than the data item in v.

Example: Figure 5.12 below, contains a binary search tree for the set S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}. In order to find a element we start at the root and compare it to the data in the current vertex (initially the root). If the element is greater we continue through the right child, if it is smaller we continue through the left child, if it is equal we have found it. If we reach a terminal vertex without founding the element, then that element is not present in S.

Figure 5.12: Binary Search Tree

Making a Binary Search Tree: We can store data in a binary search tree by randomly choosing data from *S* and placing it in the tree in the following way: The first data chosen will be the root of the tree. Then for each subsequent data item, starting at the root we compare it to the data in the current vertex *v.* If the new data item is greater than the data in the current vertex then we move to the right child, if it is less we move to the left child. If there is no such child then we create one and put the new data in it. For instance, the tree in Figure 5.13 below has been made from the following list of words choosing them in the order they occur: "IN A PLACE OF LA MANCHA WHOSE NAME I DO NOT WANT TO REMEMBER".

Figure 5.13: Another Binary Search Tree
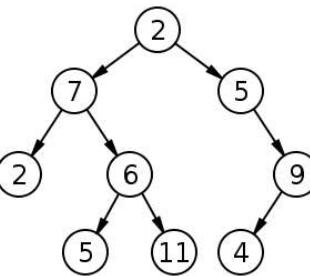
In computer science, a **binary tree** is a tree data structure in which each node has at most two child nodes, usually distinguished as "left" and "right". Nodes with children are parent nodes, and child nodes may contain references to their parents. Outside the tree, there is often a reference to the "root" node (the ancestor of all nodes), if it exists. Any node in the data structure can be reached by starting at root node and repeatedly following references to either the left or right child.

Binary trees are used to implement binary search trees and binary heaps.

Figure 5.14

1.   A directed edge refers to the link from the parent to the child (the arrows in the picture of the tree).

2.   The root node of a tree is the node with no parents. There is at most one root node in a rooted tree.

3.   A leaf node has no children.

4.   The depth of a node n is the length of the path from the root to the node. The set of all nodes at a given depth is sometimes called a level of the tree. The root node is at depth zero.

5.   The height of a tree is the length of the path from the root to the deepest node in the tree. A (rooted) tree with only one node (the root) has a height of zero.

6.   Siblings are nodes that share the same parent node.

7.   A node p is an ancestor of a node q if it exists on the path from q to the root. The node q is then termed a descendant of p.

8.   The size of a node is the number of descendants it has including itself.

9.   In-degree of a node is the number of edges arriving at that node.

10.   Out-degree of a node is the number of edges leaving that node.

11.   The root is the only node in the tree with In-degree = 0.
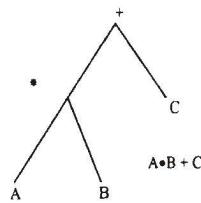
Example: Depth of tree with level =3,then, size of tree is, level+1=4

5.3.2 Traversal of a Binary Tree

Tree traversal is one of the most common operations performed on tree data structures. It is a way in which each node in the tree is visited exactly once in a systematic manner. There are many applications that essentially require traversal of binary trees. For example, a binary tree could be used to represent an arithmetic expression as shown in Figure 5.15.

Figure 5.15

The full binary tree traversal would produce a linear order for the nodes in a binary tree. There are three ways of binary tree traversal,

1. In-order traversal

2. Pre-order traversal

3. Post-order traversal
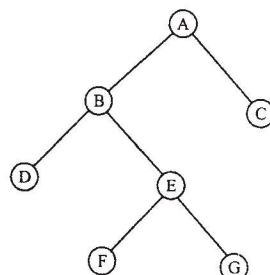
Inorder Traversal

The in-order traversal of a non-empty tree is defined as follows:

1. Traverse the left subtree inorder (L).

2. Visit the root node (N)

3. Traverse the right subtree in order (R).

In Figure 5.16 inorder traversal of a binary tree is DBFEGAC.

Figure 5.16

Pre-order Traversal

The pre-order traversal of a non-empty binary tree is defined as follows:

1. Visit the root node (N).

2. Traverse the left subtree in pre-order (L)

3. Traverse the right subtree in pre-order (R)

In Figure 5.16 the preorder traversal of a binary tree is ABDEFGC.

Postorder Traversal
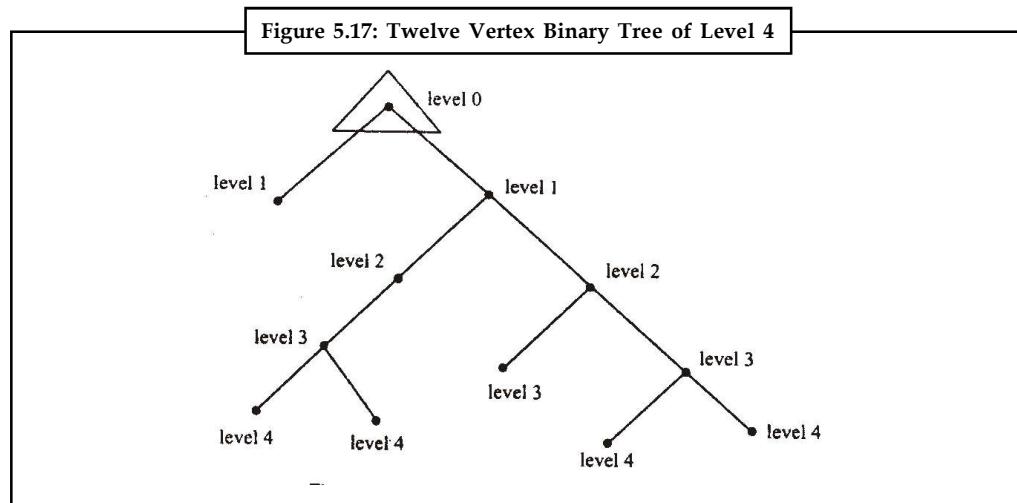
The postorder traversal of non-empty binary tree is defined as follows:

1.   Traverse the left subtree in postorder (L).

2.   Traverse the right subtree in postorder (R).

3.   Visit the root node (N).

In Figure 5.16 the postorder traversal of a binary tree is DFGEBC A.

Level of a Vertex in a Full Binary Tree

In a binary tree the distance of a vertex v_i from the root of the tree is called the level of v_i and is denoted by I_i. Thus level of the root is zero. Levels of the various vertices in the following tree have been denoted by numbers written adjacent to the respective vertices.

Figure 5.17: Twelve Vertex Binary Tree of Level 4
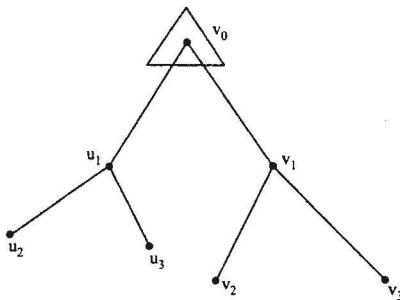
Number of Vertices of Different Levels in a Binary Tree

In a binary tree there will be two edges adjacent to the root vertex v_0. Let these edges be v_0u_1 and v_0v_1. Levels of each of the vertices u_1 and v_1 is 1. So maximum number of vertices of level 0 is $1(=2^0)$ and maximum number of vertices with level 1 is $=2^1$.

Again there can be either 0 or 2 edges adjacent to each of the vertices u_1 and v_1. Let these edges be u_1u_2, u_1u_3, v_1v_2 and v_1v_3. Levels of each of the four vertices u_2, u_3, v_2, v_3 is 2. So maximum number of vertices of level 2 is $4(= 2^2)$. In a similar way the levels of the 8 vertices that will be obtained by adding two edges to each of the four vertices u_2, u_3, v_2, v_3 shall be 3. So maximum number of vertices each of level 3 is $8(=2^3)$. Not more than two edges can be added to any of the vertices so obtained to keep the degree of that vertex as 3.

Figure 5.18

Proceeding in this way we see that the maximum number of vertices in a n level binary tree at levels 0, 1, 2, 3,... shall be 2^0. 2^1, 2^2, 2^3,... respectively.

whose sum = $2^0 + 2^1 + 2^2 + 2^3 + ... + 2^n$.

The maximum level of any vertex in a binary tree (denoted by l_{max}) is called *height of the tree.* The minimum possible height of an *n* vertex binary tree is $[\log_2(n + 1) - 1]$ which is equal *to* the

smallest integer $\geq [\log_2(n + 1) - 1]$ and Max. $l_{max} = \dfrac{n - 1}{2}$

Example: If a tree T has n vertices of degree 1, 3 vertices of degree 2, 2 vertices of 3 and 2 vertices of degree 4 find the value of *n*.

Solution: Let $|E|$ denote the number of edges in the graph T and $|V|$ denote the number of vertices in the same graph T.

$$\text{Sum of degrees of all vertices in T} = 2\,|E|$$

$$n.1 + 3.2 + 2.3 + 2.4 = 2\,|E|$$

or $$n + 6 + 6 + 8 = 2(|V| - 1)$$

or $$n + 20 = 2[(n + 3 + 2 + 2) - 1]$$

or $$n + 20 = 2n + 12$$

or $$n = 8.$$

Theorem 1: To prove that in every non-trivial tree there is at least one vertex of degree one.

Proof: Let us start at vertex v_1. If $d(v_1) = 1$, then the theorem is already proved. If $d(v_1) > 1$, then we can move to a vertex say v_2 that is adjacent to v_1. Now if $d(v_2) > 1$, again move to another vertex say v_3 that is adjacent to v_2. In this way we can continue to produce a path v_1, v_2, v_3 .... (without repetition of any vertex, in order to avoid formation of circuit as the graph is a tree). As the graph is finite, this path must end at some vertex whose degree shall be one because we shall only enter this vertex and cannot exit from it.

5.3.3 Types of Binary Trees

1. A rooted binary tree is a tree with a root node in which every node has at most two children.

2. A full binary tree (sometimes proper binary tree or 2-tree or strictly binary tree) is a tree in which every node other than the leaves has two children.

3. A perfect binary tree is a full binary tree in which all leaves are at the same depth or same level. (This is ambiguously also called a complete binary tree.)

4. A complete binary tree is a binary tree in which every level, except possibly the last, is completely filled, and all nodes are as far left as possible.

5. An infinite complete binary tree is a tree with N_0 levels, where for each level d the number of existing nodes at level d is equal to 2d. The cardinal number of the set of all nodes is N_0. The cardinal number of the set of all paths is 2^{N0}. The infinite complete binary tree essentially describes the structure of the Cantor set; the unit interval on the real line (of cardinality 2^{N0}) is the continuous image of the Cantor set; this tree is sometimes called the Cantor space.

6. A balanced binary tree is commonly defined as a binary tree in which the height of the two subtrees of every node never differ by more than 1., although in general it is a binary tree where no leaf is much farther away from the root than any other leaf. (Different balancing schemes allow different definitions of "much farther" [4]). Binary trees that are balanced according to this definition have a predictable depth (how many nodes are traversed from the root to a leaf, root counting as node 0 and subsequent as 1, 2, ..., depth). This depth is equal to the integer part of $\log_2(n)$ where n is the number of nodes on the balanced tree.

    Example:

    1. balanced tree with 1 node, $\log_2(1) = 0$ (depth = 0).

    2. balanced tree with 3 nodes, $\log_2(3) = 1.59$ (depth=1).

    3. balanced tree with 5 nodes, $\log_2(5) = 2.32$ (depth of tree is 2 nodes).

7. A rooted complete binary tree can be identified with a free magma.

8. A degenerate tree is a tree where for each parent node, there is only one associated child node. This means that in a performance measurement, the tree will behave like a linked list data structure.

9. A Tango tree is a tree optimized for fast searches.

Notes   This terminology often varies in the literature, especially with respect to the meaning "complete" and "full".

---

10. A Strictly Binary Tree: Its when the tree is fully expanded i.e., with 2 degree expansion.

5.3.4 Properties of Binary Trees

1. The number of nodes n in a perfect binary tree can be found using this formula: n = 2h + 1 –1 where h is the height of the tree.

2. The number of nodes n in a complete binary tree is minimum: n = 2h and maximum: n = 2h + 1 – 1 where h is the height of the tree.

3. The number of leaf nodes L in a perfect binary tree can be found using this formula: L = 2h where h is the height of the tree.

4. The number of nodes n in a perfect binary tree can also be found using this formula: n = 2L – 1 where L is the number of leaf nodes in the tree.

5. The number of NULL links in a Complete Binary Tree of n-node is (n+1).

6. The number of leaf nodes in a Complete Binary Tree of n-node is [n/2].

7. For any non-empty binary tree with n0 leaf nodes and n2 nodes of degree 2, n0 = n2 + 1.[5]

8. n = n0 + n1 + n2 + n4 + n3 + n5 + ... + nB-1 + nB

9. B = n - 1, n = 1 + 1*n1 + 2*n2 + 3*n3 + 4*n4 + ... + B*nB, NOT include n0

5.4 Summary

- A full binary tree (sometimes proper binary tree or 2-tree or strictly binary tree) is a tree in which every node other than the leaves has two children.

- A perfect binary tree is a full binary tree in which all leaves are at the same depth or same level. (This is ambiguously also called a complete binary tree.)

- A complete binary tree is a binary tree in which every level, except possibly the last, is completely filled, and all nodes are as far left as possible.

- An infinite complete binary tree is a tree with N_0 levels, where for each level d the number of existing nodes at level d is equal to 2d. The cardinal number of the set of all nodes is N_0. The cardinal number of the set of all paths is 2^{N_0}. The infinite complete binary tree essentially describes the structure of the Cantor set; the unit interval on the real line (of cardinality 2^{N_0}) is the continuous image of the Cantor set; this tree is sometimes called the Cantor space.

- A directed graph *G* is called **symmetric** if, for every arc that belongs to *G*, the corresponding inverted arc also belongs to *G*.

- Trees are a relatively simple type of graph but they are also very important. Many applications use trees as a mathematical representation.

- A tree is called a **rooted tree** if one vertex has been designated the **root**, in which case the edges have a natural orientation, *towards* or *away* from the root.

5.5 Keywords

N-ary Tree: An n-ary tree is a rooted tree for which each vertex which is not a leaf has at most n children. 2-ary trees are sometimes called binary trees, while 3-ary trees are sometimes called ternary trees.

Ordered Tree: An ordered tree is a rooted tree for which an ordering is specified for the children of each vertex.

Recursive Tree: A recursive tree is a labeled rooted tree where the vertex labels respect the tree order.

Terminal Vertex: A terminal vertex of a tree is a vertex of degree 1. In a rooted tree, the leaves are all terminal vertices; additionally, the root, if not a leaf itself, is a terminal vertex if it has precisely one child.

Weighted Digraph: A weighted digraph is a digraph with weights assigned for its arcs, similarly to the weighted graph.
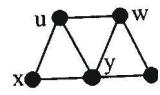
5.6 Self Assessment

Fill in the blanks:

1. A rooted tree naturally defines an …………. digraph, if all edges of the underlying tree are directed away from the root.

2. A spanning tree in a connected graph G is a …………….. which includes every vertex and a tree.

3. A ………………… may be found by the building-up method or the cutting-down method.

4. A polytree or oriented tree is a ……………… with at most one undirected path between any two vertices.

5. A labeled tree is a tree in which each ……………… is given a unique label.

6. A directed tree is a directed graph which would be a tree if the ……………. on the edges were ignored.

7. A binary tree is a tree data structure in which each node has at most ………….. child nodes, usually distinguished as "left" and "right".

8. A perfect binary tree is a full binary tree in which all leaves are at the same ……………… or same level.

5.7 Review Questions

1. By adding a new edge in every possible way to each unlabelled tree with 6 vertices, draw the 11 unlabelled trees with 7 vertices.

2. Give an example of a tree with eight vertices and

    (a) exactly 2 vertices of degree 1,

    (b) exactly 4 vertices of degree 1, and

    (c) exactly 7 vertices of degree 1.

3. Use the Handshaking Lemma to show that every tree with n vertices, where $n \geq 2$, has at least 2 vertices of degree 1.

4. (a) Find 3 different spanning trees in the graph shown using the building up method.

    (b) Find 3 more spanning trees (different from those found in part (*a*) using the cutting down method.

Answers: Self Assessment

1.	Acyclic	2.	Subgraph
3.	spanning tree	4.	directed graph
5.	vertex	6.	directions
7.	two	8.	depth

5.8 Further Readings

Books

Béla Bollobás, *Modern Graph Theory*, Springer

Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph Theory, Combinatorics, and Algorithms*, Birkhäuser

Online links

http://en.wikipedia.org/wiki/File:Tree_graph.svg

http://cslibrary.stanford.edu/110/BinaryTrees.html

Unit 6: Algorithm

CONTENTS

Objectives

Introduction

Objectives

After studying this unit, you will be able to:

- Discuss minimal spanning tree
- Explain prim's algorithm
- Describe shortest path

Introduction

In last unit you studied about the graph and tree. The graph is such a general data structure that almost all computational problems can be formulated using one of the primary graph processing algorithms. Lists and trees are subsets of graphs. The major problems in hardware synthesis, operating system scheduling and sharing, compiler optimization, software design and minimization, network communication and synchronization, and requirements and specification modeling are graph problems. There are many algorithms that can be applied to graphs. Many of these are actually used in the real world, such as Dijkstra algorithm to find shortest paths. This unit provides you clear understanding of these algorithms.

6.1 Minimal Spanning Tree
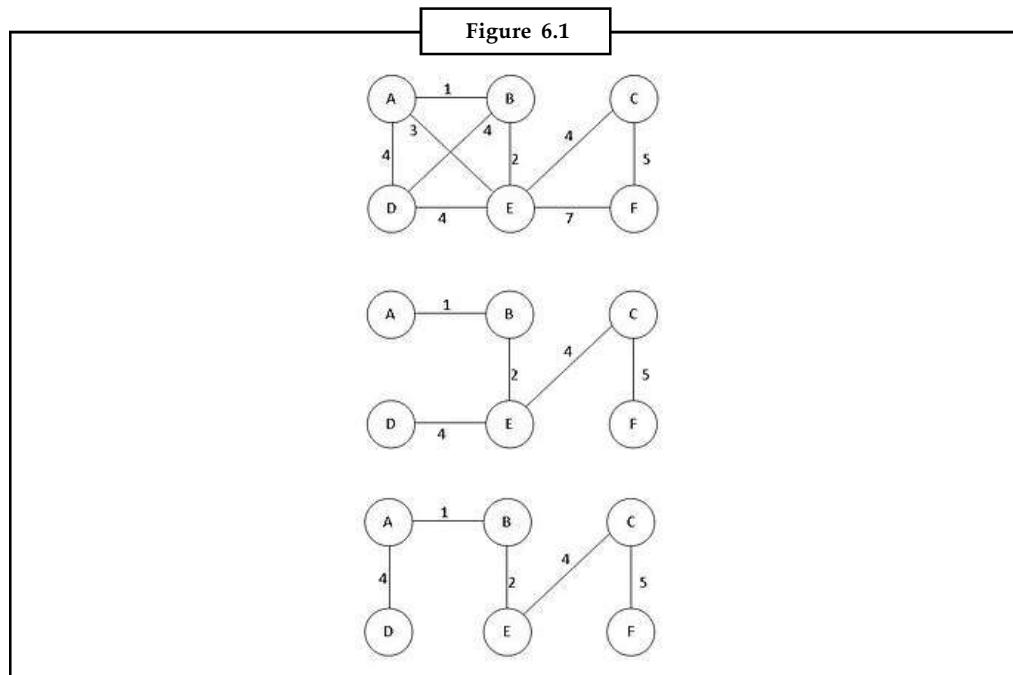
Given a connected, undirected graph, a spanning tree of that graph is a subgraph that is a tree and connects all the vertices together. A single graph can have many different spanning trees. We can also assign a weight to each edge, which is a number representing how unfavorable it is, and use this to assign a weight to a spanning tree by computing the sum of the weights of the edges in that spanning tree. A minimum spanning tree (MST) or minimum weight spanning tree is then a spanning tree with weight less than or equal to the weight of every other spanning tree. More generally, any undirected graph (not necessarily connected) has a minimum spanning forest, which is a union of minimum spanning trees for its connected components.

One example would be a cable TV company laying cable to a new neighborhood. If it is constrained to bury the cable only along certain paths, then there would be a graph representing which points are connected by those paths. Some of those paths might be more expensive, because they are longer, or require the cable to be buried deeper; these paths would be represented by edges with larger weights. A spanning tree for that graph would be a subset of those paths that has no cycles but still connects to every house. There might be several spanning trees possible. A minimum spanning tree would be one with the lowest total cost.

6.1.1 Properties

1. *Possible multiplicity:* This figure shows there may be more than one minimun spanning tree in a graph. In the figure, the two trees below the graph are two possibilities of minimum spanning tree of the given graph.

Figure 6.1
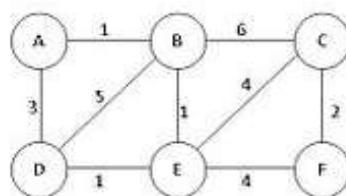
There may be several minimum spanning trees of the same weight having a minimum number of edges; in particular, if all the edge weights of a given graph are the same, then every spanning tree of that graph is minimum. If there are n vertices in the graph, then each tree has n-1 edges.
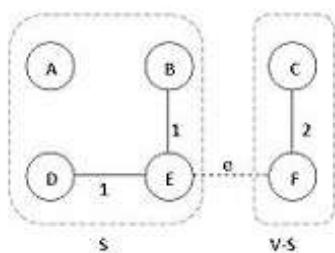
2. *Uniqueness:* If each edge has a distinct weight then there will only be one, unique minimum spanning tree. This can be proved by induction or contradiction. This is true in many realistic situations, such as the cable TV company example above, where it's unlikely any two paths have exactly the same cost. This generalizes to spanning forests as well.

    A proof of uniqueness by contradiction is as follows.

    (a) Say we have an algorithm that finds an MST (which we will call A) based on the structure of the graph and the order of the edges when ordered by weight. (Such algorithms do exist, see below.)

    (b) Assume MST A is not unique.

    (c) There is another spanning tree with equal weight, say MST B.

    (d) Let e1 be an edge that is in A but not in B.

    (e) As B is a MST, $\{e_1\} \cup B$ must contain a cycle C.

    (f) Then B should include at least one edge e_2 that is not in A and lies on C.

    (g) Assume the weight of e_1 is less than that of e_2.

    (h) Replace e_2 with e_1 in B yields the spanning tree $\{e_1\} \cup B - \{e_2\}$ which has a smaller weight compared to B.

    (i) Contradiction. As we assumed B is a MST but it is not.

    If the weight of e_1 is larger than that of e_2, a similar argument involving tree $\{e_2\} \cup A - \{e_1\}$ also leads to a contradiction. Thus, we conclude that the assumption that there can be a second MST was false.

3. *Minimum-cost subgraph:* If the weights are *positive*, then a minimum spanning tree is in fact the minimum-cost subgraph connecting all vertices, since subgraphs containing cycles necessarily have more total weight.

4. *Cycle property:* For any cycle C in the graph, if the weight of an edge e of C is larger than the weights of other edges of C, then this edge cannot belong to an MST. Assuming the contrary, i.e. that e belongs to an MST T1, then deleting e will break T1 into two subtrees with the two ends of e in different subtrees. The remainder of C reconnects the subtrees, hence there is an edge f of C with ends in different subtrees, i.e., it reconnects the subtrees into a tree T2 with weight less than that of T1, because the weight of f is less than the weight of e.

5. *Cut property:* This figure shows the cut property of MSP.T is the MST of the given graph. If S = {A,B,D,E}, thus V-S = {C,F},then there are 3 possibilities of the edge across the cut(S,V-S), they are edges BC, EC, EF of the original graph. Then, e is one of the minimum-weight-edge for the cut, therefore S □ {e} is part of the MST T.
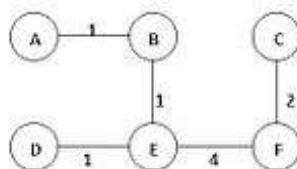
Figure 6.2

For any cut C in the graph, if the weight of an edge e of C is smaller than the weights of other edges of C, then this edge belongs to all MSTs of the graph. Indeed, assume the contrary, for example, edge BC (weighted 6) belongs to the MST T instead of edge e (weighted 4) in the left figure. Then adding e to T will produce a cycle, while replacing BC with e would produce MST of smaller weight.

5.   *Minimum-cost edge:* If the edge of a graph with the minimum cost e is unique, then this edge is included in any MST. Indeed, if e was not included in the MST, removing any of the (larger cost) edges in the cycle formed after adding e to the MST, would yield a spanning tree of smaller weight.

Task   Analyse if the weights are *positive*, then a minimum spanning tree is in fact the minimum-cost subgraph connecting all vertices. Comment

6.1.2 Algorithms

The first algorithm for finding a minimum spanning tree was developed by Czech scientist Otakar Borùvka in 1926 (see Borùvka's algorithm). Its purpose was an efficient electrical coverage of Moravia. There are now two algorithms commonly used, Prim's algorithm and Kruskal's algorithm. All three are greedy algorithms that run in polynomial time, so the problem of finding such trees is in FP, and related decision problems such as determining whether a particular edge is in the MST or determining if the minimum total weight exceeds a certain value are in P. Another greedy algorithm not as commonly used is the reverse-delete algorithm, which is the reverse of Kruskal's algorithm.

The fastest minimum spanning tree algorithm to date was developed by David Karger, Philip Klein, and Robert Tarjan, who found a linear time randomized algorithm based on a combination of Borùvka's algorithm and the reverse-delete algorithm. The fastest non-randomized algorithm, by Bernard Chazelle, is based on the soft heap, an approximate priority queue. Its running time is O(m á(m,n)), where m is the number of edges, n is the number of vertices and á is the classical

functional inverse of the Ackermann function. The function á grows extremely slowly, so that for all practical purposes it may be considered a constant no greater than 4; thus Chazelle's algorithm takes very close to linear time. If the edge weights are integers with a bounded bit length, then deterministic algorithms are known with linear running time. Whether there exists a deterministic algorithm with linear running time for general weights is still an open question. However, Seth Petie and Vijaya Ramachandran have found a provably optimal deterministic minimum spanning tree algorithm, the computational complexity of which is unknown.

More recently, research has focused on solving the minimum spanning tree problem in a highly parallelized manner. With a linear number of processors it is possible to solve the problem in $O(\log n)$ time. A 2003 paper "Fast Shared-Memory Algorithms for Computing the Minimum Spanning Forest of Sparse Graphs" by David A. Bader and Guojing Cong demonstrates a pragmatic algorithm that can compute MSTs 5 times faster on 8 processors than an optimized sequential algorithm. Typically, parallel algorithms are based on Boruvka's algorithm—Prim's and especially Kruskal's algorithm do not scale as well to additional processors.

Other specialized algorithms have been designed for computing minimum spanning trees of a graph so large that most of it must be stored on disk at all times. These external storage algorithms, for example as described in "Engineering an External Memory Minimum Spanning Tree Algorithm" by Roman Dementiev et al., can operate as little as 2 to 5 times slower than a traditional in-memory algorithm; they claim that "massive minimum spanning tree problems filling several hard disks can be solved overnight on a PC." They rely on efficient external storage sorting algorithms and on graph contraction techniques for reducing the graph's size efficiently.

The problem can also be approached in a distributed manner. If each node is considered a computer and no node knows anything except its own connected links, one can still calculate the distributed minimum spanning tree.

[edit] MST on complete graphs

Alan M. Frieze showed that given a complete graph on n vertices, with edge weights that are independent identically distributed random variables with distribution function F satisfying $F'(0) > 0$, then as n approaches $+\infty$ the expected weight of the MST approaches $\zeta(3) / F'(0)$, where ζ is the Riemann zeta function. Under the additional assumption of finite variance, Alan M. Frieze also proved convergence in probability. Subsequently, J. Michael Steele showed that the variance assumption could be dropped.

In later work, Svante Janson proved a central limit theorem for weight of the MST.

Did u know? **What is uniform random weights?**

The uniform random weights in the exact expected size of the minimum spanning tree has been computed for small complete graphs.

6.2 Prim's Algorithm
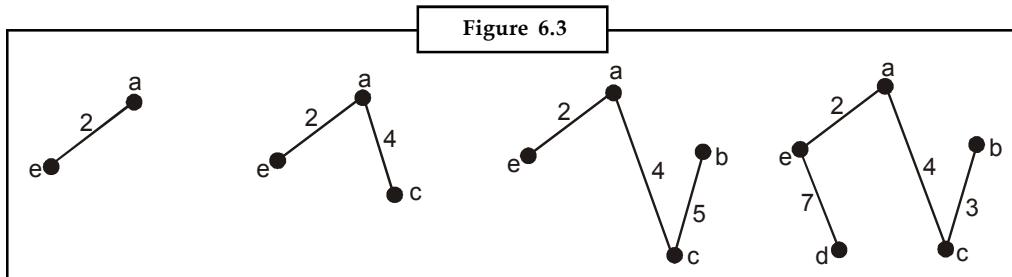
Start with a finite set of vertices, each pair joined by a weighted edge.

1.   Chose and draw any vertex

2.   Find the edge of least weight joining a drawn vertex to a vertex not currently drawn. Draw this weighted edge and the corresponding new vertex.
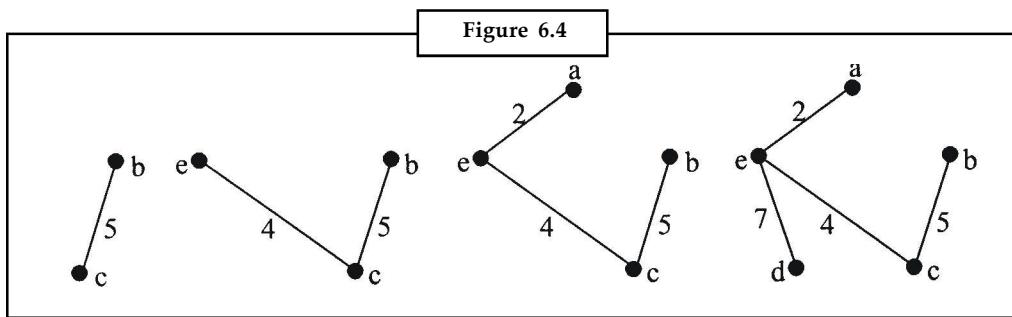
Repeat step 2 until all vertices are connected, then stop.

To construct a minimum spanning tree using Prim's algorithm we proceed thus. Choose vertex *a* to start, minimum weight edge is *ae* (weight 2), minimum weight edge from {*a,e*} to new vertex is *ac* (weight 4), minimum weight edge from {*a, c, e*} to new vertex is *cb* (weight 5), minimum weight edge from {*a, b, c, e*} to new vertex is *ed* (weight 7), all vertices now connected. Total weight is 18.

Figure 6.3

Of course the choice of a as starting what happens if we start somewhere else. Choose vertex *b* to start, minimum weight edge is *bc* (weight 5), minimum weight edge from {*b, c*} to new vertex is *ca* or *ce* (both weight 4), choose *ce*, minimum weight edge from {*b, c, e*} to new vertex is *ea* (weight 2), minimum weight edge from {*a, b, c, e*} to new vertex is *ed* (weight 7), all vertices now connected. Total weight is 18.

Figure 6.4

In this case we see that we get a different minimum spanning tree (if we had chosen *ca* instead of *ce* at stage 2 we would have got the same minimum spanning tree as previously). This is because, in this graph, there is more than one minimum spanning tree.

Now, the **travelling salesmen problem** is the problem of finding the minimum weight Hamiltonian cycles in a weighted complete graph.

No simple algorithm to solve the travelling salesmen problem is known. We can seek approximate solutions which provide upper and lower bounds for the solution of the problem. Obviously if we can find upper and lower bounds which are close enough together then we have a solution which may be good enough for practical purposes and the closer they are together, the better. So we will try to find the smallest upper bound possible and the largest lower bound.
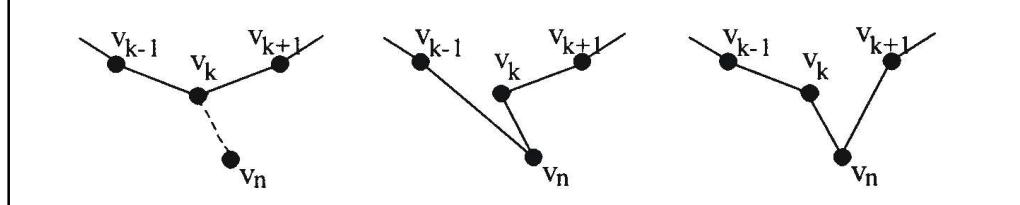
Heuristic algorithm for an upper bound on the travelling salesmen problem

1.   Start with a finite set of 3 or more vertices, each pair joined by a weight edge.

2.   Choose any vertex and find a vertex joined to it by an edge of least weight.

3.   Find the vertex which is joined to either of the two vertices identified in step (2) by an edge of least weight. Draw these three vertices and the three edges joining them to form a cycle.

4.   Find a vertex, not currently drawn, joined by an edge of least weight to any of the vertices already drawn. Label the new vertex v_n and the existing vertex to which it is joined by the edge of least weight as v_k. Label the two vertices which are joined to v_k in the existing cycle

as v_{k-1} and v_{k+1}. Denote the weight of an edge $v_p v_q$ by $w(v_p v_q)$. If $w(v_n v_{k-1}) - w(v_k v_{k-1}) <$ $w(v_n v_{k+1}) - w(v_k v_{k+1})$ then replace the edge $v_k v_{k-1}$ in the existing cycle with the edge $v_n v_{k-1}$ otherwise replace the edge $v_k v_{k+1}$ in the existing cycle with the edge $v_n v_{k+1}$. (See Figure 6.5).
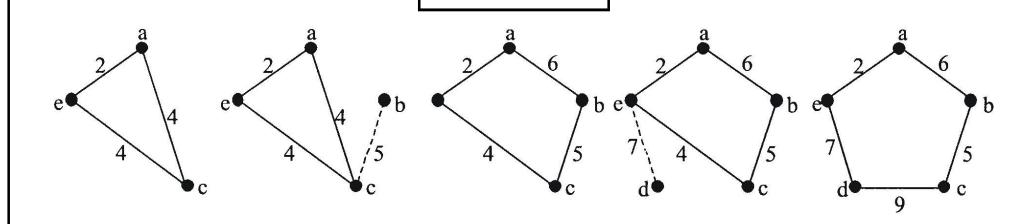
Figure 6.5

5. Repeat step 4 until all vertices are joined by a cycle, then stop. The cycle of obtained is an upper bound for the solution to the travelling salesmen problem.

Applying this to our previous problem we proceed thus. Start by choosing vertex *a*, then the edge of least weight incident to a is *ae* and the vertex adjacent to {*a*, *e*} joined by an edge of least weight is *c*. From the cycle *aeca*. Now the vertex adjacent to {*a*, *c*, *e*} joined by an edge of least weight is *b* (edge *bc*) and the vertices adjacent to *c* in the existing cycle are *a* and *e*. In this case $w(be) - w(ce) = w(ba) - w(ca)$ so we can create the new cycle by replacing either of edges *ca* or *ce*. We will choose *ca*. Now the vertex adjacent to {*a*, *b*, *c*, *e*} joined by an edge of least weight is *d* (edge *ed*) and the vertices adjacent to *e* in the existing cycle are *a* and *c*. Now $w(da) - w(ea) > w(dc) - w(ec)$ so we create the new cycle by replacing edge *ec* with edge *dc*. The cycle is *abcdea* and the weight is 29.

Figure 6.6

What would have happened if we had chosen a different starting vertex? Let us try it. Start by choosing vertex *b*, then the edge of least weight incident to *b* is *bc* and the vertex adjacent to {*b*, *c*} joined by an edge of least weight is *a* or *e*. Choose *a* and from the cycle *abca*. Now the vertex adjacent to {*a*, *b*, *c*} joined by an edge of least weight is *e* (edge *ae*) and the vertices adjacent to *a* in the existing cycle are *b* and *c*. In this case $w(eb) - w(ab) < w(ec)$ so we can create the new cycle by replacing edge *ab* with *eb*, creating the cycle *aebca*. Now the vertex adjacent to {*a*, *b*, *c*, *e*} joined by an edge of least weight is *d* (edge *de*) and the vertices adjacent to *e* in the existing cycle are *a* and *b*. Now $w(da) - w(ea) > w(db) - w(eb)$ so we create the new cycle by replacing edge *eb* with edge *db*. The cycle is *acbdea* and the weight is 26.

6.3 Shortest Path

In graph theory, the shortest path problem is the problem of finding a path between two vertices (or nodes) such that the sum of the weights of its constituent edges is minimized. An example is finding the quickest way to get from one location to another on a road map; in this case, the vertices represent locations and the edges represent segments of road and are weighted by the time needed to travel that segment.

Formally, given a weighted graph (that is, a set V of vertices, a set E of edges, and a real-valued weight function f : E → R), and one element v of V, find a path P from v to a v′ of V so that

$$\sum_{p \in P} f(p)$$

is minimal among all paths connecting v to $v′$.

The problem is also sometimes called the **single-pair shortest path problem**, to distinguish it from the following generalizations:
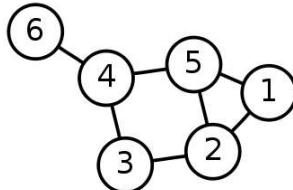
1. The **single-source shortest path problem**, in which we have to find shortest paths from a source vertex v to all other vertices in the graph.

2. The **single-destination shortest path problem**, in which we have to find shortest paths from all vertices in the graph to a single destination vertex v. This can be reduced to the single-source shortest path problem by reversing the edges in the graph.

3. The **all-pairs shortest path problem**, in which we have to find shortest paths between every pair of vertices v, $v′$ in the graph.

Notes   These generalizations have significantly more efficient algorithms than the simplistic approach of running a single-pair shortest path algorithm on all relevant pairs of vertices.

Figure 6.7

6.3.1 Applications

Shortest path algorithms are applied to automatically find directions between physical locations, such as driving directions on web mapping websites like Mapquest or Google Maps. For this application fast specialized algorithms are available.

If one represents a nondeterministic abstract machine as a graph where vertices describe states and edges describe possible transitions, shortest path algorithms can be used to find an optimal sequence of choices to reach a certain goal state, or to establish lower bounds on the time needed to reach a given state. For example, if vertices represents the states of a puzzle like a Rubik's Cube and each directed edge corresponds to a single move or turn, shortest path algorithms can be used to find a solution that uses the minimum possible number of moves.

In a networking or telecommunications mindset, this shortest path problem is sometimes called the min-delay path problem and usually tied with a widest path problem. For example, the algorithm may seek the shortest (min-delay) widest path, or widest shortest (min-delay) path.

A more lighthearted application is the games of "six degrees of separation" that try to find the shortest path in graphs like movie stars appearing in the same film.

Did u know? **You the other applications of the shortest path?**

Other applications include "operations research, plant and facility layout, robotics, transportation, and VLSI design".

6.3.2 Linear Programming Formulation

There is a natural linear programming formulation for the shortest path problem, given below. It is very trivial compared to most other uses of linear programs in discrete optimization, however it illustrates connections to other concepts.

Given a directed graph (V, A) with source node s, target node t, and cost w_{ij} for each arc (i, j) in A, consider the program with variables x_{ij}

$$\text{minimize} \sum_{ij \in A} w_{ij} x_{ij} \text{ subject to } x \geq 0 \text{ and for all } i,$$

$$\sum_{j} x_{ij} - \sum_{j} x_{ji} = \begin{cases} 1, & \text{if } i = s; \\ -1, & \text{if } i = t; \\ 0, & \text{otherwise} \end{cases}$$

This LP, which is common fodder for operations research courses, has the special property that it is integral; more specifically, every basic optimal solution (when one exists) has all variables equal to 0 or 1, and the set of edges whose variables equal 1 form an *s-t* dipath. See Ahuja et al. for one proof, although the origin of this approach dates back to mid-20th century.

The dual for this linear program is

maximize y_t - y_s subject to for all *ij*, $y_j - y_i \leq w_{ij}$

and feasible duals correspond to the concept of a consistent heuristic for the A* algorithm for shortest paths. For any feasible dual y the reduced costs $w'_{ij} = w_{ij} - y_j + y_i$ are nonnegative and A* essentially runs Dijkstra's algorithm on these reduced costs.

6.4 Dijkstra's Algorithm
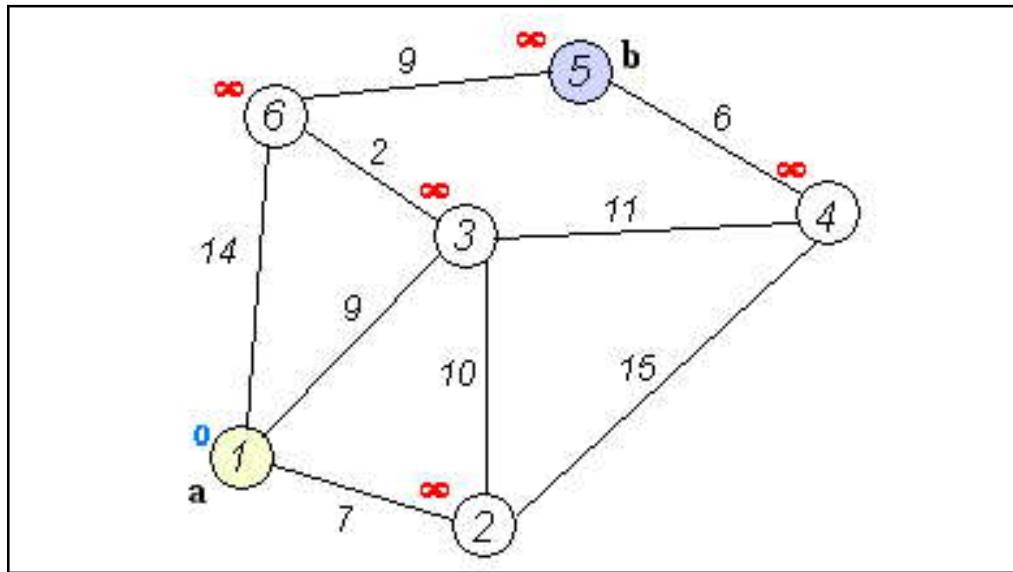
Dijkstra's algorithm, conceived by Dutch computer scientist Edsger Dijkstra in 1956 and published in 1959, is a graph search algorithm that solves the single-source shortest path problem for a graph with nonnegative edge path costs, producing a shortest path tree. This algorithm is often used in routing and as a subroutine in other graph algorithms.

For a given source vertex (node) in the graph, the algorithm finds the path with lowest cost (i.e. the shortest path) between that vertex and every other vertex. It can also be used for finding costs of shortest paths from a single vertex to a single destination vertex by stopping the algorithm once the shortest path to the destination vertex has been determined. For example, if the vertices of the graph represent cities and edge path costs represent driving distances between pairs of cities connected by a direct road, Dijkstra's algorithm can be used to find the shortest route between one city and all other cities. As a result, the shortest path first is widely used in network routing protocols, most notably IS-IS and OSPF (Open Shortest Path First).

Dijkstra's original algorithm does not use a min-priority queue and runs in $O(|V|^2)$. The idea of this algorithm is also given in (Leyzorek et al. 1957). The common implementation based on a min-priority queue implemented by a Fibonacci heap and running in $O(|E| + |V| \log |V|)$ is

due to (Fredman & Tarjan 1984). This is asymptotically the fastest known single-source shortest-path algorithm for arbitrary directed graphs with unbounded nonnegative weights.

Algorithm

Let the node at which we are starting be called the **initial node**. Let the **distance of node Y** be the distance from the **initial node** to Y. Dijkstra's algorithm will assign some initial distance values and will try to improve them step by step.

1. Assign to every node a distance value: set it to zero for our initial node and to infinity for all other nodes.

2. Mark all nodes as unvisited. Set initial node as current.

3. For current node, consider all its unvisited neighbors and calculate their *tentative* distance. For example, if current node (A) has distance of 6, and an edge connecting it with another node (B) is 2, the distance to B through A will be 6+2=8. If this distance is less than the previously recorded distance, overwrite the distance.

4. When we are done considering all neighbors of the current node, mark it as visited. A visited node will not be checked ever again; its distance recorded now is final and minimal.

5. If all nodes have been visited, finish. Otherwise, set the unvisited node with the smallest distance (from the initial node, considering all nodes in graph) as the next "current node" and continue from step 3.

Description

Notes   For ease of understanding, this discussion uses the terms **intersection**, **road** and **map** — however, formally these terms are **vertex**, **edge** and **graph**, respectively.

Suppose you want to find the shortest path between two intersections on a city map, a starting point and a destination. The order is conceptually simple: to start, mark the distance to every intersection on the map with infinity. This is done not to imply there is an infinite distance, but to note that that intersection has not yet been *visited*. (Some variants of this method simply leave the intersection unlabeled.) Now, at each iteration, select a *current* intersection. For the first iteration the current intersection will be the starting point and the distance to it (the intersection's label) will be zero. For subsequent iterations (after the first) the current intersection will be the closest unvisited intersection to the starting point—this will be easy to find.

From the current intersection, update the distance to every unvisited intersection that is directly connected to it. This is done by determining the sum of the distance between an unvisited intersection and the value of the current intersection, and relabeling the unvisited intersection with this value if it is less than its current value. In effect, the intersection is relabeled if the path to it through the current intersection is shorter than the previously known paths. To facilitate shortest path identification, in pencil, mark the road with an arrow pointing to the relabeled intersection if you label/relabel it, and erase all others pointing to it. After you have updated the distances to each neighboring intersection, mark the current intersection as *visited* and select the unvisited intersection with lowest distance (from the starting point) — or lowest label—as the current intersection. Nodes marked as visited are labeled with the shortest path from the starting point to it and will not be revisited or returned to.

Continue this process of updating the neighboring intersections with the shortest distances, then marking the current intersection as visited and moving onto the closest unvisited intersection until you have marked the destination as visited. Once you have marked the destination as visited (as is the case with any visited intersection) you have determined the shortest path to it, from the starting point, and can trace your way back, following the arrows in reverse.

In the accompanying animated graphic, the starting and destination intersections are colored in light pink and blue and labelled a and b respectively. The visited intersections are colored in red, and the current intersection in a pale blue.

Of note is the fact that this algorithm makes no attempt to direct "exploration" towards the destination as one might expect. Rather, the sole consideration in determining the next "current" intersection is its distance from the starting point. In some sense, this algorithm "expands outward" from the starting point, iteratively considering every node that is closer in terms of shortest path distance until it reaches the destination. When understood in this way, it is clear how the algorithm necessarily finds the shortest path, however it may also reveal one of the algorithm's weaknesses: its relative slowness in some topologies.

Pseudocode

In the following algorithm, the code u := vertex in *Q* with smallest dist[], searches for the vertex *u* in the vertex set *Q* that has the least dist[*u*] value. That vertex is removed from the set *Q* and returned to the user. dist_between(*u*, *v*) calculates the length between the two neighbor-nodes *u* and *v*. The variable *alt* on line 15 is the length of the path from the root node to the neighbor node *v* if it were to go through *u*. If this path is shorter than the current shortest path recorded for *v*, that current path is replaced with this *alt* path. The previous array is populated with a pointer to the "next-hop" node on the source graph to get the shortest route to the source.

```
1    function Dijkstra(Graph, source):

2    for each vertex v in Graph: // Initializations

3    dist[v] := infinity ; // Unknown distance function from source to v

4    previous[v] := undefined ; // Previous node in optimal path from source

5    end for ;
```

```
6    dist[source] := 0 ; // Distance from source to source

7    Q := the set of all nodes in Graph ;

     // All nodes in the graph are unoptimized – thus are in Q

8    while Q is not empty: // The main loop

9    u := vertex in Q with smallest dist[] ;

10   if dist[u] = infinity:

11   break ; // all remaining vertices are inaccessible from source

12   fi ;

13   remove u from Q ;

14   for each neighbor v of u: // where v has not yet been removed from Q.

15   alt := dist[u] + dist_between(u, v) ;

16   if alt < dist[v]: // Relax (u,v,a)

17   dist[v] := alt ;

18   previous[v] := u ;

19   fi ;

20   end for ;

21   end while ;

22   return dist[] ;

23 end Dijkstra.
```

If we are only interested in a shortest path between vertices *source* and *target*, we can terminate the search at line 13 if *u = target*. Now we can read the shortest path from *source* to *target* by iteration:

```
1    S := empty sequence

2    u := target

3    while previous[u] is defined:

4        insert u at the beginning of S

5        u := previous[u]
```

Now sequence *S* is the list of vertices constituting one of the shortest paths from *target* to *source*, or the empty sequence if no path exists.

A more general problem would be to find all the shortest paths between *source* and *target* (there might be several different ones of the same length). Then instead of storing only a single node in each entry of previous[] we would store all nodes satisfying the relaxation condition. For example, if both *r* and *source* connect to *target* and both of them lie on different shortest paths through *target* (because the edge cost is the same in both cases), then we would add both *r* and *source* to previous[*target*]. When the algorithm completes, previous[] data structure will actually describe a graph that is a subset of the original graph with some edges removed. Its key property will be that if the algorithm was run with some starting node, then every path from that node to any other node in the new graph will be the shortest path between those nodes in the original graph, and all paths of that length from the original graph will be present in the new graph. Then to actually find all these short paths between two given nodes we would use a path finding algorithm on the new graph, such as depth-first search.

Running Time

An upper bound of the running time of Dijkstra's algorithm on a graph with edges *E* and vertices *V* can be expressed as a function of $|E|$ and $|V|$ using Big-O notation.

For any implementation of set Q the running time is $O\left(|E|.dk_Q + |V|.em_Q\right)$, where dk_Q and em_Q are times needed to perform decrease key and extract minimum operations in set Q, respectively.

The simplest implementation of the Dijkstra's algorithm stores vertices of set Q in an ordinary linked list or array, and extract minimum from Q is simply a linear search through all vertices in Q. In this case, the running time is $O(|V|^2 + |E|) = O(|V|^2)$.

For sparse graphs, that is, graphs with far fewer than $O(|V|^2)$ edges, Dijkstra's algorithm can be implemented more efficiently by storing the graph in the form of adjacency lists and using a binary heap, pairing heap, or Fibonacci heap as a priority queue to implement extracting minimum efficiently. With a binary heap, the algorithm requires $O((|E| + |V|)\log|V|)$ time (which is dominated by $O(|E|\ log\ |V|)$, assuming the graph is connected), and the Fibonacci heap improves this to $O(|E| + |V|\log|V|)$.

⚠️

Caution Directed acyclic graphs, it is possible to find shortest paths from a given starting vertex in linear time, by processing the vertices in a topological order, and calculating the path length for each vertex to be the minimum or maximum length obtained via any of its incoming edges.

Related Problems and Algorithms

The functionality of Dijkstra's original algorithm can be extended with a variety of modifications.

Example: Sometimes it is desirable to present solutions which are less than mathematically optimal.

To obtain a ranked list of less-than-optimal solutions, the optimal solution is first calculated. A single edge appearing in the optimal solution is removed from the graph, and the optimum solution to this new graph is calculated. Each edge of the original solution is suppressed in turn and a new shortest-path calculated. The secondary solutions are then ranked and presented after the first optimal solution.

Dijkstra's algorithm is usually the working principle behind link-state routing protocols, OSPF and IS-IS being the most common ones.

Unlike Dijkstra's algorithm, the Bellman-Ford algorithm can be used on graphs with negative edge weights, as long as the graph contains no negative cycle reachable from the source vertex s. (The presence of such cycles means there is no shortest path, since the total weight becomes lower each time the cycle is traversed.)

The A* algorithm is a generalization of Dijkstra's algorithm that cuts down on the size of the subgraph that must be explored, if additional information is available that provides a lower bound on the "distance" to the target. This approach can be viewed from the perspective of linear programming: there is a natural linear program for computing shortest paths, and solutions to its dual linear program are feasible if and only if they form a consistent heuristic (speaking roughly, since the sign conventions differ from place to place in the literature). This feasible dual/consistent heuristic defines a nonnegative reduced cost and A* is essentially running Dijkstra's algorithm with these reduced costs. If the dual satisfies the weaker condition of admissibility, then A* is instead more akin to the Bellman-Ford algorithm.

> *Notes*   The process that underlies Dijkstra's algorithm is similar to the greedy process used in Prim's algorithm. Prim's purpose is to find a minimum spanning tree for a graph.

6.5 Summary

- A *binary tree* is a rooted tree in which each vertex has at most two children, designated as *left child* and *right child.* If a vertex has one child, that child is designated as either a left child or a right child, but not both.

- In graph theory, the shortest path problem is the problem of finding a path between two vertices (or nodes) such that the sum of the weights of its constituent edges is minimized.

- **Dijkstra's algorithm**, conceived by Dutch computer scientist Edsger Dijkstra in 1956 and published in 1959, is a graph search algorithm that solves the single-source shortest path problem for a graph with nonnegative edge path costs, producing a shortest path tree.

- A more general problem would be to find all the shortest paths between *source* and *target*

- Continue this process of updating the neighboring intersections with the shortest distances, then marking the current intersection as visited and moving onto the closest unvisited intersection until you have marked the destination as visited.

6.6 Keywords

All-pairs Shortest path Problem: In which we have to find shortest paths between every pair of vertices v, v' in the graph.

Dijkstra's algorithm: It is conceived by Dutch computer scientist Edsger Dijkstra in 1956.

Shortest path algorithms: These are applied to automatically find directions between physical locations, such as driving directions on web mapping websites like Mapquest or Google Maps.

Single-source Shortest path Problem: In which we have to find shortest paths from a source vertex v to all other vertices in the graph.

Single-destination Shortest path Problem: In which we have to find shortest paths from all vertices in the graph to a single destination vertex v. This can be reduced to the single-source shortest path problem by reversing the edges in the graph.

6.7 Self Assessment
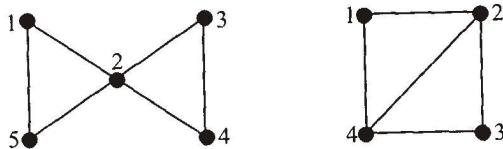
Fill in the blanks:

1. A minimum spanning tree (MST) or minimum weight spanning tree is then a spanning tree with .... than or equal to the weight of every other spanning tree.

2. If each .... has a distinct weight then there will only be one, unique minimum spanning tree.

3. If the edge of a graph with the minimum cost e is unique, then this edge is included in any ....

4. The .... is the problem of finding a path between two vertices (or nodes) such that the sum of the weights of its constituent edges is minimized.

5.   In a networking or telecommunications mindset, this shortest path problem is sometimes called the .... problem and usually tied with a widest path problem.

6.   Dijkstra's original algorithm does not use a .... and runs.

7.   Shortest path algorithms are applied to .... find directions between physical locations.

6.8 Review Questions

1.   Find all the spanning trees in each of the two graphs shown.

   Hint: How many edges does a tree with 5 vertices have?

2.   (a)   How many spanning trees has $K_{2,3}$?

   (b)   How many spanning trees has $K_{2,100}$?

   (c)   How many spanning trees has $K_{2,n}$?

3.   If G is a simple, connected, planar graph with $n(\geq 3)$ vertices and m edges, and if g is the length of the shortest cycle in G, show that

$$m \leq g(n - 2)/(g - 2)$$

   Hint: The edges around a face in a plane drawing of a planar graph are a cycle. Find a lower bound on the face degree sum of G then use the Handshaking Lemma and Euler's formula.

4.   The following tables gives the distances between 6 European cities (in hundreds of km). Find an upper bound on the solution of the travelling salesman problem for these cities starting at (a) Moscow and (b) Seville. Which is the better bound? Find a lower bound to the solution by removing (a) Moscow and (b) Seville. Which is the better lower bound?

	Berlin	London	Moscow	Paris	Rome	Seville
Berlin		8	11	7	10	15
London	8	—	18	3	12	11
Moscow	11	18	—	19	20	27
Paris	7	3	19	—	9	8
Rome	10	12	20	9	—	13
Seville	15	11	27	8	13	—

5. The following table gives the distances between English cities (in miles). Find an upper bound on the solution of the travelling salesman problem for these cities starting at (a) Bristol and (b) Leeds. Which is the better bound? Find a lower bound to the solution by removing (a) Exeter and (b) York. Which is the better lower bound?

	Bristol	Exeter	Hull	Leeds	Oxford	York
Bristol	—	84	231	220	74	225
Exeter	84	—	305	271	154	280
Hull	231	305	—	61	189	37
Leeds	220	271	61	—	169	24
Oxford	74	154	189	169	—	183
York	225	280	37	24	183	—

Answers: Self Assessment

1. weight less
2. edge
3. MST
4. shortest path problem
5. min-delay path
6. min-priority queue
7. automatically

6.9 Further Readings

Books

Béla Bollobás, *Modern Graph Theory*, Springer

Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph Theory, Combinatorics, and Algorithms*, Birkhäuser

Online links

http://en.wikipedia.org/wiki/File:Tree_graph.svg

http://cslibrary.stanford.edu/110/BinaryTrees.html

Unit 7: Boolean Algebra

CONTENTS

Objectives

Introduction

Objectives

After studying this unit, you will be able to:

- Define propositions
- Find compound propositions
- Describe basic logic operations

Introduction

We have already dealt extensively with the algebra of sets and mode mention of a related logical arithmetic. Now we will discuss Boolean algebra. Set algebra is a particular Boolean algebra and although different Boolean algebra are structurally very similar.

1.   Developed by famous mathematician **George Boole**.

2.   Useful in analysis, design of electronic computers, dial telephone switching system and many kind of electronic control devices.

7.1 Propositons

Well-defined Formula of Proposition

"A proposition is a statement that is either true or false but not both".

Example:

1. "Washington D.C. is the capital of the United States of America"

2. Toronto is the capital of Canada

3. 1 + 1 = 2   (a) Delhi is the capital of India

4. 2 + 2 = 3   (b) Haryana is a country

5. What time is it?   (c) 7 is a prime number.

6. Real this carefully (d) 2 + 3 = 5

7. x + 1 = 2   (e) Wish you a happy new year.

8. x + 7 = z   (f) How are you?

                         (g) Please wait.

                         (h) Take one disprin.

S → Statements 1 and 3 are true, whereas 2 and 4 are false, so they are propositions. Sentences 1 and 2 are not propositions because they are not statements. Sentences 3 and 4 are not propositions because they are neither true nor false.

Did u know? **What is axiom?**

An axiom is a proposition that is assumed to be true. With sufficient information, mathematical logic can often categorize a proposition as true or false, although there are various exceptions (e.g., "This statement is false").

7.2 Compound Propositions

Compound propositions involve the assembly of multiple statements, using multiple operators.

Writing Truth Tables for Compound Propositions

To write the truth table for a compound proposition, it's best to calculate the statement's truth value after each individual operator. For example, in the statement $pV\neg q \to q$, it's best to solve for $\neg q$, then for $pV\neg q$, and finally for the statement as a whole:

1. p = (T, T, F, F); q = (T, F, T, F)

2. p = (T, T, F, F); $\neg q$ = (F, T, F, T)

3. $pV\neg q$ = (T, T, F, T)

   q = (T, F, T, F)

1. $pV\neg q \to q$ = (T, F, T, F)

The Contrapositive, Inverse and Converse

The **contrapositive** of conditional statement $p \rightarrow q$ is $\neg q \rightarrow \neg p$. **A conditional is logically equivalent to its contrapositive.** In other words, if q did not occur, then we can assume p also did not occur.

The **inverse** is $\neg p \rightarrow \neg q$.

The **converse** is $q \rightarrow p$.

Task "The **converse** and **inverse** are logically equivalent." Comment

---

7.3 Basic Logical Operations

At first glance, it may not seem that the study of logic should be part of mathematics. For most of us, the word logic is associated with reasoning in a very nebulous way:

"If my car is out of gas, then I cannot drive it to work."

seems logical enough, while

"If I am curious, then I am yellow."

is clearly illogical. Yet our conclusions about what is or is not logical are most often unstructured and subjective. The purpose of logic is to enable the logician to construct valid arguments which satisfy the basic principle

"If all of the premises are true, then the conclusion must be true."

It turns out that in order to reliably and objectively construct valid arguments, the logical operations which one uses must be clearly defined and must obey a set of consistent properties. Thus logic is quite rightly treated as a mathematical subject.

Up until now, you've probably considered mathematics as a set of rules for using numbers. The study of logic as a branch of mathematics will require you to think more abstractly then you are perhaps used to doing. For instance, in logic we use variables to represent **propositions** (or premises), in the same fashion that we use variables to represent numbers in algebra. But while an algebraic variable can have any number as its value, a logical variable can only have the value **True** or **False**. That is, True and False are the "numerical constants" of logic. And instead of the usual arithmetic operators (addition, subtraction, etc.), the logical operators are "AND", "OR", "NOT", "XOR" ("eXclusive OR"), "IMPLIES" and "EQUIVALENCE". Finally, rather than constructing a series of algebraic steps in order to solve a problem,

Caution You will learn how to determine whether a statement is always true (a **tautology**) or is a **contradiction** (never true), and whether an argument is valid.

7.3.1 Truth Tables

In algebra, it is rarely possible to guess the numerical solution to a problem, and because there are an infinite number of numbers it is obvious that one cannot try all possible solutions in order to find one that solves the problem. But in logic, we only have two "numbers": True and False. Therefore, any logical statement which contains a finite number of logical variables

(which of course covers any problem we have to deal with) can be analyzed using a table which lists all possible values of the variables: a "**truth table**". Since each variable can take only two values, a statement with "n" variables requires a table with 2^n rows. Using the letters "p", "q", "r", etc., to represent logical variables, we can construct truth tables for statements involving any number of variables (although we will usually limit ourselves to at most three variables per statement to simplify the matter):

p
T
F

for statements with one variable,

p	q
T	T
T	F
F	T
F	F

for statements with two variables and

p	q	r
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

For statements with three variables (where in every case, "T" stands for True and "F" for False). The extension to more than three variables should now be obvious:

1.   For the first variable, the first half of the rows are T while the second half are F

2.   For the second variable, the rows are split into four sections: the first and third quarters are T while the second and fourth quarters are F

3.   For the third variable, the rows are split into eighths, with alternating eighths having T's and F's

4.   In general, for the nth variable, the rows are split into 2^n parts, with alternating T's and F's in each part

7.3.2 Logical Operators

We will now define the logical operators which we mentioned earlier, using truth tables. But let us proceed with caution: most of the operators have names which we may be accustomed to using in ways that are fuzzy or even contradictory to their proper definitions. In all cases, use the truth table for an operator as its exact and only definition; try not to bring to logic the baggage of your colloquial use of the English language.

The first logical operator which we will discuss is the "**AND**", or **conjunction** operator. For the computer scientist, it is perhaps the most useful logical operator we will discuss. It is a "**binary**" operator (a binary operator is defined as an operator that takes two operands; not binary in the sense of the binary number system):

p AND q

It is traditionally represented using the following symbol:

but we will represent it using the ampersand ("**&**") since that is the symbol most commonly used on computers to represent a logical AND. It has the following truth table:

p	q	p & q
T	T	T
T	F	F
F	T	F
F	F	F

Notice that p & q is only T if both p and q are T. Thus the rigorous definition of AND is consistent with its colloquial definition. This will be very useful for us when we get to Boolean Algebra: there, we will use 1 in place of T and 0 in place of F, and the AND operator will be used to "**mask**" bits.

Perhaps the quintessential example of masking which you will encounter in your further studies is the use of the "**network mask**" in networking. An IP ("Internet Protocol") address is 32 bits long, and the first n bits are usually used to denote the "network address", while the remaining 32 - n bits denote the "host address":

n bits	32 - n bits
Network Address	Host Address

Suppose that on your network, the three most significant bits in the first byte of an IP address denote the network address, while the remaining 29 bits of the address are used for the host. To find the network address, we can AND the first byte with

$1 1 1 0 0 0 0 0_2$

since

x x x y y y y y

 &

1 1 1 0 0 0 0 0

 =

x x x 0 0 0 0 0

(x & 1 = x, but x & 0 = 0). Thus masking allows the system to separate the network address from the host address in order to identify which network information is to be sent to. Note that most network numbers have more than 3 bits. You will spend a lot of time working with network masks in your courses on networking.

The **OR** (or **disjunction**) operator is also a binary operator, and is traditionally represented using the following symbol:

We will represent OR using the stroke ("**|**"), again due to common usage on computers. It has the following truth table:

p	q	p | q
T	T	T
T	F	T
F	T	T
F	F	F

p | q is true whenever either p is true, q is true or both p and q are true (so it too agrees with its colloquial counterpart).

The **NOT** (or **negation** or **inversion**) operator is a "**unary**" operator: it takes just one operand, like the unary minus in arithmetic (for instance, -x). NOT is traditionally represented using either the tilde ("~") or the following symbol:

In a programming environment, NOT is frequently represented using the exclamation point ("**!**"). Since the exclamation point is too easy to mistake for the stroke, we will use the tilde instead. Not has the following truth table:

p	~ p
T	F
F	T

~ p is the negation of p, so it again agrees with its colloquial counterpart; it is essentially the 1's complement operation.

The **XOR** (eXclusive OR) operator is a binary operator, and is not independent of the operators we have presented thus far (many texts do not introduce it as a separate logical operator). It has no traditional notation, and is not often used in programming (where our usual logical operator symbols originate), so we will simply adopt the "**X**" as the symbol for the XOR:

p	q	p X q
T	T	F
T	F	T
F	T	T
F	F	F

p X q is T if either p is T or q is T, **but not both**. We will see later how to write it in terms of ANDs, ORs and NOTs.

The implication operator (**IMPLIES**) is a binary operator, and is defined in a somewhat counterintuitive manner (until you appreciate it, that is!). It is traditional notated by one of the following symbols:

but we will denote it with an arrow ("**->**"):

p	q	p -> q
T	T	T
T	F	F
F	T	T
F	F	T

So p -> q follows the following reasoning:

1.   a True premise implies a True conclusion, therefore T -> T is T;

2.   a True premise cannot imply a False conclusion, therefore T -> F is F; and

3.   you can conclude anything from a false assumption, so F **->** anything is T.

IMPLIES (**implication**) is definitely one to watch; while its definition makes sense (after a bit of thought), it is probably not what you are used to thinking of as implication.

EQUIVALENCE is our final logical operator; it is a binary operator, and is traditionally notated by either an equal sign, a three-lined equal sign or a double arrow (" **<->**"):

p	q	p <-> q
T	T	T
T	F	F
F	T	F
F	F	T

p <-> q is T if p and q are the same (are equal), so it too follows our colloquial notion.

Just as with arithmetic operators, the logical operators follow "**operator precedence**" (an implicit ordering). In an arithmetic expression with sums and products and no parentheses, the multiplications are performed before the additions. In a similar fashion, if parentheses are not used, the operator precedence for logical operators is:

1.   First do the NOTs;

2.   then do the ANDs;

3.   then the ORs and XORs, and finally

4.   do the IMPLIES and EQUIVALENCEs.

Notes   A function in *sum of products* form can be implemented using NAND gates by replacing all AND and OR gates by NAND gates.

A function in *product of sums* form can be implemented using NOR gates by replacing all AND and OR gates by NOR gates.

A non empty set B together with two binary operations *t*, (known as addition and multiplication) and an separation *t*, *i* (complementation) on B is said to be Boolean algebra if it satisfied the following Axioms.

B₁ Commutativity : The operations are commutative *i.e.,*

$a + b = b + a$ and $a - b = b - a \; \forall \; ab \in B$

B_2 *Distributivity :* Each binary operations distributes over the other i.e.,

$$a + (b.\, c) \;=\; (a + b)\,.\,(a + c)$$

$$a\,.\,(b + c) \;=\; (a.\, b) + (a\,.\,c)$$

B_3 *Identity :* B contains distinct identity elements 0 and 1 with respect to the operation + and Respectively *i.e.,*

$$a + 0 \;=\; a,\, a.1 = a \;\forall\; a\; \varepsilon \; B.$$

B_4 *Complementation :* For every $a \in B$ there exist on element $a^1 \in B$ such that

$$a + a^1 \;=\; 1 \text{ and } aa^1 = 0$$

* A boolean algebra is generally denoted by tuples

$$(B, +, 0, ', 0, 1) \text{ or } B\ (B, +, 0, ') \text{ or by } B.$$

* Instead of binary operations + and 0 we may use other symbols such as $\cup, \cap$ (Union Intersection) or $\wedge, \vee$ (join muss to denote these operations.)

Boolean Sub Algebra

A boolean sub algebra is a non empty subset S of a boolean algebra B such that $a, b \in S \Rightarrow (a + b)$, $a\,.\,b,\, a^1 \in S$.

Example: Let B $\bullet$ (0, *a, b,* 1) Define +, $\bullet$ and ' by the tables given below :

t	0	a	b	1
0	0	0	0	0
a	0	a	0	a
b	0	0	b	b
1	0	a	b	1

$\bullet$	0	a	b	1
0	0	a	b	1
a	a	a	1	1
b	b	1	b	1
1	1	1	1	1

1	
0	1
a	b
b	a
1	0

Then B forms a boolean algebra under these operations

Example: Let B = [*p, q, r, s*] be a set addition and multiplication operations are defined on B as per table given below :

t	p	q	r	s
p	p	q	q	p
q	q	q	q	q
r	q	q	r	r
s	p	q	r	s

$\bullet$	p	q	r	s
p	p	s	s	s
p	r	r	r	s
r	s	r	r	s
s	p	s	s	s

1.   **+ and** $\bullet$ are binary operations since each element of the table is from the given set B.

2.   *Commutativity :* From the table

   (i) $p + q = q + p$

   (ii) $p.q = q.\, p \qquad \forall\; pq \in B$

3.   *Identity element:* From table it is clear that s is the additive identity

4. *Distributivity:*

   (a) $p \cdot (q + r) = p.q = p$

   and $p.q + p.r = p + s = p$

   $\Rightarrow$ $p(q + r) = p.q + p.r \; \forall \; p, q, r, s \in B.$

   (b) We can also see that

   $p + qr = (p + q)(p + r) \; \forall \; p, q, r, \in B$

5. *Complement:* From table it can be verified that each element has get its complement such that $p + r = q$, $P \cdot r = S \Rightarrow r$ is the complement of S. Hence B is a boolean algebra.

Example: Let B be a set of positive integers being divisions of 30 and operations $\vee, \wedge$ on it are defined as

$a \vee b$ = c where c is the LCM of a, b.

$a \wedge b$ = d where d is the HCF of a, b. $\forall \; a, b, c, d \; \varepsilon \; B$

then show that B is a Boolean algebra.

B = (1, 2, 3, 5, 6, 10, 15, 30)

$\vee$	1	2	3	5	6	10	15	30
1	1	2	3	5	6	10	15	30
2	2	2	6	10	6	10	30	30
3	3	6	3	15	6	30	15	30
5	5	10	15	5	30	10	15	30
6	6	6	6	30	6	30	30	30
10	1	2	3	10	30	10	30	30
15	15	30	15	15	30	30	15	30
30	30	30	30	30	30	30	30	30

$\wedge$	1	2	3	5	6	10	15	30
1	1	1	1	1	1	1	1	
2	1	2	1	1	2	2	1	2
3	1	1	3	1	3	1	3	3
5	1	1	1	5	1	5	5	5
6	1	2	3	1	6	2	3	6
10	1	2	1	3	2	10	5	10
15	1	1	3	5	3	5	13	13
30	2	2	3	5	6	10	15	30

7.4 Theorems of Boolean Algebra

Theorem 1: In a boolean algebra

1. Additive identity is unique.

2. Multiplicative identity is unique.

3. Complement of every element is unique.

Proof: Let (B, $\bullet$ +) is a boolean algebra.

1. Let if possible in B i_1 and i_2 be two additive identities then

When i_1 is additive identity $\Rightarrow a + i_1 = a$

and when i_2 is additive identity $\Rightarrow a + i_2 = a$

Hence, $\quad i_1 + i_2 = \quad i_2$ [Let i_1 is identity]

$\quad i_2 + i_1 = \quad i_1$ [Let i_2 is identity]

Since, $\quad i_1 + i_2 = \quad i_2 + i$ $\Rightarrow i_i = i_2$

2.   Suppose it is possible e_1 and e_2 be two multiplicative identity in B.

$$e_1 a = a \text{ and } e_2\, a = a$$

Also, $e_1\, e_2 = e_2$ [When e_1 is identity]

$e_2\, e_1 = e_1$ [When e_2 is identity]

Since, $e_1\, e_2 = e_2\, e_1$

$\Rightarrow$ $e_1 = e_2$

3.   Let if possible for $a \in B$. We have two different complement element b and c in b then

$$a.b = 0 \text{ and } a.c = 0$$

$$a + b = 1 \text{ and } a + c = 1$$

$$b = b + 0$$

$$= b + ac$$

$$= (b + a)\,(b + c)$$

$$b = 1 \cdot (b + c)$$

$$b = b + c \qquad\qquad\qquad …(1)$$

Similarly, $C = c + 0$

$$= c + ab$$

$$= (c + a)\,(c + b)$$

$$= c + b$$

$$c = b + c \qquad\qquad\qquad …(2)$$

(1) and (2) $\Rightarrow$ $b = c$

Theorem 2

Idempotent law:

1.   If a be an element of a boolean Algebra then

(i) $a + a = a$

(ii) $a \cdot a = a$

$$= [a + (b + c)]\, c + [a + (b + c)]\, c$$

$$= c\,[a + b(b + c)] + c_1\,[a + (b + c)]$$

$$= [ca + c\,(b + c)] + [c_1\, a + c_1\,(b + c)]$$

$$= (c\, a + c) +$$

2.   If a and b are arbitrary elements of a boolean algebra B then

(i) $(a + b)^1 = a^1\, b^1,$

(ii) $(ab)^1 = a^1 + b^1$

Proof: (*i*) Consider

$$(a + b) + (a^1\, b^1) = (aq + b + a^1)\,(a + b + b^1) = (a + a^1 + b)\,(a + b + b^1)$$

$$= (1 + b)\,(a + 1) = 1.1 = 1 \qquad\qquad\qquad …(3)$$

Again $(a + b) \cdot (a^1 b^1)$ $=$ $(a^1 \cdot b^1)(a + b) = (a^1 \cdot b^1) a + (a^1 \cdot b^1) b$

$=$ $(b^1 a^1) a + (a^1 b^1) b$

$=$ $b^1 (a^1 a) + a^1 (b^1 b)$

$=$ $b^1 . 0 + a_1 . 0$

1 and 2 $\Rightarrow$ Result $=$ $0 + 0 = 0$ ... (4)

2. $(a^1 + b^1) + (a \cdot b)$ $=$ $(a^1 + b^1 + a)(a^1 + b^1 + b) = (a^1 + a + b^1) \cdot (a^1 + b^1 + b)$

$=$ $(1 + b^1)(a^1 + 1) = 1.1 = 1$ ...(5)

Also $ab (a^1 + b^1)$ $=$ $(ab) a^1 + (ab) b^1 = a^1 (ab) + (ab)d$

$=$ $(a^1 a) b + a (bb) = 0.b + 9.0 = 0 + 0 = 0$ ...(6)

1 and 2 $\Rightarrow$ $(ab)^1$ $=$ $(a^1 + b^1)$

Cancellation Law

Theorem 3: In a boolean algebra B of $b + a$ and $b + a^1 = c + a^1$ then $b = c$. Also if $ba = ca$ and $ba^1 = ca^1$ then $b = c$

Proof: Assuming that $b + a = c + a$ and $b + a^1 = (+a)$

b $=$ $b + 0 = b + aa_1$ Let, ba $=$ ca and $ba^1 = ca^1$

$=$ $(b + a)(b + a^1)$ b $=$ $b.1 = b(a + a^1) = ba + ba^1$

$=$ $(c + a)(c + a^1)$ $=$ $ca + ca^1 = c(a + a^1) = c.1$

$=$ $c + aa^1 = c + 0$ $=$ $c.$

$=$ $c\ b = c$ b $=$ c

Logic (Digital) Circuits and Boolean Algebra
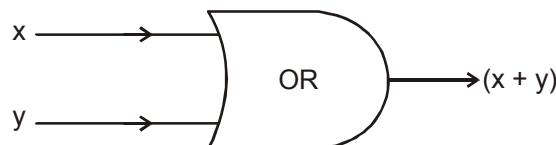
Some special type of net-works are used in digital computers for the processing of information in it. These net-works are represented by block diagrams. Logic circuits are structures which are built up from certain elementary circuits called logic gates.

Logic Gates

There are three basic logic gates.

1.   *OR gate* (OR)-Block used for OR gate is 'OR'. It converts two and more inputs into a single function given as follows:

Let x and y be two inputs, the output will be a function $(x \vee y)$ or $(x + y)$ as follows:
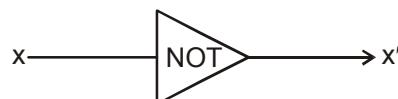
2. *AND gate* (AG) block used for AG is 'AG'. If there be two or more inputs then the output will be a function of those input gives as:
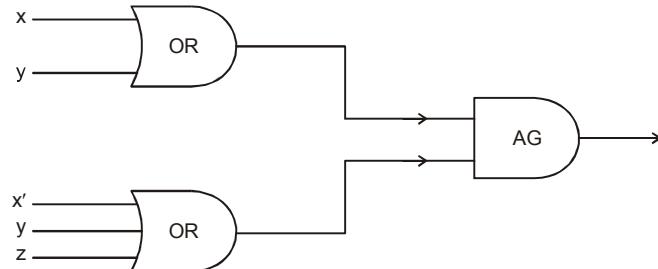
If x and y be two inputs then the output will be (x and y) *i.e.,* ($x \wedge y$) or x . y as follows:

3. *NOT gate* (or inverter). If the input be x then the output in converted into x' by an inverter.

Example: Write the Boolean function corresponding to the following network :

Required Boolean function is

$$(x + y)\,(x' + y + x')$$

Example: Design the gatting network for the function

1. $x\,y + y\,z$

2. $(x + y)'$

Solution :

1.

2.
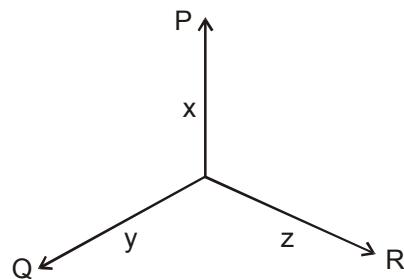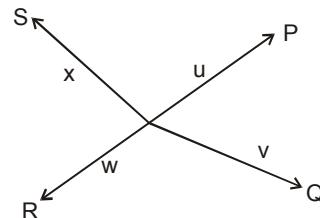
Many Terminal Networks

Any switching network having more than two terminate is called many terminal network.

If P, Q and R be the three terminals of any switching network then the transmittal of two terminals taken in pair will be denoted by T_{PO}, T_{QR}, T_{RP}.

Four terminals network is
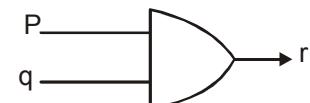
Symmetric Functions

Symmetric functions are those functions which remains unchanged by the interchange of any two variables

$$xy + x'y', x + y, x'y + xy' \quad \text{etc.}$$

OR Gate

Input		Output
P	**q**	**r (P + q)**
1	1	1
1	0	1
0	1	1
0	0	0
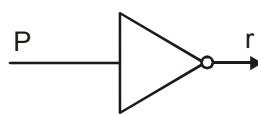
AND Gate

Input		Output
P	**q**	**P.q**
1	1	1
1	0	0
0	1	0
0	0	0
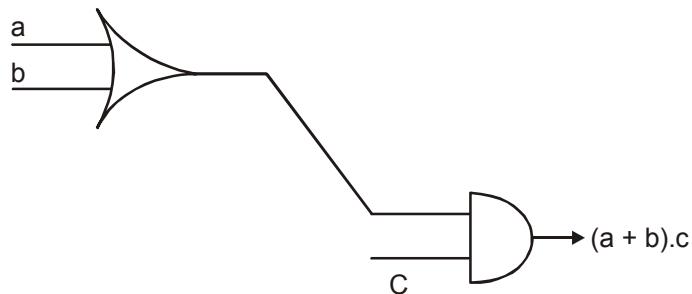
NOT Gate
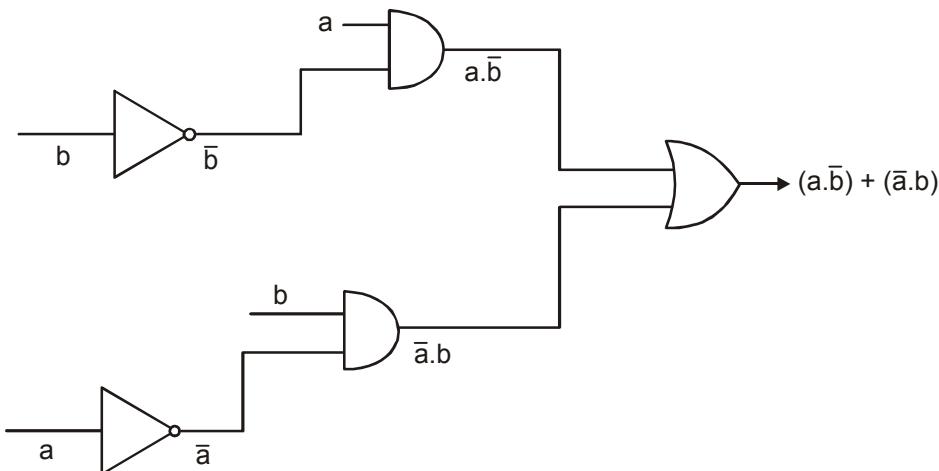
Input p	Output $r(\overline{P})$
1	0
0	1
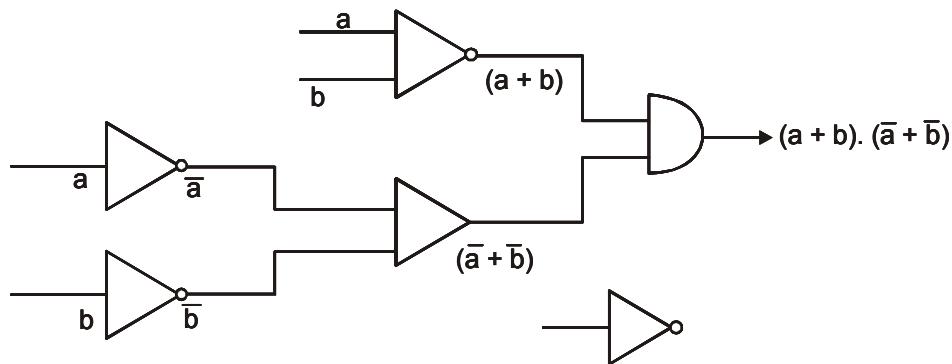
Example: 1. $(a + b) \cdot c$

2. $\left(a.\overline{b}\right) + \left(\overline{a}.b\right)$
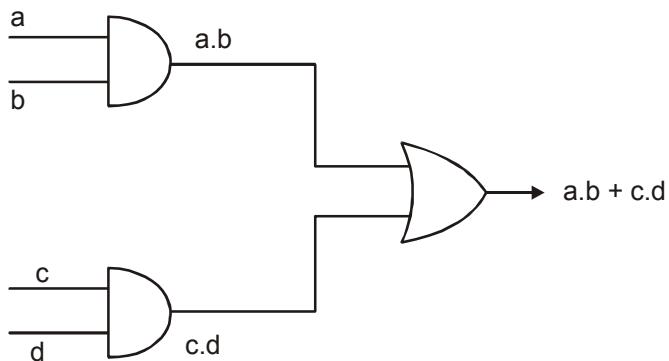
3. $(a + b).\left(\overline{a} + \overline{b}\right)$
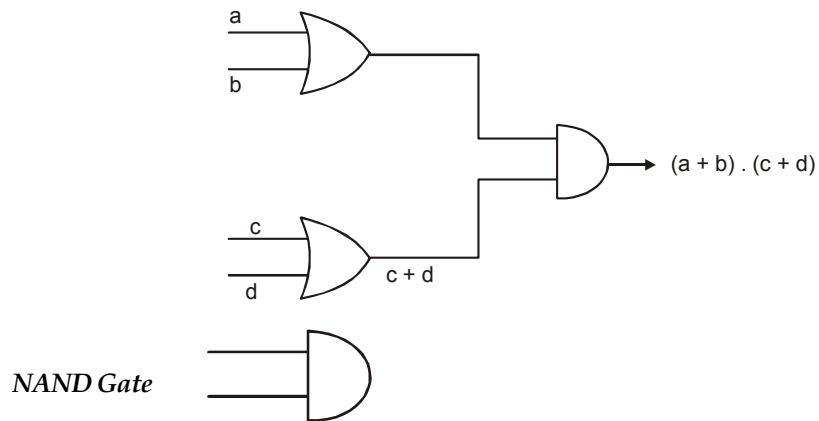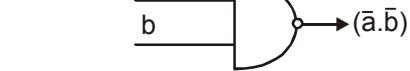
4.   AND-to-OR logic network

5.   OR to AND logic network

NAND Gate

Input		Output
a	b	
1	1	0
1	0	1
0	1	1
0	0	1
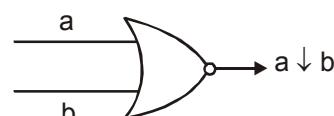
$$\Leftrightarrow \bar{a} + \bar{b}$$
$$\Leftrightarrow a \uparrow b$$

NOR Gate
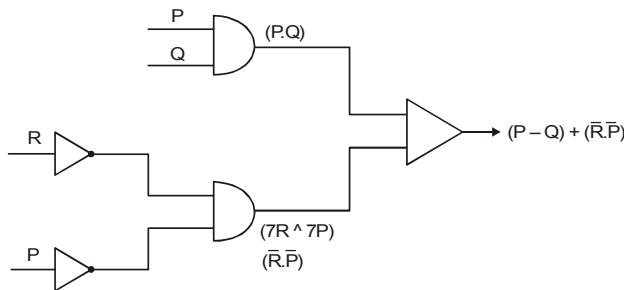
Input		Output
a	b	
1	1	0
1	0	0
0	1	0
0	0	1

$$\Leftrightarrow \bar{a} + \bar{b}$$

6. (a) $(P \wedge Q) \vee \left(7\overline{R} \wedge 7\overline{P}\right)$

Boolean Functions and Applications

Introduction: A *Boolean function* is a function from $\mathbb{Z}_n^2$ to $\mathbb{Z}_2$. For instance, consider the *exclusive-or* function, defined by the following table:

x_1	x_2	$x_1 \oplus x_2$
1	1	0
1	0	1
0	1	1
0	0	0

The exclusive-or function can interpreted as a function $\mathbb{Z}_2^2 \to \mathbb{Z}_2$ that assigns $(1, 1) \to 0$, $(1,0) \to 1$, $(0, 1) \to 1$, $(0, 0) \to 0$. It can also be written as a Boolean expression in the following way:

$$x_1 \oplus x_2 = \left(x_1 . \overline{x}_2\right) + \left(\overline{x}_2 . x_2\right)$$

Every Boolean function can be written as a Boolean expression as we are going to see next.

Disjunctive Normal Form: We start with a definition. A *minterm* in the symbols $x_1, x_2, ..., x_n$ is a Boolean expression of the form $y_1, y_2 .... y_n$, where each y_i is either x_i or $\overline{x}_i$.

Given any Boolean function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ that is not identically zero, it can be represented

$$f(x_1, ..., x_n) = m_1 + m_2 + ... + m_k,$$

where $m_1, m_2, ..., m_k$ are all the minterms $m_i = y_1 . y_2 ... y_n$ such that $f(a_1, a_2, ... a_n) = 1$, where $y_j = x_j$ if $a_j = 1$ and $y_j = \overline{x}_j$ if $a_j = 0$. That representation is called *disjunctive normal form* of the Boolean function f.
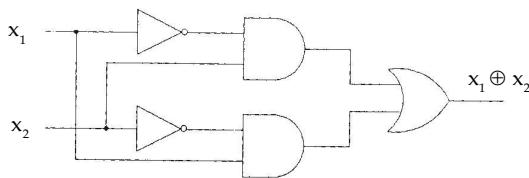
📝 *Example:* We have seen that the exclusive-or can be represented $x_1 \oplus x_2 = \left(x_1 . \overline{x}_2\right)\left(\overline{x}_1 . x_2\right)$. This provides a way to implement the exclusive-or with a combinatorial circuit as shown in figure.

Conjunctive Normal Form: A *maxterm* in the symbols $x_1, x_2, ..., x_n$ is a Boolean expression of the form $y_1 + y_2 + ... + y_n$, where each y_i is either x_i or $\overline{x}_i$.

Figure 7.1: Exclusive-Or.

Given any Boolean function $f: \mathbb{Z}_2^n \to \mathbb{Z}_2$ that is not identically one, it can be represented

$$f(x_1, ..., x_n) \;=\; M_1 . M_2 ... M_k.$$

where $M_1, M_2, \ldots, M_k$ are all the maxterms $M_i = y_1 + y_2 + ... + y_n$ such that $f(a_1, a_2, ... , a_n) = 0$, where $y_j = \overline{x}_j$ if $a_j = 0$ and $y_j = \overline{x}_j$ if $a_j = 1$. That representation is called *conjunctive normal form* of the Boolean function f.

Example: The conjunctive normal form of the exclusive-or is

$$x_1 \;\oplus\; x_2 = (x_1 + x_2) \bullet \left(\overline{x}_1 + \overline{x}_2\right).$$

Functionally Complete Sets of Gates: We have seen how to design combinatorial circuits using AND, OR and NOT gates. Here we will see how to do the same with other kinds of gates. In the following gates will be considered as functions from $\mathbb{Z}_2^n$ into $\mathbb{Z}_2$ intended to serve as building blocks of arbitrary boolean functions.

A set of gates $\{g_1, g_2, \ldots, g_k\}$ is said to be *functionally complete* if for any integer n and any function $f: \mathbb{Z}_2^2 \to Z_2$ it is possible to construct a combinatorial circuit that computes f using only the gates $g_1, g_2, ..., g_k$.
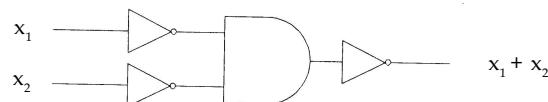
Example: The result about the existence of a disjunctive normal form for any Boolean function proves that the set of gates {AND, OR, NOT} is functionally complete. Next we show other sets of gates that are also functionally complete.

1.   The set of gates (AND, NOT} is functionally complete. Proof: Since we already know that (AND, OR, NOT} is functionally complete, all we need to do is to show that we can compute $x + y$ using only AND and NOT gates. In fact:

$$x + y = \overline{\overline{x} . \overline{y}} \, ,$$

hence the combinatorial circuit of Figure 7.2 computes $x + y$.

Figure 7.2: OR with AND and NOT
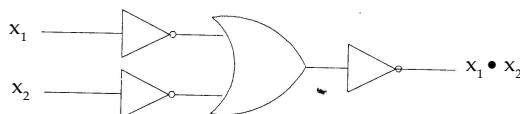
2.   The set of gates {OR, NOT} is functionally complete. The proof is similar :

$$x . y = \overline{\overline{x} + \overline{y}}$$

hence the combinatorial circuit of Figure 7.3 computes $x + y$.

Figure 7.3: AND with OR and NOT
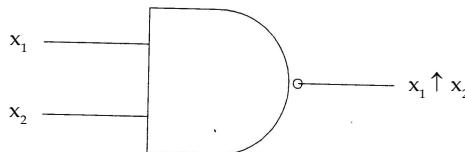
$x_1 \bullet x_2$

3.   The gate NAND, denoted $\uparrow$ and defined as

$$x_1 \uparrow x_2 \;=\; \begin{cases} 0 & \text{if } x_1 = 1 \text{ and } x_2 = 1 \\ 1 & \text{otherwise} \end{cases}$$

is functionally complete.
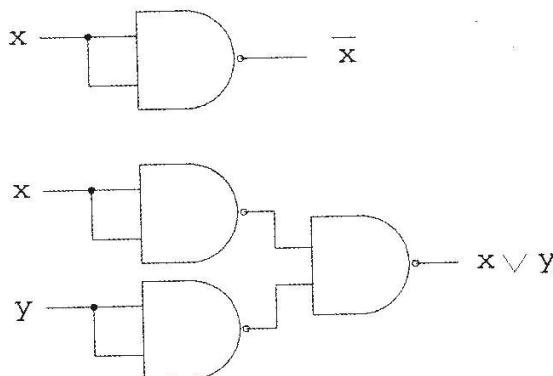
Figure 7.4: NAND Gate

$x_1 \uparrow x_2$

Proof: Note that $x \uparrow y = \overline{\overline{x} . \overline{y}}$. Hence $x = \overline{\overline{x} . \overline{y}} = x_1 \uparrow x_2$, so the NOT gate can be implemented with a NAND gate. Also the OR gate can be implemented with NAND gates: $x + y = \overline{\overline{x} . \overline{y}} = (x_1 \uparrow x_2) \uparrow (y \uparrow y)$. Since the set {OR, NOT} is functionally complete and each of its element can be implemented with NAND gates, the NAND gate is functionally complete.

Minimization of Combinatorial Circuits: Here we address the problems of finding a combinatorial circuit that computes a given Boolean function with the minimum number of gates. The idea is to simplify the corresponding Boolean expression by using algebraic properties such as $(E . a) + (E . \overline{a}) = E$ and $E + (E . a) = E$, where E is any Boolean expression. For simplicity in the following we will represent $a . b$ as $a \& b$, so for instance the expressions above will look like this: $Ea + E\overline{a} = E$ and $E + Ea = E$.

Figure 7.5: NOT and OR Functions Implemented with NAND Gates

$\overline{x}$

$x \vee y$
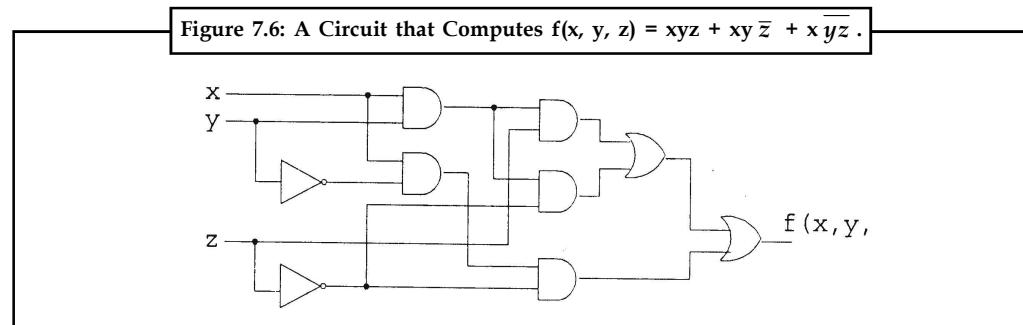
Example: Let F(*x*, *y*, *z*) the Boolean function defined by the following table:

x	*y*	*z*	*f(x, y, z)*
1	1	1	1
1	1	0	1
1	0	1	0
1	0	0	1
0	1	1	0
0	1	0	0
0	0	1	0
0	0	0	0
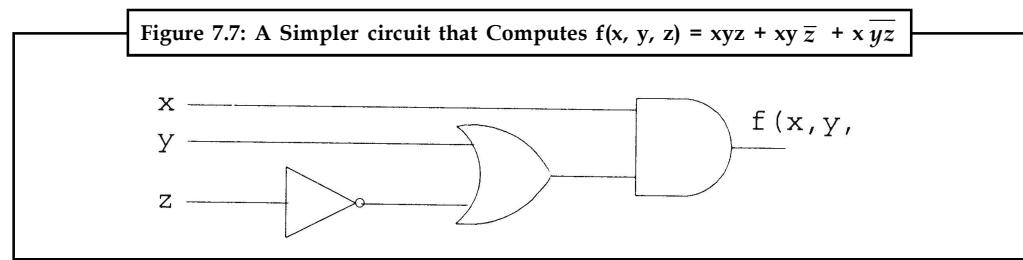
Its disjunctive normal form is $f(x, y, z} = xyz + xyz + x\,\overline{yz}$. This function can be implemented with the combinatorial circuit of Figure 7.6.

Figure 7.6: A Circuit that Computes f(x, y, z) = xyz + xy $\overline{z}$ + x $\overline{yz}$.

But we can do better if we simplify the expression in the following way:

$$f(x, y, z) = \overbrace{xyz + xy}^{xy}\,\overline{z} + x\,\overline{yz}$$

$$= xy + x\overline{yz}$$

$$= x(y + \overline{yz})$$

$$= x(y + \overline{y})(y + \overline{z})$$

$$= x(y + \overline{z})$$

which corresponds to the circuit of Figure 7.7.

Figure 7.7: A Simpler circuit that Computes f(x, y, z) = xyz + xy $\overline{z}$ + x $\overline{yz}$
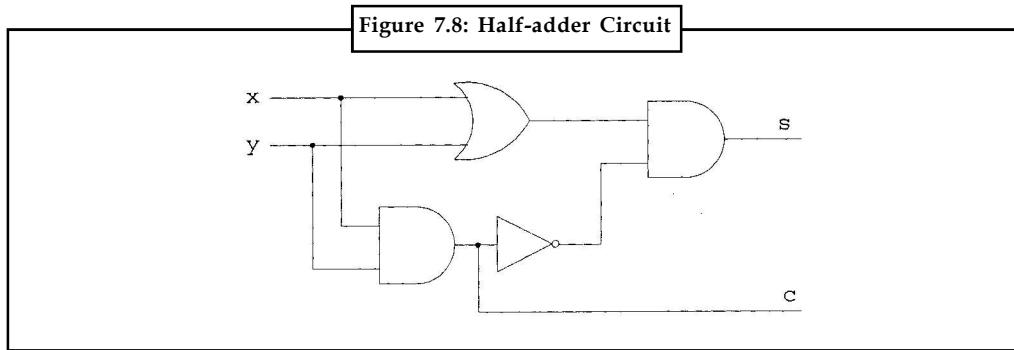
Multi-output Combinatorial Circuits: Example: Half-Adder. A half-adder is a combinatorial circuit **Notes**
with two inputs x and y and two outputs s and c, where s represents the sum of x and y and c is the
carry bit. Its table is as follows:

x	y	s	c
1	1	0	1
1	0	1	0
0	1	1	0
0	0	0	0

So the sum is $s = x \, 0 \, y$ (exclusive-or) and the carry bit is $c = x - y$. Figure 7.8 shows a half-adder
circuit.

Figure 7.8: Half-adder Circuit

Example:

The truth tables for the basic operations are:

And		
A	B	AB
0	0	0
0	1	0
1	0	0
1	1	1

Or		
A	B	A+B
0	0	0
0	1	1
1	0	1
1	1	1

Not	
A	A
0	1
1	0

For more complicated expressions, tables are built from the truth tables of their basic parts. Here are several:

1. Draw a truth table for *A+BC*.

A	B	C	BC	A+BC
0	0	0	0	0
0	0	1	0	0
0	1	0	0	0
0	1	1	1	1
1	0	0	0	1
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

2. Draw a truth table for *W(X+Y)Z*.

W	X	Y	Z	W	X+Y	W(X+Y)	W(X+Y)Z
0	0	0	0	1	0	0	0
0	0	0	1	1	0	0	0
0	0	1	0	1	1	1	0
0	0	1	1	1	1	1	1
0	1	0	0	1	1	1	0
0	1	0	1	1	1	1	1
0	1	1	0	1	1	1	0
0	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0
1	0	1	0	0	1	0	0
1	0	1	1	0	1	0	0
1	1	0	0	0	1	0	0
1	1	0	1	0	1	0	0
1	1	1	0	0	1	0	0
1	1	1	1	0	1	0	0

3. Draw a truth table for *A(B+D)*.

A	B	D	B+D	A(B+D)
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	1	0
1	0	0	0	0
1	0	1	1	1
1	1	0	1	1
1	1	1	1	1

4. Draw a truth table for *PT(P+Z)*.

P	T	Z	T	PT	P+Z	PT(P+Z)
0	0	0	1	0	0	0
0	0	1	1	0	1	0
0	1	0	0	0	0	0
0	1	1	0	0	1	0
1	0	0	1	1	1	1
1	0	1	1	1	1	1
1	1	0	0	0	1	0
1	1	1	0	0	1	0

5. Draw a truth table for *(A+B)(A+C)*.

A	B	C	A+B	A+C	(A+B)(A+C)
0	0	0	0	0	0
0	0	1	0	1	0
0	1	0	1	0	0
0	1	1	1	1	1
1	0	0	1	1	1
1	0	1	1	1	1
1	1	0	1	1	1
1	1	1	1	1	1

7.5 Summary

- Some special type of net-works are used in digital computers for the processing of information in it. These net-works are represented by block diagrams. Logic circuits are structures which are built up from certain elementary circuits called logic gates.

- It turns out that in order to reliably and objectively construct valid arguments, the logical operations which one uses must be clearly defined and must obey a set of consistent properties. Thus logic is quite rightly treated as a mathematical subject.

- The **NOT** (or **negation** or **inversion**) operator is a "**unary**" operator: it takes just one operand, like the unary minus in arithmetic (for instance, -x).

- The implication operator (**IMPLIES**) is a binary operator, and is defined in a somewhat counterintuitive manner (until you appreciate it, that is!).

7.6 Keywords

NOT: NOT is traditionally represented using either the tilde ("~").

Product of Sums: A function in *product of sums* form can be implemented using NOR gates by replacing all AND and OR gates by NOR gates.

7.7 Self Assessment

1. If *a, b, c* are elements of any Boolean algebra then prove that

    $ba = ca$ and $ba' = ca' \Rightarrow b = c$

2. In any Boolean algebra B prove that

    $a + b = 1 \Leftrightarrow a + b = b \ \forall \ a, b, \in B$

3. For any Boolean algebra B prove that

    $a + b = a + c$ and $ab = ac \Rightarrow b = c, \ \forall \ a, b, c, \in B.$

4. For any Boolean algebra B prove that

    $(a + b) . (b + c) . (c + a) = a.b + b.c + c.a$

5. If a, b, c are elements of any Boolean algebra then prove that

    $(a + b) (a' + c) = ac + a'b$

6. If a, b, c are elements of a Boolean algebra then prove that

    $a + b.c = b(b + c)$

7. If P, q, r, a, b are elements of a Boolean algebra then prove that

    (a) $Pqr + Pqr' + Pq'r + P'qr = Pq + qr + rP$

    (b) $ab + ab' + a'b + a'b' = 1$

8. In a Boolean algebra (B, +, •, '), prove that

    (a) $(a + b)' + (a + b')' = a'$

    (b) $ab + a'b' = (a' + b) (a + b')$

9. In a Boolean algebra B prove that

    (a) $ab + [c(a' + b')] = ab + c$

    (b) $(a + a'b) (a' + ab) = b$

10. In a Boolean algebra B prove that

    (a) $[x'(x + y)]' + [y(y + x')]' + [y(y' + x)][y(y' + x)]' = 1$

    (b) $(x + yz) (y' + x) (y' + z') = x(y' + z')$

11. Find whether the following statements are

    (a) $aa = a^2$ (b) $aa = a$

    (c) $0^1 = 1$ (d) $1 + a = 1$

    (e) $a + a^1 = 1$ (f) $0 + a = a$

    (g) $aa^1 = 0$ (h) $0\,a = 0$

    (i) $a (b + a) = ab + ac$ (j) $a + bc = (a + b) (a + c)$

12. Simplify the following products in boolean algebra :

    (a) $c (c + a)$ (b) $S (S^1 + t^1)$

    (c) $ab (a^1 + b^1)$ (d) $ab (ab + a^1c)$

    (e) $(P + q^1) (P^1 + q)$ (f) $a (a^1 + b + c)$

    (g) $a^1 (a + b^1 + c^1)$ (h) $(a + b) (a^1 + b^1)$

13. Factories and simplify the following in boolean algebra.

    (a) $pq + pr$ (b) $x^1 + x^1y$

    (c) $cx^1e + cde$ (d) $ab^1 + ab^1c^1$

    (e) $pq + qr + qr^1$ (f) $ab (ab + bc)$

(g) $(cd^1 + d^1)(c + d)$ (h) $(x y^1 + yz) (xz + yz^1)$

(i) $s^1 (s + t) (s + t)$ (j) $(x + yz) (y^1 + x) (y^1 + z^1)$

15.   By using the distractive law $a + bc = (a + b) (a + c)$ of boolean algebra factorise the following:

(a) $x + yz$ (b) $ax + b$

(c) $t + Pqr$ (d) $ab^1 + a^1 b$

(e) $x + x^1y$ (f) $ab + ab^1$

(g) $a + bcd$ (h) $ab + cd$

7.8 Review Questions

1.   Reduce the following Boolean products to either 0 or a fundamental product :

(a) $xyx'z$;   (b) $xyzy$;

(c) $xyz'yx$;   (d) $xyz'yx'z'$

2.   Express each Boolean expression E(x, y, z) as sum of products and then in its complete sum of products form :

(a) $E = x(xy' + x'y + y'z)$;   (b) $E = Z (x' + y) + y'$

3.   Express $E(x, y, z) = (x' + y)' + x'y$ in its complete sum of products form.

4.   Express each Boolean expression E(x, y, z) as a sum of products form :

(a) $E = y(x + yz)$;   (b) $E = x (xy + y + x'y)$;

5.   Express each set expression E(A, B, C) involving sets A, B, C as a union of intersections.

(a) $E = (A \cup B)^C \cap (C^C \cup B)$;   (b) $E = (B \cap C)^C \cap (A^C \cap C)^C$

6.   Let $E = xy' + xyz' + x'yz'$. Prove that :

(a) $xz + E = E$;   (b) $x + E \neq E$;

(c) $Z' + E \neq E$

7.   Let $E = xy' + xyz' + x'yz'$. Find
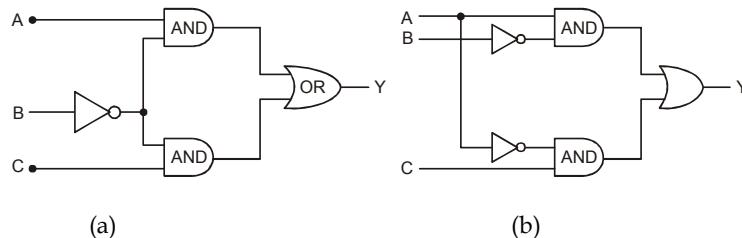
(a)   The prime implicants of E;   (b)   A minimum sum for E.

8.   Let $E = xy + y't + x'yz' + xy'zt'$. Find

(a)   Prime implicants of E;   (b)   Minimal sum for E.
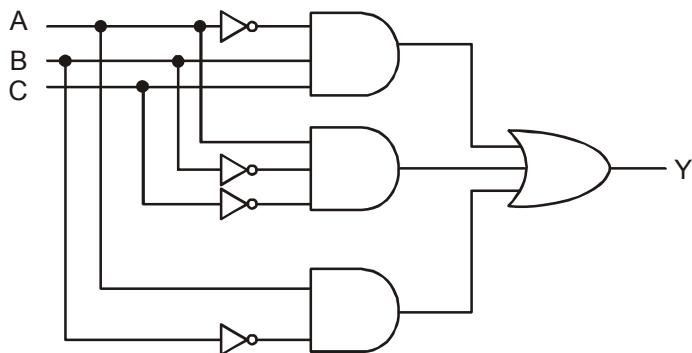
9.   Express the output X as a Boolean expression in the inputs A, B, C for the logic circuits in fig. (a) and fig. (b)

(a)   The inputs to the first AND gate are A and B' and to the second AND gate are B and C. Thus Y = AB' + B'C.

(a)   (b)
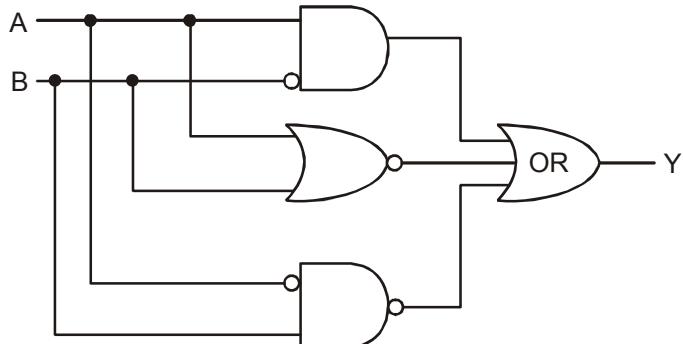
(b)   The input to the first AND gate are A and B′ to the second AND gate are A′ and C. Thus X = AB′ + A′C

10.   Express the output Y as a Boolean expression in the inputs A, B, C for the logic circuit in the figure.

11.   Express the output Y as a Boolean expression in the inputs A and B for the logic circuit in the figure.

12.   Find the output sequence Y for an AND gate with inputs A, B, C (or equivalently for Y = ABC) where :

(a)   A = 111001; B = 100101; C = 110011.

(b)   A = 11111100; B = 10101010; C = 00111100

(c)   A = 00111111; B = 11111100; C = 11000011

Answers: Self Assessment

1.   Let $ba = ca$ and $ba = ca$ ...(1)

Then, $b = b.1 = b(a + a') = ba + ba$

$= ca + ca$ [from (1)]

$= c(a + a')$

$= c.1$

$= c$

$\Rightarrow$ $b = c$

2. Let $\qquad a + b \;=\; b$ ...(1) **Notes**

Then, $\quad a' + (a + b) \;=\; (a' + a) + b$ [Associative law]

$\qquad\qquad = (a + a') + b$

$\qquad\qquad = 1 + b$

$\qquad\qquad = \; 1$

But, $\quad a' + (a + b) \;=\; a' + b$ [From (1)]

$\therefore \qquad\qquad a' + b \;=\; 1$

Conversely: Let $a' + b \;=$ 1

Then, $\qquad a + b \;=\; (a + b)\,.\,1$

$\qquad\qquad = (a + b)\,(a' + b)$ $[\because\; a' + b = 1]$

$\qquad\qquad = (b + a)\,(b + a')$ [Commutative property]

$\qquad\qquad = (b + aa')$

$\qquad\qquad = b + a$

$\qquad\qquad = \; b$

Hence, $\quad a' + b \;=\; 1 \Leftrightarrow a + b = b$

3. $\qquad\qquad b \;=\; b.1 = b\,(1 + a) = b.1 + b.a$ $[\because\; 1 + a = 1]$

$\qquad\qquad = \; b + ba$

$\qquad\qquad = \; bb + ba$ $[\because\; bb = b]$

$\qquad\qquad = \; b\,(b + a)$

$\qquad\qquad = \; b(a + b)$

$\qquad\qquad = \; b\,(a + c)$ [Given]

$\qquad\qquad = \; b.a + b.c$

$\qquad\qquad = \; a.b + bc$

$\qquad\qquad = \; ac + bc$ [Given]

$\qquad\qquad = \; (a + b).c$

$\qquad\qquad = \; c.(a + b)$

$\qquad\qquad = \; ca + c.c$

$\qquad\qquad = \; ca + c$

$\qquad\qquad = \; c(a + 1)$

$\qquad\qquad = \; c.1$

$\qquad\qquad = \; c$

Hence, $b \;=\; c$

4. $\qquad\qquad \text{LHS} \;=\; (a + b)\,.\,(b + c)\,.\,(c + a)$

$\qquad\qquad = \; ((a + b)\,(a + c)\,(b + c)$

$\qquad\qquad = \; (a + b\,.\,c)\,(b + a)$

$\qquad\qquad = \; (a + b\,.\,c)b + (a + bc)c$

$$= b(a + bc) + c(a + bc)$$

$$= b.a + bbc + ca + cbc$$

$$= a.b + (bb + bb')c + ca + ccb + cc'b$$

$$[bb' = 0; cc' = 0]$$

$$= ab + b (b + b')c + ca + c(c + c')b \qquad [b + b = 1 = c + c']$$

$$= a.b + bc + ca + cb$$

$$= ab + bc + ca = \text{RHS}$$

5. $\quad$ LHS $= (a + b) (a' + c) = (a + b)a' + (a + b)c$

$$= a' (a + b) + c(a + b)$$

$$= aa' + a'b + c(a + a' b) \qquad [\because a + a'b = (a + a')(a + b) = a + b]$$

$$= 0 + a'b + ca + ca'b$$

$$= a'b + ca'b + ca = ab + a'bc + ac$$

$$= a'b (a + c) + a.c = a'b.1 + a.c = a'b + ac$$

$$= ac + a'b = \text{RHS}$$

6. $\quad a + bc = (a + b) (a + c) = (b + a) (a + c) = (b + a).1.(a + c)$

$$= (b + a) (b + b') (a + c) = (b + ab') (a + c)$$

$$= (b + o) (a + c) \qquad [a \leq b \Rightarrow ab' = o]$$

$$= b(a + c) = \text{RHS}$$

7. (a) $\quad$ LHS $= Pqr + Pqr' + Pq'r + P'qr$

$$= Pq(r + r') + Pq'r + P'qr$$

$$= P.q.1 + Pq'r + P'qr = Pq + Pq'r + P'qr$$

$$= P(q + q'r) + P'qr = P(q + q') (q + r) + P'qr$$

$$= P.1(q + r) + P'qr = P(q + r) + P'qr = Pq + Pr + Pqr$$

$$= Pq (P + Pq)r = Pq + [(P + P') . (P + q)]r$$

$$= Pq + [1 . (P + q)]r = Pq + (P + q)r$$

$$= Pq + Pr + qr = Pq + qr + rP = \text{RHS}$$

(b) $\quad$ LHS $= ab + a.b' + a'b + a'b' = a (b + b') + a'(b + b')$

$$= a(1) + a'(1) = a + a' = 1 = \text{RHS}$$

8. (a) We know by Demorgan's law that :

$$(ab)' = a' + b'$$

Hence, $\qquad (a + b)' + (a + b') = (a + b) (a + b')'$

$$= (a + bb') = (a + o)'$$

$$= (a)' = a'$$

Hence, $(a + b)' + (a + b') = a'$

(b)   LHS $= ab + a'b' = (ab + a)(ab + b) = (a' + ab)(b' + ab)$

$\quad\quad\quad\quad = (a' + a)(a' + b)(b' + a)(b' + b) = 1(a' + b)(a + b')$

$\quad\quad\quad\quad = (a' + b)(a + b') = \text{RHS}$

9. (a)   LHS $= ab + [c(a' + b')] = ab + c(ab)' = (ab + c)(ab + (ab')')$

$\quad\quad\quad\quad = (ab + c)1 = (ab + c) = \text{RHS}$

(b)   LHS $= (a + a'b)(a' + ab) = (a + a')(a + b)(a' + a)(a' + b)$

$\quad\quad\quad\quad = 1(a + b)1(a' + b) = (a + b)(a' + b) = (b + a)(b + a)$

$\quad\quad\quad\quad = b + aa'$

$\quad\quad\quad\quad = b + o = b = \text{RHS}$

10. LHS $\quad\quad\quad = [x'(x + y)]' + [y(y + x')] + [y(y' + x)]'$

$\quad\quad\quad\quad = (x')' + (x + y) + (y)' + (y + x')' + y' + (y' + x')'$

$\quad\quad\quad\quad = x + x'y' + y' + y'x + y' + yx'$

$\quad\quad\quad\quad = x + x'y' + y' + y'x + yx'$

$\quad\quad\quad\quad = x + xy' + y' + y'x' + yx'$

$\quad\quad\quad\quad = x(1 + y') + y(1 + x') + yx'$

$\quad\quad\quad\quad = x + y' + yx' = x + x'y + y' = (x + x')(x + y) + y'$

$\quad\quad\quad\quad = x + y + y = x + 1 = 1 = \text{RHS}$

$\quad\quad$RHS $= (x + yz)(y' + x)(y' + z') = (x + y)(x + z)(x + y') + (y + z')$

$\quad\quad\quad\quad = (x + y)(x + y')(x + z)(y' + z') = (x + yy')(x + z)(y' + z')$

$\quad\quad\quad\quad = x(x + z)(y' + z') = (x + xz)(y' + z') = x(1 + z)(y' + z')$

$\quad\quad\quad\quad = x.1(y' + z') = x(y' + z') = \text{RHS}$

11. (a) Only in ordinary algebra   (b) Only in Boolean algebra

(c) Only in Boolean algebra   (d) Only in Boolean algebra

(e) Only in Boolean algebra   (f) In the both algebras

(g) In the both algebras   (h) In Boolean algebra

12. (a) $c + cd$;   (b) st';   (c) O;

(d) ab;   (e) $Pq + P'q'$;   (f) $ab + ac$;

(g) $a'b' + a'c'$;   (h) $ab' + a'b$.

13. (a) $P(q + r)$;   (b) x';   (c) cl;

(d) ab';   (e) q;   (f) ab;

(g) $cd + c'd'$;   (h) xz;   (i) o;

(j) $x(y' + z')$

14. (a) $(x + y)(x + z)$;   (b) $(a + b)(x + b)$;   (c) $(t + p)(t + q)(t + r)$;

(d) $(a + b)(a' + b')$;   (e) $x + y$;   (f) $(a + b')(b + a')$

(g) $(a + b)(a + c)(a + d)$;   (h) $(a + c)(b + c)(a + d)(b + d)$
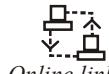
7.9 Further Readings

Books   Béla Bollobás, *Modern Graph Theory*, Springer

Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph Theory, Combinatorics, and Algorithms*, Birkhäuser

Online links   http://en.wikiversity.org/wiki/Compound_Propositions_and_Useful_ Rules#Writing_Truth_Tables_For_Compound_Propositions

http://nrich.maths.org/6023

Unit 8: Mathematical Logic

CONTENTS

Objectives

Introduction

8.1   Truth Tables

8.2   Tautologies and Contradictions

8.3   Logical Equivalence

8.4   Algebra of Propositions

8.5   Conditional and Biconditional Statements

8.6   Types of Statements

      8.6.1   Atomic Statement (Simple Statements)

      8.6.2   Molecular or Compound Statement

      8.6.3   Contingency and Satisfiable

      8.6.4   Equivalence of Formulae

8.7   Laws

8.8   Predicated Logic

8.9   Summary

8.10 Keywords

8.11 Self Assessment

8.12 Review Questions

8.13 Further Readings

Objectives

After studying this unit, you will be able to:

- Describe the tautologies and contradictions
- Define logical equivalence
- Explain the algebra of proposition
- Define conditional and biconditional statements

Introduction

Mathematical logic (also known as **symbolic logic**) is a subfield of mathematics with close connections to foundations of mathematics, theoretical computer science and philosophical logic. The field includes both the mathematical study of logic and the applications of formal logic to other areas of mathematics. The unifying themes in mathematical logic include the study of the expressive power of formal systems and the deductive power of formal proof systems.

Mathematical logic is often divided into the fields of set theory, model theory, recursion theory, and proof theory. These areas share basic results on logic, particularly first-order logic, and definability. In computer science (particularly in the ACM Classification) mathematical logic encompasses additional topics not detailed in this article; see logic in computer science for those.

Since its inception, mathematical logic has contributed to, and has been motivated by, the study of foundations of mathematics. This study began in the late 19th century with the development of axiomatic frameworks for geometry, arithmetic, and analysis. In the early 20th century it was shaped by David Hilbert's program to prove the consistency of foundational theories. Results of Kurt Gödel, Gerhard Gentzen, and others provided partial resolution to the program, and clarified the issues involved in proving consistency. Work in set theory showed that almost all ordinary mathematics can be formalized in terms of sets, although there are some theorems that cannot be proven in common axiom systems for set theory. Contemporary work in the foundations of mathematics often focuses on establishing which parts of mathematics can be formalized in particular formal systems, rather than trying to find theories in which all of mathematics can be developed.

8.1 Truth Tables

"A truth table displays the relationship between the truth values of propositions".

1. *Negation operator:* Let P be a proposition. The statement

   "It is not the case that P" is another proposition, called the negation of P. The negation of P is denoted by GP.

   Example: Find the negation of the proposition

   "Today is Friday"

   and express this in simple English.

   Solution: The negation is

   " It is not the case that today is Friday"

   The truth table for the negation of a proportion :

P	ΓP
T	F
F	T

2. *Conjunction operation:* Let P and q be two propositions. The proposition "P and 'V'" denoted by P $\wedge$ q, is the proposition that is true when both P and q are true and is false otherwise and denoted by "$\wedge$".

   Example: Find the conjunction of the propositions P and q where P is the proposition "Today is Friday" and q is the proposition " It is raining today".

   Solution: The conjunction of these Propositions P and q where P $\wedge$ q is the proposition "Today is Friday and it is raining today". This is true on rainy Fridays and is false on any day that is not a Friday and on Fridays when it does not rain.

3. *Disjunction operation:* Let P and q be propositions. The propositions "P or q" denoted by P $\vee$ q, is the proposition that is false when P and q are both false and true otherwise the proposition P $\vee$ q is called the disjunction of P and q.

The truth table for the conjunction of two propositions		
P	**q**	**P ∧ q**
T	T	T
T	F	F
F	T	F
F	F	F

The truth table for the disjunction of two propositions		
P	**q**	**P ∨ q**
T	T	T
T	F	T
F	T	T
F	F	F

4. *Exclusive OR operation:* Let P and *q* be propositions. The exclusive or of P and *q*, denoted by P ⊕ *q*, is the proposition that is true when exactly one of P and *q* is true and is false otherwise.

The truth table for the exclusive or of two propositions

P	**q**	**P ⊕ q**
T	T	F
T	F	T
F	T	T
F	F	F

5. *Implication operation:* Let P and *q* be propositions. The implication P → *q* is the proposition that is false, when P is true and *q* is false and true otherwise. In this implication P is called the hypothesis (or antecedent or premise) and *q* is called conclusion (or consequence).

The truth table for the Implication P → *q*

P	**q**	**P → q**
T	T	T
T	F	F
F	T	T
F	F	T

Some of the common ways of expressing this implication are

* "If P the *q*",   * "P is sufficient for *q*"

* "P implies *q*",   * "*q* if P"

* " If P, *q*"   * "*q* whenever P"

* "P only if"   * "*q* is necessary for P"

6. *Biconditional operator:* Let P and *q* be propositions the biconditional P → *q* is the proposition that is true when P and *q* have the same truth values and is false otherwise.

"P if and only if *q*"

The truth table for the Biconditional P ↔ q

P	**q**	**P ↔ q**
T	T	T
T	F	F
F	T	F
F	F	T

?

Did u know? **What is the difference between implication operator and biconditional operator?**

Example: Prove that p ∨ (q ∧ r) and (p ∨ q) ∧ (p ∨ r) have the same truth table i.e. they are equivalent.

Solution:

1	2	3	4	5	6	7	8
p	q	r	q ∧ r	p ∨ q	p ∨ r	p ∨ (q ∧ r)	(p ∨ q) ∧ (q ∨ r)
T	T	T	T	T	T	T	T
T	T	F	F	T	T	T	T
T	F	T	F	T	T	T	T
F	T	T	T	T	T	T	T
F	F	T	F	F	T	F	F
F	T	F	F	T	F	F	F
T	F	F	F	T	T	T	T
F	F	F	F	F	F	F	F

Since the truth values of the columns (7) and (8) are the same for all values of the component of sentences. Hence the given functions are equivalent.

Example: Using, truth table prove the DeMorgan's laws

1. ~ (p ∧ q) ≡ ~ p ∨ ~ q

2. ~ (p ∨ q) ≡ ~ p ∧ ~ q

Solution: Consider the truth table

1	2	3	4	5	6	7
p	q	p ∧ q	–p	–q	~ (p ∧ q)	(~ p ∨ ~ q)
T	T	T	F	F	F	F
T	F	F	F	T	T	T
F	T	F	T	F	T	T
F	F	F	T	T	T	F

Since the truth values of the columns (6) and (7) are same law both functions are equivalent.

Similarly we can prove (U) result.

Example: Prove that $(p \Rightarrow q) \vee r \equiv (p \vee r) \Rightarrow (q \vee r)$

Solution: Consider the truth table

1	2	3	4	5	6	7	8
p	q	r	p $\Rightarrow$ q	(p$\Rightarrow$q) $\vee$ r	p $\vee$ r	q $\vee$ r	(p $\vee$ r) $\Rightarrow$ (q $\vee$ r)
T	T	T	T	T	T	T	T
T	T	F	T	T	T	T	T
T	F	T	F	T	T	T	T
F	T	T	T	T	T	T	T
T	F	F	F	F	T	F	F
F	F	T	T	T	F	T	T
F	T	F	T	T	T	T	T
F	F	F	T	T	F	F	T

Since the truth values of the columns (5) and (8) are same hence the given sentences are logically equivalent.

Example: Simplify the following propositions:

1. $P \vee (P \wedge Q)$

2. $(P \vee Q) \wedge \sim P$

3. $\sim (P \vee Q) \vee (\sim P \wedge Q)$

Solution:

1. $P \vee (P \wedge Q)$

    $\equiv (P \wedge T) \vee (P \wedge Q)$ (since $P \wedge T = P$)

    $\equiv P \wedge (T \vee Q)$

    $\equiv P$ (since $T \vee Q = T$)

2. $(P \vee Q) \wedge \sim P$

    $\equiv \sim P \wedge (P \vee Q)$ (Commutative Law)

    $\equiv (\sim P \wedge P) \vee (\sim P \vee Q)$ (Distributive Law)

    $\equiv F \vee (\sim P \vee Q)$ (Complement Law)

    $\equiv (\sim P \vee Q)$ (Identity Law)

    $= \sim P \vee Q$

3. $\sim (P \vee Q) \vee (\sim P \wedge Q)$

    $\Rightarrow (\sim P \wedge \sim Q) \vee (\sim P \wedge Q)$

    $\equiv \sim P(\sim Q \vee Q)$ (By distributive law)

    $\equiv \sim P \wedge T$ $(\sim Q \vee Q = T)$

    $\equiv \sim P$ $(P \wedge T = P)$

Example: Construct truth table for the following functions and check whether it is a tautology or contradiction.

1. $[(p \wedge q) \vee (q \wedge r) \vee (r \wedge p)] \Leftrightarrow [(p \vee q) \wedge (q \vee r) \wedge (r \vee p)]$

Solution: Truth table corresponding to the given function is given below :

p	q	r	p ∧ q	q ∧ r	r ∧ p	p ∨ q	q ∨ r	r ∨ p	(p ∧ q) ∨(q ∧ r) ∨(r ∨ p)	(p ∨ q) ∧(q ∨ r) ∧(r ∨ p)
T	T	T	T	T	T	T	T	T	T	T
T	T	F	T	F	F	T	T	T	T	T
T	F	T	F	F	T	T	T	T	T	T
F	T	T	F	T	F	T	T	T	T	T
T	F	F	F	F	F	T	F	T	F	F
F	T	F	F	F	F	T	T	F	F	F
F	F	T	F	F	F	F	T	T	F	F
F	F	F	F	F	F	F	F	F	F	F

Since the last two columns of the above table are the same. Hence given function represents a tautology,

2. $(p \Rightarrow q) \Leftrightarrow (\sim p \vee q)$

p	q	~p	p ⇒ q	~p ∨ q	p ⇒ q ⇔ (~p ∨ q)
T	T	F	T	T	T
T T	F	F	F	F	T
F F	T	T	T	T	T
F	F	T	T	T	T

Since the last column contains all truth values T. Hence given function represents a tautology.

3. $(p \wedge q) \Rightarrow p] \Rightarrow [q \wedge \sim q]$

p	q	~q	p ∧ q	(p ∧ q) ⇒ p	q ∧ ~q	(p ∧ q) ⇒ p ⇒ [q ∧ ~q]
T	T	F	T	T	F	F
T	F	T	F	T	F	F
F	T	F	F	T	F	F
F	F	T	F	T	F	F

Since the last column contains all truth values F. Hence the given function represents a contradiction.

4. $[(p \wedge q) \Rightarrow r] \Leftrightarrow [(p \Rightarrow r) \vee (q \Rightarrow r)]$

Solution:

p	q	r	p ∧ q	(p ⇒ q)	q ⇒ r	(p ∧ q) ⇒ r	(p ⇒ q) ∨ (q ⇒ r)	(d)
T	T	T	T	T	T	T	T	T

Contd...

T	T	F	T	T	F	F	T	F
T	F	T	F	F	T	T	T	T
F	T	T	F	T	T	T	T	T
T	F	F	F	F	T	T	T	T
F	T	F	F	T	F	T	T	T
F	F	T	F	T	T	T	T	T
P	F	F	F	T	T	T	T	T

Since the last column of above table does not contain all truth values T, so, it is not a tautology.

Example: Establish the equivalence using truth tables.

1. $(p \Rightarrow q) \vee (p \Rightarrow r) \equiv p \Rightarrow (q \vee r)$

Solution: Truth table for the given statement is

p	q	r	q ∨ r	p ⇒ q	p ⇒ r	(p ⇒ q) ∨ (p ⇒ r)	p ⇒ (q ∨ r)
T	T	T	T	T	T	T	T
T	T	F	T	T	F	T	T
T	F	T	T	F	T	T	T
F	T	T	T	T	T	T	T
T	F	F	F	F	F	F	F
F	T	F	T	T	T	T	T
F	F	T	T	T	T	T	T
F	F	F	F	T	T	T	T

Since the last two columns of the above table have the same truth values. Hence the given statements are logically equivalent.

2. $(p \Rightarrow q) \Rightarrow (p \wedge q) = (\sim p \Rightarrow q) \wedge (q \Rightarrow p)$

Solution: We construct the truth table

p	q	~p	p ∧ q	p ∧ q	~p ⇒ q	q ⇒ p	(p ∧ q) ⇒ (p ∧ q)	(~p ⇒ q) ∧ (q ⇒ p)
T	T	F	T	T	T	T	T	T
T	F	F	F	F	T	T	T	T
F	T	T	T	F	T	F	F	F
F	F	T	T	F	F	T	F	F

Since the last two columns have the same truth values. Hence given statements are logically equivalent.

3. $(p \vee q) (p \wedge q) \equiv p$

Solution: Truth table is

p	q	p ∨ q	p ∧ q	(p ∨ q) ∨ (p ∧ q)
T	T	T	T	T
T	F	T	F	T
F	T	T	F	T
F	F	F	F	F

Since the first and last columns of the given statement do not contain same truth values. Hence given statements are not logically equivalent.

4.　　$p \Rightarrow (q \wedge r) \equiv (p \Rightarrow q) \wedge (p \Rightarrow r)$

Solution: Truth Tables given by

p	q	r	q	p ⇒ q	p ⇒ r	p ⇒ (q ∧ r)	(p ⇒ q) ∧ (p ⇒ r)
T	T	T	T	T	T	T	T
T	T	F	F	T	F	F	F
T	F	T	F	F	T	F	F
F	T	T	T	T	T	T	T
T	F	F	F	F	F	F	F
F	T	F	F	T	T	T	T
F	F	T	F	T	T	T	T
F	F	F	F	T	T	T	T

Since the last two columns of the above table have the same truth values hence the given statements are logically equivalent.

Example: Establish the equivalence analytically and write dual also of the given statement

$$\sim p \wedge (\sim q \wedge r) \vee (q \wedge r) \vee (p \wedge r) \equiv r.$$

Solution: Consider L.H.S.

~p ∧ (~ q ∧ r) ∨ (q ∧ r) ∨ (p ∧ r)

$\equiv [(\sim p \wedge \sim q) \wedge r) \vee [[(q \wedge r) \vee p] \vee [(q \wedge r) \vee r]$

$\equiv [(\sim p \wedge \sim q) \wedge r) \vee [\{(q \wedge r) \vee p\} \vee r]$

$\equiv [(\sim p \wedge \sim q) \wedge r) \vee [\{(q \wedge r) \wedge r\} \vee (p \wedge r)]$

$\equiv [(\sim p \wedge \sim q) \wedge r) \vee [\{(q \wedge r) \vee (p \wedge r)\}]$

$\equiv [(\sim p \wedge \sim q) \wedge r) \vee [(q \wedge p) \vee r]$

$\equiv [(\sim p \wedge \sim q)] \vee [(q \vee p)] \vee r$

$\equiv [\sim(p \vee q) \vee (p \vee q)] \wedge r$

$\equiv r$ R.H.S.

Dual of the given statement can be obtained by interchanging $\wedge$ and $\vee$ and is $\sim p \vee (\sim q \vee r) \wedge (q \vee r) \wedge (p \vee r) \equiv r$.

8.2 Tautologies and Contradictions

Tautology: A statement that is true for all possible values of its propositional variable is called a tautology. A statement that is always false is called a contradiction and a statement that can be either true or false depending on the truth values of its propositional variables is called a contingency.

Example:

P	~P	P ∨ (~P)	P ∧ (~ P)
T	F	T	F
F	T	T	F

Notes P∨ (~P) is a tautology and P ∧ (~ P) is a contradiction

8.3 Logical Equivalence

Two propositions P(p, q, ...) and Q(p, q, ...) are said to be **logically equivalent** or simply equivalent or **equal** if they have identical truth tables and is written as

$$P(p, q, ...) \equiv Q(p, q, ...)$$

As shown below truth tables of ¬ (p ∧ q) and ¬ p ∨ q are identical. Hence ¬ (p ∧ q) = ¬ p ∨ ¬ q

P	q	p ∧ q	~ (p ∧ q)
T	T	T	F
T	F	F	T
F	T	F	T
F	F	F	T
		~ (p ∧ q)	

P	q	¬ p	¬ q	¬ p ∨ ¬ q
T	T	F	F	F
T	F	F	T	T
F	T	T	F	T
F	F	T	T	T
			¬ p ∨ ¬ q	

Also when a conditional p ⇒ q & its converse q ⇒ p are both ture, the statement p & q are said to be logically equivalent. Consider the following table

P	F	p ∧ F	p ∨ F	T	p ∧ T	p ∨ T
T	F	F	T	T	T	T
T	F	F	T	T	T	T
F	F	F	F	T	F	T
F	F	F	F	T	F	T
		p ∧ F ≡ F	p ∨ F ≡ p		p ∧ T ≡ p	p ∨ T ≡ T

Hence T is the **identity** element for the operation ∧ (**conjunction**) and F is the **identity** element for the operation ∨ (**disjunction**).

Duality : Any two formulas A and A* are said to be duals of each other if one can be obtained from the other by replaying ∧ by ∨ and ∨ by ∧.

Example: 1.   The dual of (P ∧ q) ∨ r is (P ∧ q) ∧ r.

2.   The dual of (P ∨ q) ∧ r is (P ^ q) ∨ r.

Functionally complete Set : "A set of connectives is called functionally complete if every formula can be expressed in terms of an equivalent formula containing the connectives from this set".

^, ∨, ~, ⇒, ⇔

Example: Consider P ⇔ q

It is equivalent to (P ⇒ q, ∧ (q ⇒ P)

Task   Analyse the importance of logical equivalence in truth tables.

8.4 Algebra of Propositions

We can easily verify by truth tables that every proposition satisfy the following laws. If P and Q are propositions and Fare tautology and contradiction respectively.

Associative Laws

1.   (P ∨ Q) ∨ R = P ∨ (Q ∨ R)

2.   (P ∧ Q) ∧ R = P ∧ (Q ∨ R)

Commutative Laws

1.   P ∨ Q = Q ∧ P

2.   P ∧ Q = Q ∧ P

Distributive Laws

1.   P ∨ (Q ∧ R) = (P ∨ Q) ∧ (P ∨ R)

2.   P ∧ (Q ∨ R) = (P ∧ Q) ∨ (P ∧ R)

DeMorgan's Laws

1.   ~ (P ∨ Q) ≡ ~ P ∧ ~ Q

2.   ~ (P ∧ Q) ≡ ~P ∨ ~ Q

3.   ~ (P ⇒ Q) ≡ ~P ∧ ~Q

4.   ~ (P ⇔ Q) ≡ P ⇔ ~Q ~P ⇔ Q

Identity Laws

1.   P ∨ F = P

2.   P ∨ ~P ≡ F

3.   ~~P ≡ P

4.   ~T ≡ F

5.   ~F ≡ T

Idempotent Laws

1. $P \lor P \equiv P$

2. $P \land P \equiv P$

Compliment Laws

1. $P \lor \sim P = T$

2. $P \land \sim P \equiv F$

3. $\sim \sim P \equiv P$

4. $\sim T \equiv F$

5. $\sim F \equiv T$

Notes

1.   *Idempotent Law -* a set associated with itself results in the same original set. e.g. $A \land A = A$ and $A \lor A = A$

2.   *Commutative Law -* sets can be swapped around freely in a simple association. e.g. $A \land B$ is the same as $B \lor A$

3.   *Associative Law -* with operators of same type being used, brackets can go anywhere. e.g. $(A \land B) \land C$ is the same as $A \land (B \land C)$

8.5 Conditional and Biconditional Statements

Two propositions (simple or compound) are said to be logically equivalent if they have identical truth values.

If p is logically equivalent to q, we denote $p \equiv q$.

We have already constructed the truth table of $(\sim p) \lor q$ in earlier class. Recall the same truth table.

Truth Table for $(\sim p) \lor q$

p	q	~p	(~p) ∨q
T	T	F	T
T	F	F	F
F	T	T	T
F	F	T	T

Note that the last column of above table and last column of earlier table are identical. This implies $p \rightarrow q$ is logically equivalent to $(\sim p) \lor q$.

1. $p \rightarrow q \equiv (-p) \lor q$.

We explain equation 1 with the following example.

Consider the statement "either the monsoon break in June or the rivers are dried"

Let p: The monsoon does not break in june

q: rivers are dried

Clearly 'if p then q' is true if p is true and q is true. That is, the conditional statement of the above statement is "If the monsoon does not break in June then rivers are dried".

This statement is equivalent to "The monsoon breaks in the month of June or the rivers are dried" which is symbolically written as $(\sim p) \vee q$.

We know from equation 1 that

$$(\sim p \vee q) \equiv p \to q$$

Taking negation on both sides, we have

$$\sim (\sim p \vee q) \equiv \sim (p \to q)$$

Using Demorgan's law to L.H.S, we have

$$\sim (\sim p) \wedge \sim q \equiv \sim (p \to q)$$

$$\Rightarrow p \wedge \sim q \equiv \sim (p \to q) \left(\because \sim (\sim p) = p \right)$$

Thus we define

Negation of Conditional Statement

2. $\sim (p \to q) = p \wedge (\sim q)$

Write the negation of the conditional statement

"If the weather is cold then it will snow"

The negation of the given statement is "The weather is cold and it will not snow"

Contrapositive of a Conditional Statement

We know that

$$p \to q \equiv (\sim p) \vee q \text{ (C5)}$$

$$\equiv q \vee (\sim p) \text{ (commutative)}$$

$$\equiv \sim (\sim q) \vee (\sim p) \ (\because \sim (\sim q) = q)$$

$$\equiv \sim q \to \sim p \ \text{ (C5)}$$

$\sim q \to \sim p$ is called the contrapositive of $p \to q$,

3. $p \to q \equiv (-q) \to (-p)$

The conditional statement is equivalent to it contrapositive or

Example: Consider the conditional statement "If a quadrilateral is a square, then all the sides are equal".

The contrapositive of this statement is "If all the sides of a quadrilateral are not equal, then it is not a square".

The Biconditional Statement

If two simple statements p and q are connected by the connective 'if and only if', then the resulting compound statement is called the biconditional statement. Symbolically it is represented by p ↔ q.

Example: An integer is even if and only if it is divisible by 2. It is a biconditional having the truth value T.

The biconditional statement p ↔ q is true when either both p and q are true or both p and q are false.

The truth table for biconditional is given below

p	q	p ↔ q
T	T	T
T	F	F
F	T	F
F	F	T

For further reference, we write the truth values of biconditional statements as follows.

p ↔ q has the truth value T if both p and q have the same truth value.

That is

p is true and q is true ⇒ p ↔ q is true.

p is false and q is false ⇒ p ↔ q is true.

p ↔ q has truth value F if p and q has opposite truth values.

That is

If p has truth value T and q has the truth value F, p ↔ q has truth value F.

Also,

if p has the truth value F and q has the truth value T, p ↔ q has truth value F.

Another way of defining biconditional statement is given below.

If p and q are propositions, then the conjunction of conditionals p → q and q → p is called a biconditional proposition.

Let us construct the truth table of (p → q) ∧ (q → p)

Truth Table for (p → q) ∧ (q → p)

p	q	p → q	q → p	(p → q) ∧ (q → p)
T	T	T	T	T
T	F	F	T	F
F	T	T	F	F
F	F	T	T	T

The last column of above table is identical to the last column of earlier table.

∴   p ↔ p and (p → q) ∧ (q → p) are logically equivalent

8.6 Types of Statements

8.6.1 Atomic Statement (Simple Statements)

A statement which has no logical connective in it is called atomic sentences ,for example:

1.   This is my body (true)

2.   Delhi is the capital of U.P (False).

8.6.2 Molecular or Compound Statement

The sentence formed by two simple sentences using a logical connective is called a compound sentences. For example,

1.   U.P. is in India and Lucknow is the capital of U.P.

2.   Hari will study he will play.

Example: Construct the truth table for

$$\sim (P \wedge q) \Leftrightarrow (\sim P \vee \sim q)$$

P	q	P ∧ q	~(P ∧ q)	~ q	~q	~P ∨ ~ q	~ (P ∧ q) ⇔ (~ P ∨ ~ q)
T	T	T	F	F	F	F	F
T	F	F	T	F	T	T	T
F	T	F	T	T	F	T	T
F	F	F	T	T	T	T	T

8.6.3 Contingency and Satisfiable

A statement formula (expression involving prepositional variable) that is neither a tautology nor a contradiction is called a contingency

If the resulting truth value of a statement formula A $(p_1, p_2,... p_n)$ is true for at least one combination of truth values of the variables $p_1, p_2,..., p_n$ then A is said to be satisfiable

Example: Prove that the sentence 'It is wet or it is not wet' is a tautology.

Solution: Sentence can be symbolized by $p \vee \sim p$ where p symbolizes the sentence 'It is wet'. Its Truth table is given below:

p	~p	p ∨ ~ p
T	F	T
F	T	T

As column 3 contains T everywhere, therefore the given proposition is tautology. This tautology is called the **Law of the Excluded Middle.**

Example: Prove that the proposition ~ [p ∧ (~p)] is a tautology.

Solution: The truth table is given below:

P	~q	P ∧ (~ p)	~ [p ∧ (~ p)]
T	F	F	T
F	T	F	T

Since last column contains 'T' everywhere therefore the given proposition is a tautology.

Example: Show that ((p ∨ ~ q) ∧ (~ p ∨ ~ q)) ∨ q is a tautology be definition

Solution: ((p ∨ ~ q) ∧ (~ p ∨ ~ q)) ∨ q

  ⇒ ((p ∧ ~ p) ∨ ~ q) ∨ q (distributive law)

  ⇒ (F ∨ ~ q) ∨ q

  ⇒ (~ q) ∨ q   (as p ∧ ~ p = F)

  ⇒ T

Example: Prove that (~ p ∨ q) ∧ (p ∧ ~ q) is a contradiction.

Solution: Truth table for the given proposition

p	q	~p	~q	~p ∨ q	p ∧ ~ q	(~ p ∨ q) ∧ (p ∧ ~q)
T	T	F	F	T	F	F
T	F	F	T	F	T	F
F	T	T	F	T	F	F
F	F	T	T	T	F	F

Since F appears in every row of the last column, therefore the given proposition is contradiction.

8.6.4 Equivalence of Formulae

Two statements A and B in variable P_1 ...... P_n ($n \geq U$) are said to be equivalent if they acquire the same truth values for all interpretation, *i.e.*, they have identical truth values.

8.7 Laws

1. *Law of Identity:* Under this law the symbol used will carry the same sense through out the specified problems.

2. *Law of the Excluded middle:* Which express that a statement is either true or false.

3. *Law of Non-contradiction:* It expresses that no statement is true and false simultaneously.

8.8 Predicated Logic

Well formed formula of Predicate

The statement "*x* is greater than >" has two parts the first part the variable *x* is the subject of the statement the second part the predicate "is greater then >" refers to a property that the subject of the statement can have.

$P(x) \rightarrow$ Proposition function P at q.

Quantifiers: When all the variables in a propositional function are assigned values, the resulting statement has a truth value. However, there is another important way called quantification to create a proposition from a propositional function. There are two types of quantification namely **universal quantification and existential quantification.**

Many mathematically statement assert that a property is true for all values of a variable in a particular domain, called the universe of discourse. Such a statement is expressed using a universal quantification. The universal quantification of a propositional function is the proposition that asserts that $P(x)$ is true for all values of x in the universe of discourse. The universe of discourse specifies the possible values of the variable x.

Arguments: An argument (denoted by the symbol $\vdash$ which is called trunstile) is a sequence of propositions that purport to imply another propositions.

The sequence of propositions serving as evidence will be called, the **premises,** and the proposition inferred will be called the **conclusion.**

An arguments is valid if and only if whenever the conjunction of the premises is true, the conclusion is also true. If we let p_1, p_2, p_3 be the premises and p_4 the conclusion then argument p_1, $p_2, p_3, \dashv p_4$ will be valid if and only if whenever $p_1 \wedge p_2 \wedge p_3$ is true, p_4 is also. We can reduce this to the conditional $\Rightarrow$ as follows:

Del^n : If $p_1, p_2, \dots, p_n$ are premises and p is a conclusion then the argument $p_1, p_2, \dots, p_n \vdash p$ is valid if and only if $p_1 \wedge p_2 \wedge \dots p_n \Rightarrow p$ is true for all combinations of truth values of $p_1 p_2 \dots p_n$ and p.

In other word in order to decide whether an argument is valid, use the conjunction of evidences as the antecedent of conditional of which the conclusion of the argument is the consequent and see whether or not a tautology results.

If $p_1 \wedge p_2 \wedge \dots \wedge p_n \Rightarrow p$ is not a tautology then the argument $p_1 \dots p_n \vdash$ is invalid.

Example: "If the labour market is perfect then the wages all persons in a particular employment will be equal. But it is always the case that wages for such persons are not equal therefore the labour market is not perfect". Test the validity of this argument.

Solution: In the given case let

        p_1 : "The labour market is perfect"

        p_2 : "Wages of all persons in a particular employment will be equal"

      $\sim p_2$: Wages for such persons are not equal.

      $\sim p_1$: The labour market is not perfect.

The premise are $p_1 \Rightarrow p_2, \sim p_2$ and the conclusion is $\sim p_1$

The argument $p_1 \Rightarrow p_2 \sim p_2 \vdash \sim p_1$

valid if and only if $(p_1 \Rightarrow p_2) \Rightarrow \sim p_2 \Rightarrow \sim p_1$ is a tautology.

We construct the truth tables as below:

p_1	p_2	$\sim p_1$	$\sim p_2$	$p_1 \Rightarrow p_2$	$p_1 \Rightarrow p_2 \wedge \sim p_2$	$p_1 \Rightarrow p_2 \wedge \sim p_2 \Rightarrow \sim p_1$
T	T	F	F	T	F	T
T	F	F	T	F	F	T
F	T	T	F	T	F	T
F	F	T	T	T	T	T

is follows that $p_1 \Rightarrow p_2 \wedge \sim p_2 \Rightarrow \sim p_1$ is a tautology

Hence the argument is valid.

Example: Test the validity of the following argument. "If Ashok wins then Ram will be happy. If Kamal wins Raju will be happy. Either Ashok will win or Kamal will win. However if Ashok wins, Raju will not be happy and if Kamal wins Ram will not be happy. So Ram will be happy if and only if Raju is not happy."

Solution: Here let

       p_1 : Ashok wins

       p_2 : Ram is happy

       p_3 : Kamal wins

       p_4 : Raju is happy

The premises are $p_5 : p_1 \Rightarrow p_2$, $p_6 : p_3 \rightarrow p_4$, $p_7 : p_1 \vee p_2$

       $p_8 : p_1 \Rightarrow \sim p_4$
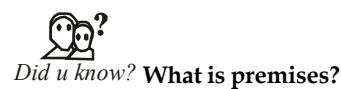
       $p_9 : p_3 \Rightarrow \sim p_2$

The conclusion is $p_1 : p_2 \Leftrightarrow \sim p_4$

The above argument is valid if $p_1 : p_2 \Leftrightarrow \sim p_4$ is a tautology so we construct the truth table.

1	2	3	4	5	6	7	8	9	10	11	12	13
p_1	p_2	p_3	p_4	$p_1 \Rightarrow p_2$	$p_3 \Rightarrow p_4$	$p_1 \vee p_3$	$\sim p_4$ or	$p_1 \Rightarrow \sim p_4$	$p_3 \Rightarrow \sim p_2$	$P_2 \Rightarrow \sim p_4$	$p_1 \Rightarrow p_2 \wedge p_3 \Rightarrow p_4 \wedge (p_1 \wedge p_3) \Rightarrow p_1 \Rightarrow \sim p_4 \wedge p_3 \Rightarrow p_2$	$(12) \Rightarrow (13)$
T	T	T	T	T	T	F	F	F	F	F	F	T
T	T	T	T	T	T	F	F	T	F	F	F	T
T	F	T	F	F	F	T	T	T	T	F	F	T
T	F	F	F	F	T	T	T	T	T	F	F	T
F	T	T	T	T	T	T	F	T	F	F	F	T
F	T	F	T	T	T	F	F	T	T	F	F	T
F	F	T	F	T	F	T	F	T	T	F	F	T
F	F	F	F	T	F	F	T	T	T	F	F	T

Since the given statement is a tautology. Hence the argument is valid.

?
Did u know? **What is premises?**

Since the given statement is a tautology. Hence the argument is valid.

8.9 Summary

- "A proposition is a statement that is either true or false but not both".

- The statement "It is not the case that P" is another proposition, called the negation of P. The negation of P is denoted by GP.

- The propositions "P or q" denoted by $P \vee q$, is the proposition that is false when P and q are both false and true otherwise the proposition $P \vee q$ is called the disjunction of P and q.

- The proposition "P and 'V'" denoted by $P \wedge q$, is the proposition that is true when both P and q are true and is false otherwise and denoted by "$\wedge$". It is called conjunction.

- The implication $P \rightarrow q$ is the proposition that is false, when P is true and q is false and true otherwise. In this implication P is called the hypothesis (or antecedent or premise) and q is called conclusion (or consequence).

- A statement formula (expression involving prepositional variable) that is neither a tautology nor a contradiction is called a contingency.

- Two statements A and B in variable P_1 ...... P_n $(n \geq U)$ are said to be equivalent if they acquire the same truth values for all interpretation, i.e., they have identical truth values

- A statement that is true for all possible values of its propositional variable is called a tautology. A statement that is always false is called a contradiction and a statement that can be either true or false depending on the truth values of its propositional variables is called a contingency.

8.10 Keywords

Duality: Any two formulas A and A* are said to be duals of each other if one can be obtained from the other by replaying $\wedge$ by $\vee$ and $\vee$ by $\wedge$.

Proposition: A proposition is a statement that is either true or false but not both

Quantifiers: When all the variables in a propositional function are assigned values, the resulting statement has a truth value.

Tautology: A statement that is true for all possible values of its propositional variable is called a tautology.

Truth Table: A truth table displays the relationship between the truth values of propositions

8.11 Self Assessment

1. Let P be "It is cold" and let q be "It is training". Give a simple verbal sentence which describes each of the following statements.

   (a) $\neg P$ (b) $s (s^1 + 1)$

   (c) $P \vee q$ (d) $q \vee \neg P$

2. Let P be "Ram reads Hindustan Times", let q be "Ram reads Times of India," and let r be "Ram reads NBT! write each of the following in symbolic form:

    (a)   Ram reads Hindustan Times on Times of India not NBT.

    (b)   Ram reads Hindustan Times and Times of India, on he does not read Hindustan Times and NBT.

    (c)   It is not true that Ram reads Hindustan Times but not NBT.

    (d)   It is not true that Ram reads NBT on Times of India but not Hindustan Times.

3. Determine the truth value of each of the following statements :

    (a)   4 + 2 = 5 and 6 + 3 = 9   (b)   3 + 2 = 5 and 6 + 1 = 7

    (c)   4 + 5 = 9 and 1 + 2 = 4   (d)   3 + 2 = 5 and 4 + 7 = 11

4. Find the truth table of $\neg P \wedge q$.

5. Verify that the propostion $P \vee \neg (P \vee q)$ is tautalogy.

6. Show that the propostions $\neg (P \wedge q)$ and $\neg P \vee \neg q$ are logically equivalent.

7. Rewrite the following statements without using the conditonal :

    (a)   If it is cold, he means a hat.

    (b)   If productivity increases, then wages rise.

8. Determine the contraposition of each statement :

    (a)   If John is a poet, then he is poor;

    (b)   Only if he studies more then he will pass the test.

9. Show that the following arguments is a fallacy :

    $P \rightarrow q, \neg P + \neg q$.

10. Write the negation of each statement as simply as possible.

    $P \rightarrow q, \neg q + \neg P$.

11. Let A = {1, 2, 3, 4, 5}. Determine the truth table value of each of the following statements :

    (a) $(\exists x \in A) (x + 3 = 10)$ (b) $(\forall \times \in A) (x + 3 > 10)$

    (c) $(\exists x \in A) (x + 3 < 5)$ (d) $(\forall \times \in A) (x + 3 \leq 7)$

12. Determine the truth table value of each of the following statements where U = {1, 2, 3} is the universal set :

    (a) $\exists x \, \forall y, x^2 < y + 1$;   (b) $\forall x \, \exists y, x^2 + y^2 < 12$

    (c) $\forall x \, \forall y, x^2 + y^2 < 12$

13. Negate each of the following statements :

    (a) $\exists x \, \forall y, P(x, y)$ (b) $\exists x \, \forall y, P(x, y)$

    (c) $\exists x \, \exists y \, \exists z, P(x, y, z)$

14. Let P(x) denote the sentence "x + 2 > 5". State whether on not P(x) is a propositional function on each of the following sets :

    (a)   N, the set of positive integers;

    (b)   M = (– 1, – 2, – 3, ...)

    (c)   C, the set of complex numbers.

15. Negative each of the following statements :

    (a)   All students live in the domintonics

    (b)   All Mathematics majors are values.

    (c)   Some students are 25 (years) on older.

8.12 Review Questions

1. Consider the conditional proposition P ↓ q. The simple propositions q → P, ¬ P → ¬ q and ¬ q → ¬ P are called respectively the converse inverse and contra positive of the conditional P → q. Which if any of these proposition are logically equivalent to P → q?

2. Write the negation of each statement as simple as possible.

    (a)   If she works, she will down money.

    (b)   He swims if and only if the water is warm.

    (c)   If it shows, then they do not drive the can.

3. Verify that the proposition (P ∧ q) ∧ ¬ (P ∨ q) is a contradiction.

4. Let P denote the is rich and let q denote "He is happy". Write each statement in symbolic form using P and q. Note that "He is poor" and he is unhappy" are equivalent to. ¬P and ¬q respectively.

    (a)   If he is rich, then he is unhappy

    (b)   He is neither rich nor happy

    (c)   It is necessary to be poor in order to be happy.

    (d)   To be poor is to be unhappy.

5. Find the truths tables for :

    (a)   P ∨ ¬ q   (b)   P ∧ ¬ q

6. Show that :

    (a)   P ∧ q logically implies P ↔ q

    (b)   P → ¬ q does not logically imply P ↔ q.

7. Let A = {1, 2, ..., 9, 10}. Consider each of the following sentences. If it is a statement, then determine its truth value. If it is a propositional function, determine its truth set.

    (a)   (∀ x∈A) (∀ y∈A) (x + y < 14)   (b)   (∀ y∈A) (x + y < 14)

    (c)   (∀ x∈A) (∀ y∈A) (x + y < 14)   (d)   (∃y∈A) (x + y < 14)

8   Negative each of the following statement :

    (a)   If the teacher is absent, then some students do not complete their homework.

    (b)   All the students completed their homework and the teacher is present.

    (c)   Some of the students did not compute their homework on teacher is absent.

Answers: Self Assessment

1. In each case, translate ∧, ∨ and ~ to read "and", "or", and "It is false that" on "hat", respectively and then simplify the English sentence.

    (a)   It is not cold

    (b)   It is cold and training

    (c)   It is cold on it is training

    (d)   It is training or it is cold.

2. Vise ∨ for "an", ∨ for "and", (or, its logical equivalent, "but"), and ¬ for "not" (negation).

    (a)   (P ∨ q) ∧ ¬ r

    (b)   (P ∧ q) ∨ ¬ (P ∧ r)

    (c)   ¬ (P ∧ ¬ r)

    (d)   ¬ [(r ∨ q) ∧ ¬ P]

3. The statement "P and q" is true only when both sub statements are true. Thus

    (a)   False

    (b)   True

    (c)   False

    (d)   True

4.

P	q	¬ P	¬ P ∧ q
T	T	F	F
T	F	F	F
F	T	T	T
F	F	T	F

5. Construct the truth table of P ∨ ¬ (P ∧ q). Since the truth values of P ∨ ¬ (P ∧ q) is T for all values of P and q, the proposition is a tautology.

P	q	P ∨ q	¬ (P ∧ q)	P ∨ ¬ (P ∨ q)
T	T	T	F	T
T	F	F	T	T
F	T	F	T	T
F	F	F	T	T

6. Construct the truth tables for ¬ (P ∧ q) and ¬ P ∨ ¬ q. Since the truth tables are the same, the propostions ¬ (P ∧ q) and ¬ P ∧ q are logically equivalent and we can write

    ¬ (P ∧ q) ≡ P ∨ ¬ q

P	q	P ∧ q	¬ (P ∧ q)	P	q	¬ P	q	¬ P ∧ ¬ q
T	T	T	F	T	T	F	F	F
T	F	F	T	T	F	F	T	T
F	T	F	T	F	T	T	F	T
F	F	F	T	F	F	T	T	T

7. Recall that "If P then q" is equivalent to "Not P or q"; that is, P → q ≡ ¬ P ∨ q. Hence,

    (a) It is not cold on he means a hot.

    (b) Productivity does not increase on wages rise.

8. (a) The contrapositive of P → q is ¬ q → ¬ P. Hence the contrapositive of the given statement is

         If John is not poor, then he is not a poet.

    (b) The given statement is equivalent to "If Marc Passes the test, then he studied." Hence its contrapostive is

    If Marc does not study, then he will not pass the test.

9. Construct the truth table for [(P → q) ∧ ¬ P] → ¬ q.

    Since the proposition [(P → q) ∧ ¬ P] → ¬ q is not a tautology, the argument is a fallacy. Equivalently, the argument is a fallacy since in third line of the truth table P → q and ¬ P are true but ¬ q is false.

P	q	P → q	¬ P	(P → q) ∧ ¬ P	¬ q	[(P → q) ∧ ¬ P] → ¬ q
T	T	T	F	F	F	T
T	F	F	F	F	T	T
F	T	T	T	T	F	F
F	F	T	T	T	T	T

Construct the truth table for [(P → q) ∧ ¬ q] → P. Since the proposition [(P → q) ∧ ¬ q] → ¬ P is a tautalogy, the argument is valid.

P	q	(P → q)	∧ ¬ q] →	¬ P						
T	T	T	T	T	F	F	T	T	F	T
T	F	T	F	F	F	T	F	T	F	T
F	T	F	T	T	F	F	T	T	T	F
F	F	F	T	F	T	T	F	T	T	F
Step		1	2	1	3	2	1	4	2	

10. Construct the truth tables of the premises and conclusion as shown below. Now, P → ¬ q, r → q, and r and true simultaneously only in the fifth now of the table, where P is also true. Hence, the argument is valid.

	P	q	r	P → q	r → q	q
1	T	T	T	F	T	F
2	T	T	F	F	T	F
3	T	F	T	T	F	F
4	T	F	F	T	T	F
5	F	T	T	T	T	T
6	F	T	F	T	T	T
7	F	F	T	T	T	T
8	F	F	F	T	T	T

11. (a) False. For no number in A is a solution to x + 3 = 10.

    (b) True. For every number in A satisfies x + 3 < 10.

(c) True. For if $x_0 = 1$, then $x_0 + 3 <$, i.e., 1 is a solution.

(d) False. For if $x_0 = 5$, then $x_0 + 3$ is not less than on equal.

(e) In other words, 5 is not a solution to the given condition.

12. (a) True. For if x = 1, then 1, 2 and 3 are all solutions to $1 < y + 1$.

(b) True. For each x_0, let y = 1; then $x_0^2 + 12$ is a true statement.

(c) False. For if $x_0 = 2$ and $y_0 = 3$, then $x_0^2 + y_0^2 < 12$ is not a true statement.

13. Use $\neg\, \forall\, x\, P(x) \equiv \exists\, x \neg P(x)$ and $\neg\, \exists\, x\, P(x) \equiv \forall\, x\, P(x)$;

(a) $\neg\, (\exists x, \forall\, y, P(x, y)) \equiv \forall\, x, \exists y, \neg\, P(x, y)$

(b) $\neg\, (\forall\, x, \forall\, y, P(x, y)) \equiv \exists x, \exists y, \neg\, P(x, y)$

(c) $\neg\, (\exists y\, \exists x\, \forall\, z, P(x, y, z)) \equiv \forall\, y, \forall\, x\, \exists z, \neg\, P(x, y, z)$

14. (a) Yes.

(b) Although P(x) is false for every element in M, P(x) is. still al propostional function on M.

(c) No. Note that 2i + 2 > 5 does not have meaning in other words in equalities are not defined for complex number.

15. (a) At least one student does not live in the donmitonic.

(b) At least one Mathematics major is female.

(c) None of the students is 25 on older.

8.13 Further Readings

Book

Anuranjan Misra, *Discrete Mathematics*, Acme Learning Pvt Ltd.

Online links

http://en.wikipedia.org/wiki/Truth_table

http://en.wikipedia.org/wiki/Logical_equivalence

Unit 9: Hasse Diagrams and Posets

CONTENTS

Objectives

Introduction

Objectives

After studying this unit, you will be able to:

- Understand partially ordered sets

- Know about hasse diagram

- State the meaning of posets

- Describe consistent enumeration

Introduction

In this unit, we shall discuss the ordered set, partially ordered set and lattices, which is a special kind of an ordered set.

Partially Ordered Set (on Poset)

A relation R on a set S is called a partial order on partial ordering, if it is

1. *Reflexive*: For any $x \in$ S, xRx or $(x, x) \in$ R

2. *Anti symmetric*: If aRb and bRa then $a = b$

3. *Transitive*: If aRb and aRc then aRc.

A set S together with a partial orderning R is called a partially ordered set (or poset) and it is denoted by (S, R). For example, the relation $\leq$ (less than or equal to) on the set N of positive integer. The partial order relation is usually denoted by $\leq$ and $a \leq b$ is real as a precedes b.

Example: Show that the relation ≥ (greater than or equal to) is a partial orderning on the set of integers.

Solution: We have $a \geq a$ for every integer a. Therefore, the relation ≥ is reflexive. Also $a \geq b$ and $b \geq a$ implies $a = b$. Therefore, ≥ is antisymmetric. Finally $a \geq b$ and $b \geq c$ implies $a \geq c$. Therefore, is ≥ transitive. Hence ≥ is a partial ordering on the set of integers.

Example: Consider P(S) as the power set, *i.e.*, the set of all subsets of a given set S. Show that the inclusion relation ⊆ is a partial ordering on the power set P(S).

Solution: Since

1. $A \subseteq A$ for all $A \subseteq S$, ⊆ is reflexive.

2. $A \subseteq B$ and $B \subseteq A \Rightarrow A = B$, ⊆ is antisymmetric.

3. $A \subseteq B$ and $B \subseteq A \Rightarrow A \subseteq C$, ⊆ is transitive.

It follow that ⊆ is a partial ordering on P(S) and (P(S), ⊆) is a poset

9.1 Totally Ordered (or Linearly Ordered) Set

Did u know? Lexicographical Order

If every two elements of a poset (S, ≥) are comparable then S is called a totally (or linearly) ordered set and the relation ≤ is called a total order or partial order. A totally ordered set is also called a chain.

Example: 1.   Consider the set N of positive integers ordered by divisibility. Then 21 and 7 are comparable since 7/21 on the other hand, 3 and 5 are non comparable since neither 3/5 non 5/3. Thus N is hot line only ordered by divisibility. Observe that A = {2, 6, 12, 36} is a linearly ordered subset of N, since 2/6, 6/12 and 12/36.

2.   The set N of positive integers with the usual order ≤ is linearly ordered and hence every ordered subset of N is also line only ordered.

3.   The power set P(A) of a set A with two or bone elemens is not linearly ordered by set inclusion. For instance suppose *a* and *b* belong to A. Then {*a*} and {*b*} are non comparable. Observe that the empty set φ, {*a*} and A do form a linearly ordered subset of P(A), since $\phi \subseteq \{A\} \subseteq A$. Similarly, φ, {*b*} and A form a linearly ordered subset of P(A).

Well ordered set: A poset (S, ≤) is called well ordered set if on is a total ordering and every non-empty subset of S has a least element.

Power sets and order: There are a number of ways to define an order relation on the cartesian product of given ordered sets. Two of these ways follow :

1.   *Product order :* Suppose S and T are ordered sets. Then the following, is an order relation on the product set S × T called the product order :

$$(a, b) \leq (a', b') \quad \text{if} \quad a \subseteq a' \text{ and } b \leq b'$$

2. *Lexicographical order* **:** Suppose S and T are line only ordered sets. Then the following is an order relation on the product set S × T, called the Lexicographical or dictionary order:

$$(a, b) < (a', b') \quad \text{if} \quad a < b \text{ on if } a = a' \text{ and } b < b'$$

This order can be extended to $S_1 \times S_2 \times ... \times S_n$ as follows :

$(a_1, a_2, ..., a_n) < (a'_1, a'_2, ..., a'_n)$ if $a_i = a'_1$ for $i = 1, 2, ..., K – 1$ and $a_K < a'_K$.

Example: $(3, 5) < (4, 8)$ is lexicographic ordering constructed from the usual relation $\leq$ on Z.

Kleene closure and order: Let A be a (non empty) linearly ordered alphabet. Recall that A*, called the Kleene closure of A, cosists of all words W on A, and $|W|$ denotes the length W. Then the following are two order relations on A*.

1. *Alphabetical (Lexicographical order)* **:** The order is no dobut familar with the usual alphabetical ordering of A*, *i.e.*,

    a. $\lambda < W$, where λ is the empty word and W is any non empty word.

    b. Suppose $u = av'$ and V = bv' are distinct non empty words where $a, b \in$ A and $u', v' \in$ A*, then :

    $$u < v \quad \text{if} \quad |u| < |v| \text{ or if } |u| = |v| \text{ but } u \text{ precedes } v \text{ alphabetically.}$$

For example, "to" precedes "and" since $|to| = 2$ but $|and| = 3$. However, "an" precedes "to" since they have the same length; but "an" preceds "to" alphabetically. This order is also called the full semigroup order.

Minimal and maximal elements: Let (S, $\leq$) be a poset. An element a in S is called a minimal element of S if there is no element b in such that $a < b$. Similarly, an element a in S is called maximal element in S if there is no element b is S such that $b < a$. Minimal and maximal elements can be spotted easily by a Hasse diagram.

Supremum and infimum: Let A be a subset of a partially ordered set S. An element M in S is called an upper bound of A in M succeeds every element of A, *i.e.*, if for every x in A, we have $x \leq$ M.

If an upper bound of A preceds every upper bound of A then it is called the supremum of A and it is denoted by sup (A).

Similarly an element M in a poset S is called a lower bound of a subset A of S if m preceds every element of A, *i.e.*, for every x in A we have :

$$M \leq x$$

If a lower bound of A succeeds every other lower bound of A then it is called the infimum of A and it is denoted by inf (A).

Example: Find the supremum and infimum of {1, 2, 4, 5, 10} if thing exist in the poset (Z⁺, 1).

Solution: Is the only lower bound for the given set. Therefore 1 is the infimum of the given set. 20 is the least upper bound of the given set. Therefore 20 is the supremum of the given set.

9.2 Properties of Posets

An element x of a poset (X, R) is called *maximal* if there is no element $y \in$ X satisfying $x <$ R y. Dually, x is *minimal* if no element satisfies $y <$ R x.

In a general poset there may be no maximal element, or there may be more than one. But in a finite poset there is always at least one maximal element, which can be found as follows: choose any element x; if it is not maximal, replace it by an element y satisfying $x <$ R y; repeat until a

maximal element is found. The process must terminate, since by the irreflexive and transitive laws the chain can never revisit any element. Dually, a finite poset must contain minimal elements.

An element x is an *upper bound* for a subset Y of X if $y \in$ R x for all $y \leq Y$. *Lower bounds* are defined similarly. We say that x is a *least upper bound* or *l.u.b.* of Y if it is an upper bound and satisfies $x <$ R x' for any upper bound x'. The concept of *a greatest lower bound* or *g.l.b.* is defined similarly.

A *chain* in a poset (X, R) is a subset C of X which is totally ordered by the restriction of R (that is, a totally ordered subset of X). An *antichain* is a set A of pairwise incomparable elements.

Infinite posets (such as $\mathbb{Z}$), as we remarked, need not contain maximal elements. *Zorn 's Lemma* gives a sufficient condition for maximal elements to exist:

Let (X, R) be a poset in which every chain has an upper bound. Then X contains a maximal element.

As well known, there is no "proof of Zorn's Lemma, since it is equivalent to the Axiom of Choice (and so there are models of set theory in which it is true, and models in which it is false). Our proof of the existence of maximal elements in finite posets indicates why this should be so: the construction requires (in general infinitely many) choices of upper bounds for the elements previously chosen (which is form a chain by construction).

The *height* of a poset is the largest cardinality of a chain, and its *width* is the largest cardinality of an antichain. We denote the height and width of (X, R) by $h(X)$ and $w(X)$ respectively (suppressing as usual the relation R in the notation).

In a finite poset (X, R), a chain C and an antichain A have at most one element in common. Hence the least number of antichains whose union is X is not less than the size $h(X)$ of the largest chain in X. In fact there is a partition of X into $h(X)$ antichains. To see this, let A_1 be the set of maximal elements; by definition this is an antichain, and it meets every maximal chain. Then, let A_2 be the set of maximal elements in $X \backslash A_1$, and iterate this procedure to find the other antichains.

There is a kind of dual statement, harder to prove, known as

Dilworth's Theorem

Let (X, R) be a finite poset. Then there is a partition of X into $w(X)$ chains.

An *up-set* in a poset (X, R) is a subset Y of X such that, if $y \in Y$ and $y \leq$ R z, then $z \in Y$. The set of minimal elements in an up-set is an antichain. Conversely, if A is an antichain, then

$$\uparrow (A) = \{x \in X : a <$$ R $x \text{ for some } a \in A\}$$

is an up-set. These two correspondences between up-sets and antichains are mutually inverse; so the numbers of up-sets and antichains in a poset are equal.

Down-sets are, of course, defined dually. The complement of an up-set is a down-set; so there are equally many up-sets and down-sets.

Task   "Infinite posets need not contain elements" Comment.

9.3 Hasse Diagrams

Let x and y be distinct elements of a poset (X, R), We say that y *covers* x. if $[x, y]_R = \{x,y\}$, that is, $x <$ R y but no element z satisfies $x <$ R $z <$ R y. In general, there may be no pairs x and y such that y covers x (this is the case in the rational numbers, for example). However, locally finite posets are determined by their covering pairs:

Proposition 1: Let (X, R) be a locally finite poset, and $x, y \in$ X. Then $x <$ R y if and only if there exist elements $z_0,..., z_n$ (for some non-negative integer n) such that $z_0 = x$, $z_n = y$, and z_{i+1} covers z_{i+1} for $i = 0, ..., n – 1$.

The *Hasse diagram* of a poset (X, R) is the directed graph whose vertex set is X and whose arcs are the covering pairs (*x, y*) in the poset. We usually draw the Hasse diagram of a finite poset in the plane in such a way that, if *y* covers *x,* then the point representing *y* is higher than the point representing *x.* Then no arrows are required in the drawing, since the directions of the arrows are implicit.

Figure 9.1: A Hasse Diagram

For example, the Hasse diagram of the poset of subsets of {1, 2, 3} is shown in Figure 9.1.

Linear Extensions and Dimension

One view of a partial order is that it contains partial information about a total order on the underlying set. This view is borne out by the following theorem. We say that one relation *extends* another if the second relation (as a set of ordered pairs) is a subset of the first.

Figure 9.2: A Crown

Theorem 2: Any partial order on a finite set X can be extended to a total order on X.

This theorem follows by a finite number of applications of the next result.

Proposition 2: Let R be a partial order on a set X, and let *a, b* be incomparable elements of X. Then there is a partial order R' extending R such that (*a, b*) ∈ R' (that is, *a < b* in the order R').

A total order extending R in this sense is referred to as a *linear extension* of R. (The term "linear order" is an alternative for "total order".)

This proof does not immediately shows that every infinite partial order can be extended to a total order. If we assume Zorn's Lemma, the conclusion follows. It cannot be proved from the Zermelo-Fraenkel axioms alone (assuming their consistency), but it is strictly weaker than the Axiom of Choice, that is, the Axiom of Choice (or Zorn's Lemma) cannot be proved from the Zermelo-Fraenkel axioms and this assumption. In other words, assuming the axioms consistent, there is a model in which Theorem 2 is false for some infinite poset, and another model in which Theorem 2 is true for all posets but Zorn's Lemma is false.

The theorem gives us another measure of the size of a partially ordered set. To motivate this, we use another model of a partial order. Suppose that a number of products are being compared using several different attributes. We regard object *a* as below object *b* if *b* beats *a* on every attribute. If each beats the other on some attributes, we regard the objects as being incomparable.

This defines a partial order (assuming that each attribute gives a total order). More precisely, given a set S of total orders on X, we define a partial order R on X by $x < Ry$ if and only if $x < Sy$ for every $s \in$ S. In other words, R is the intersection of the total orders in S.

Theorem 3: Every partial order on a finite set X is the intersection of some set of total orders on X.

Now we define the *dimension* of a partial order R to be the smallest number of total orders whose intersection is R. In our motivating example, it is the smallest number of attributes which could give rise to the observed total order R.

The *crown* on *2n* elements $a_1, ..., a_n, b_1, ..., b_n$ is the partial order defined as follows: for all indices $i \neq j$, the elements a_i and a_j are incomparable, the elements b_i and b_j are incomparable, but $a_i < b_j$, and for each *i*, the elements a_i and b_i are incomparable. Figure 9.2 shows the Hasse diagram of the 6-element crown.

Now we have the following result:

Proposition 3: The crown on *2n* elements has dimension *n*.

9.4 The Möbius Function

Lef R be a partial order on the finite set X. We take any linear order extending $\mathbb{R}$, and write $X = \{x_1, ..., x_n\}$, where $x_1 < ... < x_n$ (in the linear order S): this is not essential but is convenient later.

The *incidence algebra A*(R) of R is the set of all functions $f: X \times X \to \mathbb{R}$ which satisfy $f(x, y) = 0$ unless $x < R y$ holds. We could regard it as a function on R, regarded as a set of ordered pairs. Addition and scalar multiplication are defined pointwise; multiplication is given by the rule

$$(fg)(x, y) \ = \ \sum_z f(x,z)g(z,y)$$

If we represent *f* by the $n \times n$ matrix A_f with (*i*, *f*) entry $f(x_i, x_j)$, then this is precisely the rule for matrix multiplication. Also, if $x \not< R y$, then there is no point *z* such that $x \leq R z$ and $z \leq R y$, and so $(fg)(x, y) = 0$. Thus, A(R) is closed under multiplication and does indeed form an algebra, a subset of the matrix algebra $M_n(\mathbb{R})$. Also, since *f* and *g* vanish on pairs not in R, the sum can be restricted to the interval $[x, y]_R = \{z : x < R z < R y\}$:

Incidentally, we see that the (*i*, *j*) entry of A_f is zero if $i > j$, and so consists of upper triangular matrices. Thus, an element $f \in A$ (R) is invertible if and only if $f(x,x) \neq 0$ for all $x \in X$.

The *zeta-function* ζ_R is the matrix representing the relation R as defined earlier; that is, the element of A(R) defined by

$$\zeta_R(x, y) \ = \ \begin{cases} 1 & \text{if } x \leq Ry, \\ 0 & \text{otherwise} \end{cases}$$

Its inverse (which also lies in A(R)) is the *Möbius function* μ_R of R. Thus, we have, for all $(x,y) \in$ R,

$$\sum_{x \in |x,y|_R} \mu(x,z) \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise} \end{cases}$$

This relation allows the Möbius function of a poset to be calculated recursively. We begin with $\mu_R(x, x) = 1$ for all $x \in X$. Now, if $x < R y$ and we know the values of $\mu(x, z)$ for all $z \in [x, y]_R \setminus \{y\}$, then we have

$$\mu_R(x, y) \ = \ \sum_{z \in |x,y|_R \{y\}} \mu(x,z)$$

In particular, $\mu_R(x, y) = -1$ if y covers x.

The definition of the incidence algebra and the Möbius function extend immediately to locally finite posets, since the sums involved are over intervals $[x, y]_R$.

The following are examples of Möbius functions.

1.   The subsets of a set:

$$\mu(A, B) = (-1)^{|B/A|} \text{ for } A \subseteq B;$$

2.   The subspaces of a vector space $V \subseteq GF(q)^n$:

$$\mu(U, W) = (-1)^k q^{\binom{k}{2}} \quad \text{for} U \subseteq W, \text{ where } k = \dim U - \dim W.$$

3.   The (positive) divisors of an integer n:

$$\mu(a, b) = \begin{cases} (-1)^r & \text{if } \dfrac{b}{a} \text{ is the product of } r \text{ distinct primes;} \\ 0 & \text{otherwise} \end{cases}$$

In number theory, the classical Möbius function is the function of one variable given by $\mu(n) = \mu(1, n)$ (in the notation of the third example above).

The following result is the *Möbius inversion* for locally finite posets. From the present point of view, it is obvious.

Theorem 4: $f = g\,\zeta \Leftrightarrow g = f\mu$. Similarly, $f = \zeta g \Leftrightarrow g = \mu f$.

Example: Suppose that f and g are functions on the natural numbers which are related by the identity $f(n) = \Sigma_{d|n} g(d)$. We may express this identity as $f = g\zeta$, where we consider f and g as vectors and where ζ is the zeta function for the lattice of positive integer divisors of n. Theorem 7 implies that $g = f\mu$, or

$$g(n) = \sum_{d|n} \mu(d,n)\, f(d) = \sum_{d|n} \mu\left(\frac{d}{n}\right) f(d)$$

which is precisely the classical Möbius inversion.

Example: Suppose that f and g are functions on the subsets of some fixed (countable) set X which are related by the identity $f(A) = f(A) = \Sigma_{B \subseteq A} g(B)$. We may express this identity as $f = \zeta g$ where ζ, is the zeta function for the lattice of subsets of X. Theorem 7 implies that $g = \mu f$, or

$$g(A) = \sum_{B \supseteq A} \mu(A, B)\, f(B) = \sum_{B \supseteq A} (-1)^{|B \setminus A|} f(B)$$

which is a rather general form of the inclusion/exclusion principle.

Case Study

Show that the posets given below are lattices obtain the Hasse diagrams.

(*i*) $(S_6, |)$ (*ii*) $(S_8, |)$ (*iii*) $(S_{24}, |)$ (*iv*) $(S_{30}, |)$

where S_n is the set of all divisions of n and $|$ denotes division.

Solution: We have S_6 = {1, 2, 3, 6}

If we take any two elements of S_6 then their lower bound and upper bound will also be in S_6. Therefore, $(S_6, 1)$ will be a lattice.

Similarly,

$$S_8 \;=\; \{1, 2, 4, 8\}$$

$$S_{24} \;=\; \{1, 2, 3, 4, 6, 8, 12, 24\}$$

$$S_{30} \;=\; \{1, 2, 3, 5, 6, 10, 15, 30\}$$

We can show easily that there are lattices. The Hasse diagram of these lattices are given below :

9.5 Summary

- If every two elements of a poset (S, ≥) are comparable then S is called a totally (or linearly) ordered set and the relation ≤ is called a total order or partial order. A totally ordered set is also called a chain.

- An element x of a poset (X, R) is called *maximal* if there is no element $y \in X$ satisfying $x < R$ y. Dually, x is *minimal* if no element satisfies $y < R$ x.

- Let x and y be distinct elements of a poset (X, R), We say that y *covers x.* if $[x, y]_R = \{x,y\}$, that is, $x < R$ y but no element z satisfies $x < R$ $z < R$ y.

- Lef R be a partial order on the finite set X. We take any linear order extending R, and write X = $\{x_1, \ldots, x_n\}$, where $x_1 < \ldots < x_n$ (in the linear order S): this is not essential but is convenient later.

9.6 Keywords

Hasse Diagrams: A Hasse diagram is a type of mathematical diagram used to represent a finite partially ordered set, in the form of a drawing of its transitive reduction.

Partially Ordered Sets: A partially ordered set (or poset) formalizes the intuitive concept of an ordering, sequencing, or arrangement of the elements of a set.

9.7 Self Assessment

1. Show that the inclusion $\subseteq$ relation is a partial ordering on the power set of a set S.

2. Draw the Hasse diagram for the partial ordering {[A, B], A $\subseteq$ B} on the power set P(S) for S = {1, 2, 3}.

3. Prerequisites in a college is a partial ordering of available classes. We say A < B if course A is prerequisite for course B. We consider the maths courses and third prerequisites as given below :

Course	Prerequisite
Math I	None
Math II	Math I
Math III	Math I
Math IV	Math III
Math V	Math II
Math VI	Math V
Math VII	Math II, Math III
Math VIII	Math VI, Math IV

    Draw the Hasse diagram for the partial ordering of these courses.

4. Find the lower and upper bound of the subset {1, 2, 3} and {1, 3, 4, 6} in the poset with the Hasse diagram given below:

5. Let N = {1, 2, 3, ...} be ordered by divisibility. State whether the following subsets are totally ordered—

    (i)   {2, 5, 24}   (ii)   {3, 5, 15}

    (iii)   {5, 15, 30}   (iv)   {2, 4, 8, 32}

6. Show that every finite lattice L is bounded.

7. Show that the poset ({1, 2, 3, 4, 5} 1) is not a lattice.

8. Show that the poset ({1, 3, 6, 12, 24} 1) is a lattice.

9.   Determine whether the posets represented by the following Hasse diagrams are lattices: **Notes**

          (i)   (ii)   (iii)

10.   Write the dual of the following:

       (i) $(a \wedge b) \vee a = a \wedge (b \vee a)$

       (ii) $(a \wedge b) \vee c = (a \vee c) \wedge (b \vee c)$

11.   Give an example of an infinite lattice L with finite length.

12.   Let A = {1, 2, 3, 4, 5, 6} be ordered as in the figure.

       (a)   Find all minimal and maximal elements of A.

       (b)   Does A have a first or last element?

       (c)   Find all linearly ordered subsets of A, each of which contains at least three elements.

13.   Let B = {a, b, c, d, e, f} be ordered as in the figure.

       (a)   Find all minimal and maximal elements of B.

       (b)   Does B have a first on last element?

       (c)   List two and find the number of consistent environmentations of B into the set {1, 2, 3, 4, 5, 6}.

14. Let C = {1, 2, 3, 4} be ordered as shown in the figure. Let L(C) denote the collection of all non empty linearity ordered. Subsets of C ordered by set inclusion. Draw a diagram L(C)

15. State whether each of the following is true or false and, if it is a false, give a counter example.

    (a) If a poset S has only one maximal element a then *a* is a last element.

    (b) If a finite poset S has only one maximal element *a*, then *a* is a last element.

    (c) If a linearly ordered set S has only one maximal element *a* then *a* is a last element.

16. Let S = {*a, b, c, d, e*} be ordered as in the figure.

    (a) Find all minimal and maximal elements of S.

    (b) Does S have any first on last element?

    (c) Find all subsets S in which C is a minimal element.

    (d) Find all subsets of S in which C is a first element.

    (e) List all linearly ordered subsets with three on more elements.

9.8 Review Questions

1.. Let S = {*a, b, c, d, e, f*} be ordered as shown in the figure

   (a) Find all minimal and maximal elements of S.

   (b) Does S have any first on last element.

   (c) List all linearity ordered subsets with three on more elements

2. Let M = {2, 3, 4, ...} and let $M^2 = M \times M$ be ordered as follows :

   $$(a, b) \leq (c, d) \qquad \text{If} \quad a/c \text{ and } b \leq d$$

   Find all minimal and maximal elements of M × M.

3. Consider the set R of real numbers with the usual order ≤. Let A = {$x : x \in Q$ and $5 < x^3 < 27$}.

    (a) Is A bounded above an below ?

    (b) Do Sup (A) and inf (A) exist ?

4. Suppose the union 5 of the sets A = {$a_1, a_2, a_3, ...$}, B = {$b_1, b_2, b_3, ...$} C = {$c_1, c_2, c_3, ...,$} is ordered as follows :

    S = {A; B; C} = {$a_1, a_2, ..., b_1, b_2, ..., c_1, c_2, ...$}

    (a) Find all limit elements of S.

    (b) Show that S is not isomorphic to N = {1, 2, ...} with the usual order ≤.

5. Let A = {a, b, c} be linearly ordered by $a < b < c$ and let N = {1, 2, ...} be given the usual order ≤.

    (a) Show that S = {A; N} is isomorphic to N.

    (b) Show that S′ = {N; A} is not isomorphic to N.

6. Show that the divisibility relation | is a partial ordering on the set of positive integers.

7. Find the maximal and minimal elements of the poset {2, 4, 5, 10, 12, 20, 25}

8. Find the greatest lower bound and least upper bound of the set {1, 2, 4, 5, 10} in the poset (Z^+, 1).

9. Show that the poset {1, 2, 4, 8, 16} is a lattice.

10. Draw the Hasse diagram for divisibility on {1, 2, 3, 5, 7, 11, 13}.

11. Show that the poset given below are lattices.

    (i) (S_9, 1),   (ii) (S_{12}, 1),   (iii) (S_{15}, 1)

    Also obtain their Hasse diagram.

Answers: Self Assessment

1. For each subset A of S we have A ⊆ A. Therefore ⊆ is reflexive. Also A ⊆ B and B ⊆ A ⇒ A = B.

    Therefore ⊆ is antisymmetric. Finally A ⊆ B and B ⊆ C implies A ⊆ C. Therefore ⊆ is transitive. Hence ⊆ is a partial ordering and so (P(S), ⊆) is a poset.

2. The required Hasse diagram can be obtained from the digraph of the given poset by deleting all the loops and all the edges that occur from transitive property, *i.e.*, (φ, {1, 23}), (φ, {1, 33}), (φ, (2, 33)), (φ {1, 2, 33}), ({1}{1, 2, 33}), ({2}, {1, 2, 33}) and deleting arrows which will be as given below:

3. We put Math I in the bottom of the diagram as it is the only course with no prerequisite. Math II and Math III only require Math I, so we have Math I << Math II and Math I << Math III.

Therefore, we draw lines starting upward from Math I to Math II and from Math I to Math III. Continuing this process we draw the complete Hasse diagram of the given partial ordering as shown below:

4. The upper bounds of such set {1, 2, 3} are 5, 6, 7, 8, 9 and its lower bound is only 1. The upper bound of {1, 3, 4, 6} are 6, 8 and 9 and its lower bound is only 1.

5. (i) Here 2 divides 6 and 6 divides 24, therefore, the given set is totally ordered.

(ii) Here 3 and 5 are not comparable. Therefore, the given set is not totally ordered.

(iii) Here 5 divides 15 which divides 30. Therefore, the given set is totally ordered.

(iv) Here 2 divides 4, 4 divides 8 and 8 divides 32. Therefore, the given set is totally ordered.

6. Let $L = \{a_1, a_2, ..., a_n\}$ is a finite lattice then $a_1 \wedge a_2 ... \wedge a_n$ and $a_1 \vee a_2 ... \vee a_n$ are lower and upper bounds of L respectively. Thus the lower and upper bounds of L exists. Hence the finite lattice is bounded.

7. 2 and 3 have no upper bound consequently they do not have a least upper bound. Hence the given set is not a lattice.

8. Every two elements of the given poset have a least upper bound as well as a greatest lower bound which are longer and smaller elements respectively. Hence the given poset is a lattice.

9. (i) The poset represented by the given Hasse diagram is a lattice as every pair of elements of it has a least upper bound as well as greatest lower bound.

(ii) The poset represented by the given Hasse diagram is not a lattice because the elements 2 and 3 have no least upper bound. The elements 4, 5 and 6 are the upper bounds but none of the sets three elements preceeds other two with respect to the ordering of the given poset.

(iii) The poset represented by the given Hasse diagram is a lattice as every pair of elements of it has a least upper bound as well as a greatest lower bound.

10. The dual will be obtained by replacing $\wedge$ by $\vee$ and $\vee$ by $\wedge$. Thus dual statements will be

(i) $(a \vee b) \wedge a = a \vee (b \wedge a)$

(ii) $(a \vee b) \wedge c = (a \wedge c) \vee (b \wedge c)$

11.   Let L = {0, 1, a_1, a_2, ..., a_n, ...} be ordered

i.e., $0 < a_n < 1 \ldots n \in \mathbb{N}$

Then N is an infinite lattice having finite length.

12.   (a)   Minimal 4 and 6 maximal 1 and 2

      (b)   First, none; last, none

      (c)   {1, 3, 4}, {1, 3, 6} {2, 3, 4}, {2, 3, 6}, {2, 5, 6}

13.   (a)   Minimal *d* and *f*; maximal *a*

      (b)   First, none; last, a

      (c)   There are eleven: *dfebca, dfecba, dfceba, fdebca, fdecda, fdceba, fedbca, fcdeba, fecdba, fcedba*

14.

15.   (a)   False, Example: NU{*a*} where 1 << *a*, and N ordered by ≤.

      (b)   True (*c*) True

16.   (a)   Minimal, *a* : maximal *d* and *e*

      (b)   First *a*; last, none

      (c)   Any subset which contains *c* and omits *a*; that is : *c, cb, cd, ce, cbd, cbe, cde, cbde*,

      (d)   *abd acd, ace*

9.9 Further Readings

Books   Béla Bollobás, *Modern graph theory*, Springer

          Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph theory, Combinatorics, and Algorithms*, Birkhäuser

Online links   http://en.wikipedia.org/wiki/Hasse_diagram

                http://en.wikipedia.org/wiki/Partially_ordered_set

Unit 10: Supremum and Infimum

CONTENTS

Objectives

Introduction

Objectives

After studying this unit, you will be able to:

- Know supremum
- Understand infimum
- Get aware isomorphic order Sets
- Describe lattices
- Explain bounded lattices

Introduction

We have already seen two equivalent forms of the completeness axiom for the reals: the montone convergence theorem and the statement that every Cauchy sequence has a limit. The second of these is useful as it doesn't mention the order relation < and so applies to the complex numbers for instance. (It turns out that the set of complex numbers is also complete in the Cauchy sense.) This web page describes a third approach due to Dedekind. This uses the order relation < of $\mathbb{R}$ but applies to arbitrary subsets of $\mathbb{R}$ rather than sequences. There are a number of places (including results about the radius of convergence for power series, and several results in more advanced analysis) where Dedekind's approach is particularly helpful.

10.1 Supremum

In mathematics, given a subset S of a partially ordered set T, the **supremum** (sup) of S, if it exists, is the least element of T that is greater than or equal to each element of S. Consequently, the supremum is also referred to as the **least upper bound** (**lub** or **LUB**). If the supremum exists, it may or may not belong to S. If the supremum exists, it is unique.

Suprema are often considered for subsets of real numbers, rational numbers, or any other well-known mathematical structure for which it is immediately clear what it means for an element to be "greater-than-or-equal-to" another element. The definition generalizes easily to the more abstract setting of order theory, where one considers arbitrary partially ordered sets.

The concept of supremum is not the same as the concepts of *minimal* upper bound, maximal element, or greatest element.

Did u know? Is supremum is in a precise sense dual to the concept of an infimum?

10.1.1 Supremum of a Set of Real Numbers

In analysis, the **supremum** or **least upper bound** of a set S of real numbers is denoted by sup(S) and is defined to be the smallest real number that is greater than or equal to every number in S. An important property of the real numbers is its completeness: every nonempty subset of the set of real numbers that is bounded above has a supremum that is also a real number.

Examples:

$$\sup\{1, 2, 3\} = 3$$

$$\sup\{x \in \mathbb{R} : 0 < x < 1\} = \sup\{x \in \mathbb{R} : 0 \leq x \leq 1\} = 1$$

$$\sup\left\{(-1)^n - \frac{1}{n} : n \in \mathbb{N}*\right\} = 1$$

$$\sup\{a + b : a \in A \text{ and } b \in B\} = \sup(A) + \sup(B)$$

$$\sup\{x \in \mathbb{Q} : x^2 < 2\} = \sqrt{2}$$

In the last example, the supremum of a set of rationals is irrational, which means that the rationals are incomplete.

One basic property of the supremum is

$$\sup\{f(t)+g(t):t\in A\}\le\sup\{f(t):t\in A\}+\sup\{g(t):t\in A\}$$

for any functionals *f* and *g*.

If, in addition, we define sup(*S*) = – ∞ when *S* is empty and sup(*S*) = + ∞ when *S* is not bounded above, then every set of real numbers has a supremum under the affinely extended real number system.

$$\sup\mathbb{Z}\ =\ \infty$$

$$\sup\phi\ =\ -\infty$$

If the supremum belongs to the set, then it is the greatest element in the set. The term *maximal element* is synonymous as long as one deals with real numbers or any other totally ordered set.

To show that *a* = sup(*S*), one has to show that *a* is an upper bound for *S* and that any other upper bound for *S* is greater than *a*.

Notes One could alternatively show that *a* is an upper bound for *S* and that any number less than *a* is not an upper bound for *S*.

10.1.2 Suprema within Partially Ordered Sets

Least upper bounds are important concepts in order theory, where they are also called joins (especially in lattice theory). As in the special case treated above, a supremum of a given set is just the least element of the set of its upper bounds, provided that such an element exists.

Formally, we have: For subsets *S* of arbitrary partially ordered sets (*P*, ≤), a **supremum** or **least upper bound** of *S* is an element *u* in *P* such that

1. $x \le u$ for all *x* in *S*, and

2.   for any *v* in *P* such that $x \le v$ for all *x* in *S* it holds that $u \le v$.

Thus the supremum does not exist if there is no upper bound, or if the set of upper bounds has two or more elements of which none is a least element of that set. It can easily be shown that, if *S* has a supremum, then the supremum is unique (as the least element of any partially ordered set, if it exists, is unique): if u_1 and u_2 are both suprema of *S* then it follows that $u_1 \le u_2$ and $u_2 \le u_1$, and since ≤ is antisymmetric, one finds that $u_1 = u_2$.

If the supremum exists it may or may not belong to *S*. If *S* contains a greatest element, then that element is the supremum; and if not, then the supremum does not belong to *S*.

The dual concept of supremum, the greatest lower bound, is called infimum and is also known as meet.

If the supremum of a set *S* exists, it can be denoted as sup(*S*) or, which is more common in order theory, by ∨ *S*. Likewise, infima are denoted by inf(*S*) or ∧ *S*. In lattice theory it is common to use the infimum/meet and supremum/join as binary operators; in this case a ∨ b = sup {a, b} (and similarly for infima).

A complete lattice is a partially ordered set in which *all* subsets have both a supremum (join) and an infimum (meet).

In the sections below the difference between suprema, maximal elements, and minimal upper bounds is stressed. As a consequence of the possible absence of suprema, classes of partially ordered sets for which certain types of subsets are guaranteed to have least upper bound become especially interesting. This leads to the consideration of so-called completeness properties and to numerous definitions of special partially ordered sets.

Did u know? What is meet?

10.1.3 Comparison with other Order Theoretical Notions

Greatest Elements

The distinction between the supremum of a set and the greatest element of a set may not be immediately obvious. The difference is that the greatest element must be a member of the set, whereas the supremum need not. For example, consider the set of negative real numbers (excluding zero). This set has no greatest element, since for every element of the set, there is another, larger, element. For instance, for any negative real number x, there is another negative real number $x/2$, which is greater. On the other hand, every real number greater than or equal to zero is certainly an upper bound on this set. Hence, 0 is the least upper bound of the negative reals, so the supremum is 0. This set has a supremum but no greatest element.

In general, this situation occurs for all subsets that do not contain a greatest element.

Caution If a set does contain a greatest element, then it also has a supremum given by the greatest element.

Maximal Elements

For an example where there are no greatest but still some maximal elements, consider the set of all subsets of the set of natural numbers (the powerset). We take the usual subset inclusion as an ordering, i.e. a set is greater than another set if it contains all elements of the other set. Now consider the set S of all sets that contain at most ten natural numbers. The set S has many maximal elements, i.e. elements for which there is no greater element. In fact, all sets with ten elements are maximal. However, the supremum of S is the (only and therefore least) set which contains all natural numbers. One can compute the least upper bound of a subset A of a powerset (i.e. A is a set of sets) by just taking the union of the elements of A.

Minimal Upper Bounds

Finally, a set may have many minimal upper bounds without having a least upper bound. Minimal upper bounds are those upper bounds for which there is no strictly smaller element that also is an upper bound. This does not say that each minimal upper bound is smaller than all other upper bounds, it merely is not greater. The distinction between "minimal" and "least" is only possible when the given order is not a total one. In a totally ordered set, like the real numbers mentioned above, the concepts are the same.

As an example, let *S* be the set of all finite subsets of natural numbers and consider the partially ordered set obtained by taking all sets from *S* together with the set of integers **Z** and the set of positive real numbers **R+**, ordered by subset inclusion as above. Then clearly both **Z** and **R+** are greater than all finite sets of natural numbers.

Notes   Neither is **R+** smaller than **Z** nor is the converse true: both sets are minimal upper bounds but none is a supremum.

10.1.4 Least-upper-bound Property

The **least-upper-bound property** is an example of the aforementioned completeness properties which is typical for the set of real numbers. This property is sometimes called **Dedekind completeness**.

If an ordered set *S* has the property that every nonempty subset of *S* having an upper bound also has a least upper bound, then *S* is said to have the least-upper-bound property. As noted above, the set **R** of all real numbers has the least-upper-bound property. Similarly, the set **Z** of integers has the least-upper-bound property; if *S* is a nonempty subset of **Z** and there is some number *n* such that every element *s* of *S* is less than or equal to *n*, then there is a least upper bound *u* for *S*, an integer that is an upper bound for *S* and is less than or equal to every other upper bound for *S*. A well-ordered set also has the least-upper-bound property, and the empty subset has also a least upper bound: the minimum of the whole set.

An example a set that *lacks* the least-upper-bound property is **Q**, the set of rational numbers. Let *S* be the set of all rational numbers *q* such that $q^2 < 2$. Then *S* has an upper bound (1000, for example, or 6) but no least upper bound in **Q**. For suppose $p \in \mathbf{Q}$ is an upper bound for *S*, so $p^2 > 2$. Then $q = (2p+2)/(p + 2)$ is also an upper bound for *S*, and $q < p$. (To see this, note that $q = p - (p^2 - 2)/(p + 2)$, and that $p^2 - 2$ is positive.) Another example is the Hyperreals; there is no least upper bound of the set of positive infinitesimals.

There is a corresponding 'greatest-lower-bound property'; an ordered set possesses the greatest-lower-bound property if and only if it also possesses the least-upper-bound property; the least-upper-bound of the set of lower bounds of a set is the greatest-lower-bound, and the greatest-lower-bound of the set of upper bounds of a set is the least-upper-bound of the set.

If in a partially ordered set *P* every bounded subset has a supremum, this applies also, for any set *X*, in the function space containing all functions from *X* to *P*, where $f \leq g$ if and only if $f(x) \leq g(x)$ for all *x* in *X*. For example, it applies for real functions, and, since these can be considered special cases of functions, for real *n*-tuples and sequences of real numbers.

Task   Analyse the difference between greatest-lower-bound and greatest-lower-bound

---

Example: Let A = 1 - 1 n : n ∈ ℕ. Then A has no largest element, i.e., max A doesn't exist, but sup A = 1 since 1 is an upper bound and any c < 1 is less than some 1 - 1 n by the Archimedean property. Note that sup A is not an element of A .

Let A = 1 n : n ∈ ℕ . Then A has no smallest element, i.e., min A doesn't exist, but inf A = 0 since 0 is a lower bound and any d > 0 is greater than some 1 n by the Archimedean property. Note that inf A is not an element of A .

Let A = [2 , 3] . In this case A does have largest and smallest elements, and sup A = 3 and inf A = 2.

Let A be the empty set. Then by convention every b $\in$ $\mathbb{R}$ is both an upper bound and a lower bound for A . So A does not have least upper bound or greatest lower bound.

Let A = $\mathbb{N}$. Then A does not have any upper bound, by the Archimedean property. But A does have a lower bound, such as -1 . The greatest lower bound is determined by your convention on the set of natural numbers. If you prefer 0 $\in$ $\mathbb{N}$ then inf $\mathbb{N}$ = 0 . Otherwise you will have inf $\mathbb{N}$ = 1.

Let A = $\mathbb{Z}$. Then A does not have any upper bound nor any lower bound, by the Archimedean property again.

Theorem 1: (Completeness of reals, supremum form)

Let A $\subseteq$ $\mathbb{R}$ be non-empty and bounded above. Then there is a least upper bound of A , sup A .

Proof:

This is proved by a variation of a proof of an earlier result that every real number has a monotonic sequence of rationals converging to it.

We start by defining a sequence (a n) using induction on n . Start with any a 0 $\in$ $\mathbb{Q}$ with some x $\in$ A such that a 0 < x . This just uses the assumption that A is nonempty. Now inductively assume that a n is defined and a n < x for some x $\in$ A , i.e., a n is not an upper bound for A . If a n + 1 n is an upper bound for A then let a n + 1 = a n . Otherwise, let a n + 1 = a n + k n where k = K 0 - 1 $\in$ $\mathbb{N}$ and K 0 is chosen to be the least natural number such that a n + K 0 n$\geq$ x for all x $\in$ A . Such a K exists since A is bounded above by some b $\in$ $\mathbb{R}$ and we need only take K $\in$ $\mathbb{N}$ so that a n + K n $\geq$ b , using the Archimedean Property. So since there is some such K there must be a least such number, K 0 .

By construction, (a n) is a nondecreasing sequence of rationals and bounded above by b , an upper bound for A . It follows that (a n) converges to some l $\in$ $\mathbb{R}$. We shall show that this l is the least upper bound of A .

Subproof

Suppose l is not an upper bound of A .

Subproof

Then there is x $\in$ A such that l < x . But this gives a contradiction, for if n $\in$ $\mathbb{N}$ is such that x - l > 1 n we consider the n th stage of the construction, where we chose a n + 1 = a n + k n with k greatest so that there is some y $\in$ A with a n + k n < y. But a n + k n $\leq$ l < x - 1 n so a n + k + 1 n $\leq$ l + 1 n < x contradicting this choice of k .

Thus l is after all an upper bound. To see it is the least upper bound, we again suppose otherwise.

Subproof

Suppose l is not the least upper bound of A .

Subproof

Then there is some m < l such that every x $\in$ A has x $\leq$ m . This again is impossible. Since a n $\rightarrow$ l , there must be some a n with m < a n $\leq$ l . (To see this, put ε = l - m and a n with a n - l < ε). But by construction of a n there is always some x $\in$ A with a n < x. Therefore m < x and m is not after all an upper bound for A.

This completes the proof.

Theorem 2: (Completeness of reals, infimum form)

Let $A \subseteq \mathbb{R}$ be non-empty and bounded below. Then there is a greatest lower bound of A , inf A.

Proof

Let $c \in \mathbb{R}$ be a lower bound for A and let B = - x : x $\in$ A . Then b = - c is an upper bound for B and hence by the previous result B has a least upper bound, l . It follows easily that - l is a greatest lower bound for A , for if x $\in$ A then l $\geq$ - x as - x $\in$ B so - l $\leq$ x , and if m > - l then - m < l so - m is not an upper bound for B , so there is x $\in$ B with x > - m hence - x < m and clearly - x $\in$ A .

These results are equivalent to the monotone convergence theorem. To see this, suppose (a n) is a bounded nondecreasing sequence. Then let A = a n : n $\in$ $\mathbb{N}$. By the fact that the sequence is bounded and by completeness theorem above, l = sup A exists, and it is a nice exercise to show that a n $\to$ l as n $\to$ ∞ .

10.2 Infimum

In mathematics, the **infimum** (plural **infima**) of a subset *S* of some partially ordered set *T* is the greatest element of *T* that is less than or equal to all elements of *S*. Consequently the term **greatest lower bound** (also abbreviated as **glb** or **GLB**) is also commonly used. Infima of real numbers are a common special case that is especially important in analysis. However, the general definition remains valid in the more abstract setting of order theory where arbitrary partially ordered sets are considered.

Did u know? Is infimum is in a precise sense dual to the concept of a supremum.

10.2.1 Infima of Real Numbers

In analysis the infimum or greatest lower bound of a subset *S* of real numbers is denoted by inf(*S*) and is defined to be the biggest real number that is smaller than or equal to every number in *S*. If no such number exists (because *S* is not bounded below), then we define inf(*S*) = –∞. If *S* is empty, we define inf(*S*) = ∞ (see extended real number line).

An important property of the real numbers is that every set of real numbers has an infimum (any bounded nonempty subset of the real numbers has an infimum in the non-extended real numbers).

Examples:

$$\inf \{1, 2, 3\} \ = \ 1.$$

$$\inf \{x \in \mathbb{R} : 0 < x < 1\} \ = \ 0$$

$$\inf \{x \in \mathbb{Q} : x^3 > 2 \} \ = \ \sqrt[3]{2}$$

$$\inf \{(-1)^n + 1/n : n = 1, 2, 3, ...\} \ = \ = -1.$$

If a set has a smallest element, as in the first example, then the smallest element is the infimum for the set. (If the infimum is contained in the set, then it is also known as the minimum). As the last three examples show, the infimum of a set does not have to belong to the set.

The notions of infimum and supremum are dual in the sense that

$$\inf (S) \ = \ -\sup (-S),$$

where

$$-S \ = \ \{-s/s \in S\}.$$

10.2.2 Infima in Partially Ordered Sets

The definition of infima easily generalizes to subsets of arbitrary partially ordered sets and as such plays a vital role in order theory. In this context, especially in lattice theory, greatest lower bounds are also called **meets**.

Formally, the *infimum* of a subset S of a partially ordered set $(P, \leq)$ is an element a of P such that

1. $a \leq x$ for all x in S, (a is a lower bound) and

2. for all y in P, if for all x in S, $y \leq x$, then $y \leq a$ (a larger than any other lower bound).

Any element with these properties is necessarily unique, but in general no such element needs to exist. Consequently, orders for which certain infima are known to exist become especially interesting. The dual concept of infimum is given by the notion of a *supremum* or *least upper bound*. By the duality principle of order theory, every statement about suprema is thus readily transformed into a statement about infima.

Task Make a difference between supremum and least upper bound.

10.3 Isomorphic Sets

In abstract algebra, an **isomorphism** is a bijective map f such that both f and its inverse f^{-1} are homomorphisms, i.e., *structure-preserving* mappings. In the more general setting of category theory, an **isomorphism** is a morphism $f: X \to Y$ in a category for which there exists an "inverse" $f^{-1}: Y \to X$, with the property that both $f^{-1}f = \mathrm{id}_X$ and $ff^{-1} = \mathrm{id}_Y$.

Informally, an isomorphism is a kind of mapping between objects that shows a relationship between two properties or operations. If there exists an isomorphism between two structures, we call the two structures **isomorphic**.

Notes In a certain sense, isomorphic structures are **structurally identical**, if you choose to ignore finer-grained differences that may arise from how they are defined.

10.3.1 Purpose

Isomorphisms are studied in mathematics in order to extend insights from one phenomenon to others: if two objects are isomorphic, then any property which is preserved by an isomorphism and which is true of one of the objects, is also true of the other. If an isomorphism can be found from a relatively unknown part of mathematics into some well studied division of mathematics, where many theorems are already proved, and many methods are already available to find answers, then the function can be used to map whole problems out of unfamiliar territory over to "solid ground" where the problem is easier to understand and work with.

An operation-preserving isomorphism

Suppose that on these sets *X* and *Y*, there are two binary operations $*$ and $\lozenge$ which happen to constitute the groups $(X, *)$ and $(Y, \lozenge)$. Note that the operators operate on elements from the domain and range, respectively, of the "one-to-one" and "onto" function f. There is an isomorphism from *X* to *Y* if the bijective function $f : X \to Y$ happens to produce results, that sets up a correspondence between the operator $*$ and the operator $\lozenge$.

$$f(u) \lozenge f(v) = f(u * v)$$

for all *u*, *v* in *X*.

10.3.2 Applications

In abstract algebra, two basic isomorphisms are defined:

1. Group isomorphism, an isomorphism between groups

2. Ring isomorphism, an isomorphism between rings. (Note that isomorphisms between fields are actually ring isomorphisms)

Just as the automorphisms of an algebraic structure form a group, the isomorphisms between two algebras sharing a common structure form a heap. Letting a particular isomorphism identify the two structures turns this heap into a group.

In mathematical analysis, the Laplace transform is an isomorphism mapping hard differential equations into easier algebraic equations.

In category theory, Iet the category *C* consist of two classes, one of *objects* and the other of morphisms. Then a general definition of isomorphism that covers the previous and many other cases is: an isomorphism is a morphism $f : a \to b$ that has an inverse, i.e. there exists a morphism $g : b \to a$ with $fg = 1_b$ and $gf = 1_a$. For example, a bijective linear map is an isomorphism between vector spaces, and a bijective continuous function whose inverse is also continuous is an isomorphism between topological spaces, called a homeomorphism.

In graph theory, an isomorphism between two graphs *G* and *H* is a bijective map f from the vertices of *G* to the vertices of *H* that preserves the "edge structure" in the sense that there is an edge from vertex *u* to vertex *v* in *G* if and only if there is an edge from $f(u)$ to $f(v)$ in *H*. See graph isomorphism.

In mathematical analysis, an isomorphism between two Hilbert spaces is a bijection preserving addition, scalar multiplication, and inner product.

In early theories of logical atomism, the formal relationship between facts and true propositions was theorized by Bertrand Russell and Ludwig Wittgenstein to be isomorphic. An example of this line of thinking can be found in Russell's Introduction to Mathematical Philosophy.

In cybernetics, the Good Regulator or Conant-Ashby theorem is stated "Every Good Regulator of a system must be a model of that system". Whether regulated or self-regulating an isomorphism is required between regulator part and the processing part of the system.

Did u know? What is automorphism?

10.3.3 Relation with Equality

In certain areas of mathematics, notably category theory, it is valuable to distinguish between *equality* on the one hand and *isomorphism* on the other. Equality is when two objects are "literally the same", while isomorphism is when two objects "can be made to correspond via an isomorphism". For example, the sets

$$A = \{x \in Z \mid x^2 < 2\} \text{ and } B = \{-1, 0, 1\}$$

are *equal* – they are two different presentations (one in set builder notation, one by an enumeration) of the same subset of the integers. By contrast, the sets $\{A, B, C\}$ and $\{1, 2, 3\}$ are not *equal* – the first has elements that are letters, while the second has elements that are numbers. These are isomorphic as sets, since finite sets are determined up to isomorphism by their cardinality (number of elements) and these both have three elements, but there are many choices of isomorphism – one isomorphism is

$$A \mapsto 1, B \mapsto 2, C \mapsto 3, \text{ while another is } A \mapsto 3, B \mapsto 2, C \mapsto 1,$$

and no one isomorphism is better than any other. Thus one cannot *identify* these two sets: one can choose an isomorphism between them, but any statement that identifies these two sets *depends* on the choice of isomorphism.

A motivating example is the distinction between a finite-dimensional vector space V and its dual space $V^* = \{\varphi : V \to \mathbf{K}\}$ of linear maps from V to its field of scalars $\mathbf{K}$. These spaces have the same dimension, and thus are isomorphic as abstract vector spaces (since algebraically, vector spaces are classified by dimension, just as sets are classified by cardinality), but there is no "natural" choice of isomorphism $V \simeq V^*$. If one chooses a basis for V, then this yields an isomorphism: For all $u, v \in V$,

$$v \xrightarrow{\sim} \phi_v \in V^* \text{ such that } \phi_v(u) = v^T u.$$

This corresponds to transforming a column vector (element of V) to a row vector (element of V^*) by transpose, but a different choice of basis gives a different isomorphism: the isomorphism "depends on the choice of basis". More subtly, there *is* a map from a vector space V to *its double dual* $V^{**} = \{x : V^* \to \mathbf{K}\}$ that does not depend on the choice of basis: For all $v \in V$ and $\varphi \in V^*$,

$$v \xrightarrow{\sim} x_v \in V^{**} \text{ such that } x_v(\phi) = \phi(v).$$

This leads to a third notion, that of a natural isomorphism: while V and V^{**} are different sets, there is a "natural" choice of isomorphism between them. This intuitive notion of "an isomorphism that does not depend on an arbitrary choice" is formalized in the notion of a natural transformation; briefly, that one may *consistently* identify, or more generally map from, a vector space to its double dual, $V \xrightarrow{\sim} V^{**}$, for *any* vector space in a consistent way. Formalizing this intuition is a motivation for the development of category theory.

If one wishes to draw a distinction between an arbitrary isomorphism (one that depends on a choice) and a natural isomorphism (one that can be done consistently), one may write $\approx$ for an unnatural isomorphism and $\cong$ for a natural isomorphism, as in $V \approx V^*$ and $V \cong V^{**}$. This convention is not universally followed, and authors who wish to distinguish between unnatural isomorphisms and natural isomorphisms will generally explicitly state the distinction.

Generally, saying that two objects are *equal* is reserved for when there is a notion of a larger (ambient) space which these objects live within. Most often, one speaks of equality of two subsets of a given set (as in the integer set example above), but not of two objects abstractly presented.

Notes**

For example,

The 2-dimensional unit sphere in 3-dimensional space

$$S^2 := \left\{(x, y, z) \in \mathbf{R}^3 \big| x^2 + y^2 + z^2 = 1\right\} \text{ and the Riemann sphere } \hat{C}$$

which can be presented as the one-point compactification of the complex plane $\mathbf{C} \cup \{\infty\}$ *or* as the complex projective line (a quotient space)

$$P_C^1 := \left(C^2 / \{(0, 0)\}\right) / (C^*)$$

are three different descriptions for a mathematical object, all of which are isomorphic, but which are not *equal* because they are not all subsets of a single space: the first is a subset of $\mathbf{R}^3$, the second is $\mathbf{C} \cong \mathbf{R}^2$ plus an additional point, and the third is a subquotient of $\mathbf{C}^2$

10.4 Lattices

A *lattice* is a poset (X, R) with the properties

1. X has an upper bound 1 and a lower bound 0;

2. for any two elements $x, y \in X$, there is a least upper bound and a greatest lower bound of the set $\{x, y\}$.

A simple example of a poset which is not a lattice is the poset

In a lattice, we denote the l.u.b. of $\{x, y\}$ by $x \vee y$, and the g.l.b. by $x \wedge y$. We commonly regard a lattice as being a set with two distinguished elements and two binary operations, instead of as a special kind of poset.

Lattices can be axiomatised in terms of the two constants 0 and 1 and the two operations $\vee$ and $\wedge$. The result is as follows, though the details are not so important for us. The axioms given below are not all independent. In particular, for finite lattices we do not need to specify 0 and 1 separately, since 0 is just the meet of all elements in the lattice and 1 is their join.

Properties of Lattices: Let X be a set, $\wedge$ and $\vee$ two binary operations defined on X, and 0 and 1 two elements of X. Then $(X, \vee, \wedge, 0, 1)$ is a lattice if and only if the following axioms are satisfied:

1. *Associative laws*: $x \wedge (y \wedge z) = (x \wedge y) \wedge z$ and $x \vee (y \vee z) = (x \vee y) \vee z$;

2. *Commutative laws*: $x \wedge y = y \wedge x$ and $x \vee y = y \vee x$;

3. *Idempotent laws*: $x \wedge x = x \vee x = x$;

4. $x \wedge (x \wedge y) = x = x \vee (x \wedge y)$;

5. $x \wedge 0 = 0, x \vee 1$.

A *sublattice* of a lattice is a subset of the elements containing 0 and 1 and closed under the operations $\vee$ and $\wedge$. It is a lattice in its own right.

The following are a few examples of lattices.

1. The subsets of a (fixed) set:

$$A \wedge B = A \cap B$$
$$A \vee B = A \cup B$$

The subspaces of a vector space:

$$U \wedge V = U \cap V$$

$$U \vee V = \text{span} (U \cup V)$$

The partitions of a set:

$$R \wedge T = R \cap T$$

$$R \vee T = \overline{R \cup T}$$

Here $\overline{R \cup T}$ is the partition whose classes are the connected components of the graph in which two points are adjacent if they lie in the same class of either R or T.

Distributive and Modular Lattices

A lattice is *distributive* if it satisfies the *distributive laws*

(D) $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ and $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ for x, y, z.

A lattice is *modular* if it satisfies the *modular law*

(M) $x. \vee (y \wedge z) = (x \vee y) \wedge z$ for all x, y, z such that $x \neq z$.

Figure 10.1 represents a lattice, N_5, which is not modular, as well as a modular lattice, M_3, which is not distributive.

Figure 10.1: Two lattices

N_5 M_3

Not only are N_5 and M_3 the smallest lattices with these properties, they are, in a certain sense, the only lattices with these properties. The following theorem states this more precisely.

Theorem 3: A lattice is modular if and only if it does not contain the lattice N_5 as a sublattice. A lattice is distributive if and only if it contains neither the lattice N_5 nor the lattice M_j as a sublattice.

The poset of all subsets of a set S (ordered by inclusion) is a distributive lattice: we have $0 = \theta$, $1 = S$, and l.u.b. and g.l.b. are union and intersection respectively. Hence every sublattice of this lattice is a distributive lattice.

Conversely, every finite distributive lattice is a sublattice of the lattice of sub-sets of a set. We describe how this representation works. This is important in that it gives us another way to look at posets.

Let (X, R) be a poset. Recall that an down-set in X is a subset Y with the property that, if $y \in Y$ and $z \leq R\, y$, then $z \in Y$.

Let L be a lattice. A non-zero element $x \in L$ is called *join-irreducible* if, whenever $x = y \vee z$, we have $x = y$ or $x = z$.

Theorem 4:

1. Let (X, R) be a finite poset. Then the set of down-sets. in X, with the operations of union and intersection and the distinguished elements 0 = θ and 1 = X, is a distributive lattice.

2. Let L be a finite distributive lattice. Then the set X of non-zero join-irreducible elements of L is a sub-poset of L.

3. These two operations are mutually inverse.

Meet-irreducible elements are defined dually, and there is of course a dual form of Theorem 10.

A *lattice* is a poset (X, R) with the properties

Notes For finite lattices we do not need to specify 0 and 1 separately, since 0 is just the meet of all elements in the lattice and 1 is their join.

Did u know? What is sub lattice?

10.5 Bounded Lattices

A bounded lattice is an algebraic structure $\| = (L, \wedge, \vee, 0, 1)$, such that $(L, \wedge, \vee)$ is a lattice, and the constants $0, 1 \in L$ satisfy the following:

1. for all $x \in L, x \wedge 1 = x$ and $x \vee 1 = 1$,

2. for all $x \in L, x \wedge 0 = 0$, $x \wedge 0$ and $x \wedge 0 = x$. .

The element 1 is called the upper bound, or top of $\|$ and the element 0 is called the lower bound or bottom of $\|$.

There is a natural relationship between bounded lattices and bounded lattice-ordered sets. In particular, given a bounded lattice, $(L, \wedge, \vee, 0, 1)$, the lattice-ordered set $(L, \leq)$ that can be defined from the lattice $(L, \wedge, \vee)$ is a bounded lattice-ordered set with upper bound 1 and lower bound 0. Also, one may produce from a bounded lattice-ordered set $(L, \leq)$ a bounded lattice $(L, \wedge, \vee, 0, 1)$ in a pedestrian manner, in essentially the same way one obtains a lattice from a lattice-ordered set. Some authors do not distinguish these structures, but here is one fundamental difference between them: A bounded lattice-ordered set $(L, \leq)$ can have bounded subposets that are also lattice-ordered, but whose bounds are not the same as the bounds of $(L, \leq)$; however, any subalgebra of a bounded lattice $\| = (L, \wedge, \vee, 0, 1)$ is a bounded lattice with the same upper bound and the same lower bound as the bounded lattice $\|$.

For example, let $X = \{a. b. c\}$, and let $\| = (L, \wedge, \vee, 0, 1)$ be the power set of X, considered as a bounded lattice:

1. $L = \left\{\phi, \{a\}, \{b\}, \{c\}, \{a. b\}, \{a. c\}, \{b. c\}, X\right\}$

2. $0 = \phi$ and $= X$

3. $\wedge$ is union: for $A. B \in L$, $A \vee B = A \cup B$

4. $\vee$ is intersection: for $A. B \in L$, $A \wedge B = A \cap B.$

Let $Y = \{a, b\}, ,$ and let $K = (K. \wedge. \vee. 0'. 1')$ be the power set of Y, also considered as a bounded lattice:

1. $K = \{\phi, \{a\}, \{b\}, Y\}$

2. $0' = \phi$ and $l' = Y$

3. $\vee$ is union: for $A. B \in L$, $A \vee B = A \cup B$

4. $\wedge$ is intersection: for $A. B \in L$, $A \vee B = A \cap B.$

Then the lattice-ordered set $(K, \le)$ that is defined by setting $A \le B$ iff $A \subseteq B$ is a substructure of the lattice-ordered set $(L, \le)$ that is defined similarly on L. Also, the lattice $(K, \wedge, \vee)$ is a sublattice of the lattice $(l, \wedge, \vee)$. However, the bounded lattice $K = (K, \wedge, \vee, 0', 1')$ is not a subalgebra of the bounded lattice $\parallel = (L, \wedge, \vee, 0, 1)$, precisely because $l \ne l'$

Case Study

Show that every chain and $a, b, c \in$ L, we consider the cases :

(*i*) $a \le b$ or $a \le c$ and   (*ii*) $a \ge b$ or $a \ge c$

Now we shall show that distributive law is satisfied by a, b, c :

For case (*i*) we have

$a \wedge (b \vee c) = a$ and $(a \wedge b) \vee (a \wedge c) = a$

For case (*ii*), we have

$a \wedge (b \vee c) = b \vee c$ and $(a \wedge b) \vee (a \wedge c) = b \wedge c$

Thus, we have

$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$

This shows that a chain is a distributive lattice.

10.6 Summary

- In mathematics, given a subset S of a partially ordered set T, the **supremum** (sup) of S, if it exists, is the least element of T that is greater than or equal to each element of S.

- You have seen another form of the completeness axiom for the reals, which is useful in that it doesn't involve any sequences and may be applied to subsets of $"\mathbb{R}$.

- In mathematics, the **infimum** (plural **infima**) of a subset S of some partially ordered set T is the greatest element of T that is less than or equal to all elements of S.

- In a lattice, we denote the l.u.b. of $\{x, y\}$ by $x \vee y$, and the g.l.b. by $x \wedge y$. We commonly regard a lattice as being a set with two distinguished elements and two binary operations, instead of as a special kind of poset.

- A bounded lattice is an algebraic structure $L = (L. \wedge. \vee. 0. 1)$, such that $(L. \wedge. \vee)$ is a lattice.

10.7 Keywords

Automorphism: An automorphism is an isomorphism from a mathematical object to itself.

Bounded Lattice: A lattice is a partially ordered set (also called a poset) in which any two elements have a unique supremum (the elements' least upper bound; called their join) and an infimum (greatest lower bound; called their meet).

10.8 Self Assessment

1. Consider the logarithm function: For any fixed base b, the logarithm function $\log_b$ maps from the positive real numbers $\mathbf{R}^+$ onto the real numbers $\mathbf{R}$; formally:

$$\log_b : \mathbf{R}^+ \to \mathbf{R}$$

2. Consider the group $\mathbf{Z}_6$, the integers from 0 to 5 with addition modulo 6. Also consider the group $\mathbf{Z}_2 \times \mathbf{Z}_3$, the ordered pairs where the x coordinates can be 0 or 1, and the y coordinates can be 0, 1, or 2, where addition in the x-coordinate is modulo 2 and addition in the y-coordinate is modulo 3.

3. If one object consists of a set X with a binary relation R and the other object consists of a set Y with a binary relation S then an isomorphism from X to Y is a bijective function $f : X \to Y$ such that

$$S(f(u), f(v)) \Leftrightarrow R(u, v)$$

4. Show that the lattices given by the following diagrams are not distributive.

5. If X be the interval (–3, 2). Then calculate the upper bound and lower bound.

6. If X is finite set then prove sup X = max X.

7. $\qquad \sup \{1, 1/2, 1/3, 1/4, ...\}$ and $\inf \{1, 1/2, 1/3, 1/4, ...\} = 0$

$$\sup \{..., -2, -1, 0, 1, 2, ...\} = \infty \quad \text{and} \quad \inf \{..., -2, -1, 0, 1, 2, ...\} = -\infty$$

$$\sup \{-4\} \cup (0, \infty) \quad \text{and} \quad \inf \{-4\} \cup (0, \infty) = -4$$

For convenience, we define the supremum of the empty set to be $-\infty$ and the infimum of the empty set to be ∞. Calculate the limit of $f(x)$.

8. Calculate the limit if, $f(x) = 2x - 3$

9. Calculate the limit, if $f(x) = 1/x^2$

10. Calculate the limit, if $f(x) = x/|x|$

11. Calculate the limit, if $f(x) = x \sin(1/x)$

12. Calculate the limit, of sin (1/x) if $f(x) - \sin(1/x)$

13. Calculate the limit, if $f(x) = 1/x$ as $x \to \infty$.

14 The dual concept of supremum, the greatest lower bound, is called ……………….

15 An isomorphism is a kind of mapping between objects that shows a relationship between two properties or ……………….

10.9 Review Questions

1. Consider the lattice L in the figure.

   (a) Find all sublattices with five elements.

   (b) Find all join-irreducible elements and atoms.

   (c) Find complements of *a* and *b*, if they exist

   (d) Is L distributive? Complemented?

2. Consider the lattice M in the figure below :

   (a) Find join-irreducible elements

   (b) Find the atoms.

   (c) Find complements of *a* and *b* if exist

   (d) Is M distributive? Complemented?

3. Let *A* and *B* be bounded non-empty sets. Which of the following statements would be equivalent to saying that $(A) \le \inf (B)$?

   (a) For every $b \in B$ there exists an $a \in A$ such that $a \le b$.

   (b) For every $a \in A$ and every $b \in B$, we have $a \le b$.

   (c) For every $a \in A$ there exists $b \in B$ such that $a \le b$.

   (d) There exists $a \in A$ such that $a \le b$ for all $b \in B$.

(e)   There exists $b \in B$ such that $a < b$ for all $a \in A$.

(f)   There exists $a \in A$ and $b \in B$ such that $a < b$.

(g)   For every $b \in B$ and $a > 0$ there exists an $a \in A$ such that $a < b + \epsilon$.

4.   Let A and B be bounded non-empty sets such that inf $(A) \leq$ sup (B). Which of the following statements must be true?

(a)   There exists $b \in B$ such that $a \leq b$ for all $a \in A$.

(b)   For every $a > 0$ there exists $a \in A$ and $b \in B$ such that $a < b + \epsilon$.

(c)   For every $b \in B$ there exists an $a \in A$ such that $a \leq b$.

(d)   There exists $a \in A$ and $b \in B$ such that $a \leq b$.

(e)   For every $a \in A$ there exists a $b \in B$ such that $a \leq b$.

(f)   There exists $a \in A$ such that $a \leq b$ for all $b \in B$.

(g)   For every $a \in A$ and every $b \in B$, we have $a \leq b$.

5.   Let A and B be non-empty sets. Which of the following statements would be equivalent to saying that sup $(A) \leq$ inf (B)?

(a)   There exists $a \in A$ such that $a \leq b$ for all $b \in B$.

(b)   For every epsilon there exists $a \in A$ and $b \in B$ such that $a < b + \epsilon$.

(c)   For every $a \in A$ and every $b \in B$, we have $a \leq b$.

(d)   For every $a \in A$ there exists a $b \in B$ such that $a \leq b$.

(e)   There exists $a \in A$ and $b \in B$ such that $a \leq b$.

(f)   For every $b \in B$ there exists an $a \in A$ such that $a \leq b$.

(g)   There exists $b \in B$ such that $a \leq b$ for all $a \in A$.

6.   Let A and B be non-empty sets. Which of the following statements would be equivalent to saying that sup $(A) \leq$ sup (B)?

(a)   For every $a \in A$ there exists a $b \in B$ such that $a \leq b$.

(b)   For every $a \in A$ and epsilon > 0 there exists a $b \in B$ such that $a < b + \epsilon$.

(c)   There exists $a \in A$ such that $a \leq b$ for all $b \in B$.

(d)   For every epsilon there exists $a \in A$ and $b \in B$ such that $a < b + \epsilon$.

(e)   For every $a \in A$ and every $b \in B$, we have $a \leq b$.

(f)   For every $b \in B$ there exists an $a \leq b$ such that $a \in A$.

(g)   There exists $b \in B$ such that $a \leq b$ for all $a \in A$.

7.   Let A and B be non-empty sets. The best we can say about $(A \cup B)$ is that

(a)   It is the maximum of sup (A) and sup (B).

(b)   It is the minimum of sup (A) and sup (B).

(c)   It is greater than or equal to sup (A) and greater than or equal to sup (B).

(d)   It is less than or equal to sup (A) and less than or equal to sup (B).

(e)   It is strictly greater than at least one of sup (A) and sup (B).

(f)   It is equal to at least one of sup (A) and sup (B).

(g)   It is strictly greater than sup (A) and strictly greater than sup (B).

8.   Let *A* and *B* be non-empty sets. The best we can say about (A ∩ B) is that

   (a)   It is strictly less than sup (A) and strictly less than sup (B).

   (b)   It is equal to at least one of sup (A) and sup (B).

   (c)   It is less than or equal to sup (A), and less than or equal to sup (B).

   (d)   It is greater than or equal to sup (A) and greater than or equal to sup (B).

   (e)   It is greater than or equal to at least one of sup (A) and sup (B).

   (f)   It is the minimum of sup (A) and sup (B).

   (g)   It is the maximum of sup (A) and sup (B).

9.   Let *A* be a non-empty set. If sup (A) = + ∞, this means that

   (a)   There exists an $a \in A$ such that $a > M$ for every real number *M*.

   (b)   There exists a real number *M* such that $a > M$ for every $a \in A$.

   (c)   *A* is the empty set.

   (d)   There exists $a \in A$ and a real number *M* such that $a > M$.

   (e)   For every real number *M* and every $a \in A$, we have $a > M$.

   (f)   For every real number *M*, there exists an $a \in A$ such that $a > M$.

   (g)   There exists an $a \in A$ such that $a > M$ for every real number M.

10.   Let *A* be a set. If inf (A) = + ∞, this means that

   (a)   There exists an $a \in A$ such that $a > M$ for every real number M.

   (b)   For every real number *M* and every $a \in M$, we have $a > M$.

   (c)   *A* is the empty set.

   (d)   For every real number *M*, there exists an $a \in A$ such that $a > M$.

   (e)   There exists an $a \in A$ such that $a > M$ for every real number *M*.

   (f)   There exists $a \in A$ and a real number *M* such that $a > M$.

   (g)   There exists a real number *M* such that $a > M$ for every $a \in A$.

11.   Let *A* be a non-empty set. If *A* is not bounded, this means that

   (a)   For every real number *M*, there exists an $a \in A$ such that $a > M$.

   (b)   There exists an $a \in A$ such that $|a| > M$ for every real number *M*.

   (c)   sup (A) = – ∞ and inf (A) = + ∞.

   (d)   sup (A) = + ∞ and inf (A) = – ∞.

   (e)   For every positive number *M*, there exists an $a \in A$ such that $a > M$, and for every negative number –*M*, there exists $a' \in A$ such that $a' < -M$.

   (f)   There exists a real number *M* such that $|a| > M$ for every $a \in A$.

   (g)   For every real number *M*, there exists an $a \in A$ such that $|a| > M$.

12. Let A and B be bounded non-empty sets. Which of the following statements would be equivalent to saying that sup (A) = inf (B)?

    (a) For every $\in > 0$, there exists $a \in A$ and $b \in B$ such that $b - a < \in$.

    (b) There exists $a \in A$ and $b \in B$ such that $a \leq b$. However, for any $a > 0$, there does not exist $a \in A$ and $b \in B$ for which $a + \in \leq b$.

    (c) There exists a real number L such that $a \leq L \leq b$ for all $a \in A$ and $b \in B$.

    (d) For every $a \in A$ and every $b \in B$, we have $a \leq b$.

    (e) For every $a \in A$ and every $b \in B$, we have $a \leq b$. Also, for every $\in > 0$, there exists $a \in A$ and $b \in B$ such that $b - a < \in$.

    (f) For every $a \in A$ there exists a $b \in B$ such that $a + \in < b$.

    (g) For every $a \in A$ there exists a $b \in B$ such that $a + \in < b$. Also, for every $b \in B$ there exists an $a \in A$ such that $b + \in < a$.

13. Let A be a set, and let L be a real number. If sup (A) = L, this means that

    (a) There exists a sequence x_n of elements in A which are less than L, but converges to L.

    (b) There exists a sequence x_n of elements in A which converges to L.

    (c) There exists an $\in > 0$ such that every real number a between $L - \in$ and L lies in A.

    (d) L lies in A, and L is larger than every other element of A.

    (e) $a \leq L$ for every $a \in A$.

    (f) $a \leq L$ for every $a \in A$. Also, for every $\in > 0$, there exists an $a \in A$ such that $L - \in < a \leq L$.

    (g) Every number less than L lies in A, and every number greater than L does not lie in A.

14. Let A be a set, and let L be a real number. If sup $(A) \leq L$, this means that

    (a) There exists an $a \in A$ such that $a \leq L - \in$ for every $\in > 0$.

    (b) There exists an $\in > 0$ such that every element of A is less than $L - \in$.

    (c) For every $\in > 0$ there exists an $a \in A$ such that $a \leq L - \in$.

    (d) $a \leq L$ for every $a \in A$.

    (e) For every $\in > 0$ and every $a \in A$, we have $a \leq L - \in$.

    (f) Every number less than or equal to L lies in A.

    (g) $a \leq L$ for every $a \in A$. Also, for every $\in > 0$, there exists an $a \in A$ such that $L - \in < a \leq L$.

15. Let A be a set, and let L be a real number. If sup $(A) < L$, this means that

    (a) $a < L$ for every $a \in A$. Also, for every $\in > 0$, there exists an $a \in A$ such that $L - \in < a < L$.

    (b) There exists an $\in > 0$ such that every element of A is less than $L - \in$.

    (c) Every number less than or equal to L lies in A.

    (d) There exists an $a \in A$ such that $a < L - \in$ for every $\in > 0$.

    (e) For every $\in > 0$ and every $a \in A$, we have $a < L - \in$.

    (f) For every $\in > 0$ there exists an $a \in A$ such that $a < L - \in$.

    (g) $a < L$ for every $a \in A$.

Answers: Self Assessment

1. This mapping is one-to-one and onto, that is, it is a bijection from the domain to the codomain of the logarithm function. In addition to being an isomorphism of sets, the logarithm function also preserves certain operations. Specifically, consider the group $(\mathbf{R}^+, \times)$ of positive real numbers under ordinary multiplication. The logarithm function obeys the following identity:

$$\log_b(xy) = \log_b(x) + \log_b(y).$$

   But the real numbers under addition also form a group. So the logarithm function is in fact a group isomorphism from the group $(\mathbf{R}^+, \times)$ to the group $(\mathbf{R}, +)$. Logarithms can therefore be used to simplify multiplication of real numbers. By working with logarithms, multiplication of positive real numbers is replaced by addition of logs. This way it is possible to multiply real numbers using a ruler and a table of logarithms, or using a slide rule with a logarithmic scale.

2. These structures are isomorphic under addition, if you identify them using the following scheme:

   $(0,0) \to 0$

   $(1,1) \to 1$

   $(0,2) \to 2$

   $(1,0) \to 3$

   $(0,1) \to 4$

   $(1,2) \to 5$

   or in general $(a,b) \to (3a + 4b) \mod 6$. For example note that $(1,1) + (1,0) = (0,1)$ which translates in the other system as $1 + 3 = 4$. Even though these two groups "look" different in that the sets contain different elements, they are indeed **isomorphic**: their structures are exactly the same. More generally, the direct product of two cyclic groups $\mathbf{Z}_m$ and $\mathbf{Z}_n$ is isomorphic to $\mathbf{Z}_{mn}$ if and only if m and n are coprime.

3. S is reflexive, irreflexive, symmetric, antisymmetric, asymmetric, transitive, total, trichotomous, a partial order, total order, strict weak order, total preorder (weak order), an equivalence relation, or a relation with any other special properties, if and only if R is.

   For example, R is an ordering $\leq$ and S an ordering $\sqsubseteq$, then an isomorphism from X to Y is a bijective function $f : X \to Y$ such that

   $$f(u) \sqsubseteq f(v) \Leftrightarrow u \leq v.$$

   Such an isomorphism is called an *order isomorphism* or (less commonly) an *isotone isomorphism*.

   If $X = Y$ we have a relation-preserving automorphism.

4. (*i*) We have

   $$a_1 \wedge (a_2 \vee a_3) = (a_1 \wedge a_2) \wedge (a_1 \wedge a_3)$$

   But $$a_2 (a_1 \wedge a_3) = a_2 \wedge 1 = a_2$$

   $$(a_2 \wedge a_1) \vee (a_2 \wedge a_3) = 0 \vee a_3 = a_3$$

   Thus, $$a_2 \wedge (a_1 \vee a_3) \neq (a_2 \wedge a_1) \vee (a_2 \wedge a_3)$$

   Hence the given lattices is not distributive.

(*ii*)   We have

$$b_1 \wedge (b_2 \vee b_3) \; = \; b_1 \wedge 1 = b_1$$

$$(b_1 \wedge b_2) \vee (b_1 \wedge b_3) \; = \; 0 \vee 1 = 0$$

Thus, $b_1 \wedge (b_2 \vee b_3) \; \neq \; (b_1 \wedge b_2) \vee (b_1 \wedge b_3)$

Hence the given lattice is not distributive.

5.   Let X be the interval (–3, 2). Then 17 (or any other number ≥ 2) is an upper bound for X and sup X = 2. Also, – 936.5 (or any other number $\leq$ – 3) is a lower bound for X and inf X = – 3

6.   If X is a finite set then sup X = max X, the largest number contained in X, and inf X = min X, the smallest number contained in X. If is any set that has a largest element, then sup X = max X and if X is a any set that has a smallest element then inf X = min X.

7.   Let *a* be a real number and *f* be a function.

We assume that *f* is defined near *a*, that is, *f* is defined in some interval containing *a*, except possible at *a* itselt.

   For each positive number δ, let $A(\delta) \; = \; \sup\{f(x):0<|x-a|<\delta\}$ and define

$$\lim_{x \to a} \sup f(x) \; = \; \inf\{A(\delta):\delta>0\}.$$

Also,   let $B(\delta) \; = \; \inf\{f(x):0<|x-a|<\delta\}$ and define

$$\lim_{x \to a} \inf f(x) \; = \; \sup\{B(\delta):\delta>0\}$$

Since $A(\delta) \geq B(\delta)$ for each δ > 0, it is always true that

$$\lim_{x \to a} \sup f(x) \; \geq \; \lim_{x \to a} \inf f(x).$$

If it happens that

$$\lim_{x \to a} \sup f(x) \; = \; \lim_{x \to a} \inf f(x) = L$$

we say the limit of *f(x)* as *x* approaches *a* is *L* and write

$$\lim_{x \to a} f(x) \; = \; L.$$

If it happens that

$$\lim_{x \to a} \sup f(x) \; > \; \lim_{x \to a} \inf f(x)$$

we say the limit of $f(x)$ as *x* approaches a does not exist.

8.   Let *a* = 2.

Since $y = 2x - 3$ is increasing, the supermum of $2x - 3$ when $0<|x-2|<\delta$ is the value at the right endpoint of $(2-\delta,2)\cup(2, 2+\delta)$. That is, $2(2+\delta)-3=1+2\delta$. Therefore, $A(\delta) = 1 + 2\delta$ and

$$\lim_{x \to 2} \sup 2x - 3 \; = \; \inf\{1+2\delta:\delta>0\}=1.$$

Similarly, $\qquad B(\delta) = 2 - 2\delta$, and

Since $\qquad \lim\limits_{x\to 2}\sup 2x - 3 = \lim\limits_{x\to 2}\inf 2x - 3 = 1$ we conclude that

$$\lim\limits_{x\to 2} 2x - 3 = 1.$$

9. Let $a = 0$.

The function $1/x^2$ blows up near zero so $A(\delta) = \infty$ for all $\delta > 0$ and

$$\lim\limits_{x\to 0}\sup 1/x^2 = \inf\{\infty : \delta > 0\} = \infty.$$

The infimum of $1/x^2$ when $0 < |x - 2|Z < \delta$ is $B(\delta) = 1/\delta^2$ so

$$\lim\limits_{x\to 0}\inf 1/x^2 = \inf\{1/\delta^2 : \delta > 0\} = \infty.$$

Since $\qquad \lim\limits_{x\to 0}\sup 1/x^2 = \lim\limits_{x\to 0}\inf 1/x^2 = \infty$ we conclude that

$$\lim\limits_{x\to 0} 1/x^2 = \infty.$$

10. Let $a = 0$.

Since $\qquad\qquad x/|x| = 1$ whatever $x > 0, A(\delta) = 1$ for all $\delta > 0$

$$\lim\limits_{x\to 0}\sup x/|x| = \inf\{1 : \delta > 0\} = 1.$$

Since $\qquad\qquad x/|x| = -1$ whatever $x < 0, B(\delta) = -1$ for all $\delta > 0$. and

$$\lim\limits_{x\to 0}\inf x/|x| = \sup\{-1 : \delta > 0\} = -1.$$

Since $\lim\limits_{x\to 0}\sup x/|x| = 1 > -1 = \lim\limits_{x\to 0}\inf x/|x|$ we conclude that the limit of $x/|x|$ as x approaches 0 does not exist.

11. Let $a = 0$.

when $0 < |x| < \delta, x\sin(1/x) \le |x\sin(1/x)| \le \delta$ and so $A(\delta) \le \delta$. Therefore,

$$\lim\limits_{x\to 0}\sup x\sin(1/x) \le \inf\{\delta : \delta > 0\} = 0.$$

Also, when $0 < |x| < \delta, x\sin(1/x) \ge -|x\sin(1/x)| \ge -\delta$ and so $B(\delta) \ge -\delta$. Therefore,

$$\lim\limits_{x\to 0}\inf x\sin(1/x) \ge \sup\{-\delta : \delta > 0\} = 0.$$

Since $0 \le \lim\limits_{x\to 0}\inf x\sin(1/x) \le \lim\limits_{x\to 0}\sup x\sin(1/x) \le 0,$ both must equal 0. Thus,

$$\lim\limits_{x\to 0}\inf x\sin(1/x) = 0.$$

One-Sided Limits and Infinite Limits

Limit from the right—replace the condition $0 < |x - a| < \delta$ by $0 < x - a < \delta$ to define

$$\limsup_{x \to a^+} f(x), \quad \liminf_{x \to a^+} f(x), \quad \lim_{x \to a^+} f(x).$$

Limit from the left—replace the condition $0 < |x - a| < \delta$ by $0 < a - x < \delta$ to define

$$\limsup_{x \to a^-} f(x), \quad \liminf_{x \to a^-} f(x), \quad \lim_{x \to a^-} f(x).$$

Limit as $x \to \infty$ – replace the condition $0 < |x - a| < \delta$ by $x > \delta$ to define

$$\limsup_{x \to -\infty} f(x), \quad \liminf_{x \to -\infty} f(x), \quad \lim_{x \to -\infty} f(x).$$

12. Consider the limit as x approaches zero from the right. On any interval $(0, \delta)$ the maximum value of $\sin(1/x)$ is 1 and the minimum is –1 so

$$\limsup_{x \to 0^+} \sin(1/x) \quad = \quad 1 > -1 = \liminf_{x \to 0^+} \sin(1/x).$$

We conclude that the limit of $\sin(1/x)$ as x approahces zero from the right does not exist.

13. On the interval (δ, ∞), $y = 1/x$ is decreasing. The supermum is $1/\delta$ and the infimum is 0. Therefore,

$$\limsup_{x \to \infty} 1/x \quad = \quad \inf\{1/\delta : \delta > 0\} = 0$$

and

$$\liminf_{x \to \infty} 1/x \quad = \quad \sup\{0, \delta > 0\} = 0.$$

We conclude that

$$\lim_{x \to \infty} 1/x \quad = \quad 0.$$

Identities:

$$\limsup_{x \to \infty} f(x) \quad = \quad \max\left(\limsup_{x \to a^-} f(x), \limsup_{x \to a^+} f(x) \right)$$

$$\limsup_{x \to \infty} f(x) \quad = \quad \max\left(\liminf_{x \to a^-} f(x), \liminf_{x \to a^+} f(x) \right)$$

Also, $\lim_{x \to a} f(x)$ exists if and only if both one-sided limits exist and they are equal

14. infimum

15. operations

10.10 Further Readings

Books

Béla Bollobás, Modern graph theory, Springer

Martin Charles Golumbic, Irith Ben-Arroyo Hartman, Graph theory, combinatorics, and algorithms, Birkhäuser

Online links

http://web.mat.bham.ac.uk/R.W.Kaye/seqser/supinf

http://en.wikipedia.org/wiki/Lattice_(order)

http://mathworld.wolfram.com/OrderIsomorphic.html

Unit 11: Probability Theory

CONTENTS

Objectives

Introduction

Objectives

After studying this unit, you will be able to:

- Know about sample space events
- Find finite probability spaces

Introduction

A probability is a quantitative measure of uncertainty. The statistician IJ Good suggests, "The theory of probability is much older than the human species, since the assessment of uncertainty incorporates the idea of learning from experience, which most creatures do." Development of probability theory in Europe is often associated with gamblers, who pursued their interests in the famous European casinos, such as the one at Monte Carlo. It is also associated with astrology. Theory of probability has its origin in the seventeenth century. The king of France, who was fond of gambling, consulted the great mathematicians Pascal and Fermat for solving gambling problems. These two scientists tried to solve these problems using probability concepts, which laid the foundation of the probability theory. Later, during eighteenth century, great mathematicians like Bernoulli, De-Moivre, Laplace and several others made valuable contributions to the theory. There is hardly any discipline left, in which probability theory is

not used. The probability theory tries to measure the possibility of an outcome in numerical terms. Thus, probability of an outcome is a numeric measure of the possibility or chance of the occurrence of that outcome. Understanding and interpreting the problem logic is more important in probability. There could be a number of methods to solve a given problem.

11.1 Definitions of Probability

Probability of an event can be explained or defined in four different ways. Although these are four approaches usually based on the way we look at the probability, there is hardly any difference in the application of the probability theory. These approaches overlap in concepts.

Notes   Whatever the kind of probability involved, the same set of mathematical rules, theorems and postulates hold for manipulating and analyzing probability.

11.1.1 The A-priori Probability (Mathematical or Classical Probability)

This is also called as mathematical probability or objective probability or classical probability. This probability is based on the idea that certain occurrences are equally likely. For example, if unbiased dice is rolled, numbers 1 to 6 are equally likely to appear on the top face.

This definition is based on the concept of an experiment, sample space of the experiment and event space.

Experiment

An Experiment is a procedure that yields one of a given set of possible outcomes.

E.g. Experiment of rolling a dice gives a set of possible outcomes as number on top face can be from 1 to 6. No other outcome is possible. Similarly tossing a coin is an experiment with two possible outcomes, namely Heads and Tails. (We exclude a freak possibility of the coin standing on its side!)

Sample

The Sample space of the experiment is the set of all possible outcomes.

Example: In case of an experiment of rolling a dice, the sample space is {1, 2, 3, 4, 5, 6}. Also in the experiment of rolling two die, the sample space is, {(1,1), (1,2), (1,3), …(2,1), (2,2), …., (6,6)}

Event

An Event is a subset of the sample space

E.g. Getting an even number on the top face of a rolled dice is an event. In this case, the event set is, {2, 4, 6}.

Equi-probable Sample Space

Consider a random experiment, which has *n* mutually exclusive outcomes. The sample space *S* is said to be an equi-probable sample space if all the outcomes are equally likely.

Laplace's Definitions of the Probability

For an event with finitely many possible outcomes, the probability is defined as follows:

The probability of an event *E* which is a subset of a finite sample '*S*' of equally likely outcomes,

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

Another way of defining mathematical probability is as follows:

If there are *n* mutually exclusive, collectively exhaustive and equally likely outcomes of an experiment and if *m* of them are favourable to an event *E*, then the probability of occurrence of *E*, denoted by *P(E)* is defined as,

$$P(E) \quad = \quad \frac{m}{n} \quad \text{Where, } 0 \leq m \leq n \qquad \text{Thus, } P(E) \leq 1$$

Laplace's definition is more complete

This definition is based on a-priori knowledge of equally likely outcomes and that total outcomes are finite. For example, draw of cards from a shuffled pack of 52 cards, or throw of a dice, or toss of a coin. If any of the assumptions are not true, then classical definition given above does not hold. For example, toss of a biased coin, or throw of dice by 'Shakuni Mama' in the epic '*Mahabharat*'. This definition also has a serious drawback: How do we know that the outcomes are equally likely? Since, it cannot be proved mathematically or logically, this definition is not complete.

Limitations of Classical Probability

1.   It requires finite sample space.

2.   It requires equi-probable sample space.

3.   It requires a-priori knowledge of number of outcomes in a sample space as well as in the event.

Did u know? What is random experiment?

11.1.2 The A-posteriori Probability (Statistical or Empirical or Experimental probability)

This is another type of objective probability. It is also called as experimental probability. It overcomes the shortcoming of the previous definition of probability. This definition is stated as: "Suppose that an experiment, whose sample space is S, is repeatedly performed under exactly the same conditions, and if n represents sufficiently large number of trials made to see whether

an event E occurs or not and m represents the number of trials in which E is observed to occur, then the probability of event E, denoted as P(E) is defined as,

P(E) = $\lim_{n \to \infty} \dfrac{m}{n}$, provided the circumstance for trial to trial remain the same". Thus, P(E) is defined as the limiting proportion of number of times that E occurs, that is a limiting frequency. This definition also has few drawbacks. How do we know that the ratio $\dfrac{m}{n}$ will converge to some constant limiting value that will be the same every time we carry out the experiment? If we actually carry out an experiment of flipping a coin with event of getting heads, we do not observe any systematic series so as to prove mathematically that the ratio $\dfrac{m}{n}$ converges to $\dfrac{1}{2}$. Thus, this definition also has a limited use.

11.1.3 Subjective Probability (Personal Probability or Probability as a Matter of belief)

The subjective probability of an event is the degree of confidence placed in the occurrence of an event by an individual based on all evidence available to him. This is a most simple and natural interpretation of probabilities. It could be considered as a measure of the individual's belief in the statement that he or she is making. This probability depends on the personal judgment, and hence, is called as personal or subjective probability.

Example: Probability of forecast of rainfall, or the estimate of sales, or physician assessing probability of patient's recovery are some cases of subjective probability.

11.1.4 Axiomatic Definition of Probability (Modern Probability)

In earlier definitions that we have discussed, we had to make certain assumptions. However, to assume that $\dfrac{m}{n}$ will necessarily converge to some constant value every time the experiment is performed, or the event is equally likely; seem to be very complex assumptions. In fact, would it not be more reasonable to assume a set of simpler and logically self-evident axioms (axioms means assumptions on which a theory is based) and base the probability definition on these? This is the modern axiomatic approach to probability theory. Russian mathematician A.N. Kolmogorov developed this modern concept that combines both the objective and subjective concepts of probability.

Consider an experiment whose sample space is *S*. For each event *E* of the sample space *S*, we assume that a number *P(E)* that we refer to as probability of event *E* if it satisfies the following axioms.

Axiom-I: $0 \le P(E) \le 1$

Axiom-II: $P(S) = 1$

Or. $\sum_{i=1}^{\infty} P(E_i) = 1$ Where E_i are mutually exclusive and collectively exhaustive events.

Axiom III: For any sequence of mutually exclusive events $E_1, E_2, \ldots$ etc.

$$(\text{i.e. } E_i \cap E_j = \phi \text{ if } i \ne j)$$

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

This axiom can also be written as $P(E_1 \text{ or } E_2) = P(E_1) + P(E_2)$ where E_1 and E_2 are mutually exclusive events.

11.2 Properties of Probability

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

Where, E^c is a complement of event E. That is $(E \cap E^c) = \phi$ and $(E \cup E^c) = S$

Proof:

$P(E) + P(E^c) = 1$ by Axiom II. Hence, we get the result.

If $E \subset F$, then $P(E) \leq P(F)$

Proof:

$$\text{Now, } F = E \cup (E^c \cap F)$$

$$\text{Hence, } P(F) = P(E) + P(E^c \cap F) \text{ By Axiom III.}$$

But by Axiom I, $P(E^c \cap F) \geq 0$.

Hence, $P(F) \geq P(E)$, Proved.

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

Proof:

Note $E \cup F$ can be written as union of two disjoint events E and $E^c \cap F$.

Therefore, $\qquad\qquad\qquad P(E \cup F) = P(E \cup (E^c \cap F))$

$$= P(E) + P(E^c \cap F) \qquad \text{By Axiom III} \quad ...(1)$$

But, $\qquad\qquad\qquad\qquad\qquad P(F) = P(E \cap F) \cup (E^c \cap F)$

$$= P(E \cap F) + P(E^c \cap F) \text{ By Axiom III} \qquad ...(2)$$

Substituting (2) in (1) we get the result,

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

$P(\Phi) = 0$

Proof:

Φ is an empty set or impossible event. We know that $\Phi^C = S$. Therefore,

$$P(\Phi) = 1 - P(\Phi^C) = 1 - P(S) = 1 - 1 = 0$$

$P(E^C \cup F) \leq P(E) + P(F)$

Proof:

We have prove that $\qquad P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Now by Axiom I, $0 \leq P(E \cap F)$

Therefore, $P(E \cup F) \leq P(E) + P(F)$

$P(E^C \cup F) = P(F) - P(E \cap F)$

Proof:

From set theory, F can be written as a union of two disjoint events $E \cap F$ and $E^C \cup F$. Hence, by Axiom III, we have,

$$P(F) = P(E \cap F) + P(E^C \cap F).$$

By rearranging the terems we get the result.

11.3 Basic Principles of Probability of Multiple Events

There are two important basic principles, of probability of multiple events. These are the 'product rule' and the 'sum rule'. These are very similar to the basic principles of counting. Using these two rules and other combinatorial processes, we can derive probability of any combinations of the events.

11.3.1 The Product Rule (Sequential Rule or AND Logic)

Suppose that a procedure can be broken down into a sequence of two events. If the probability of the first event is p_1 and the probability of the second event is p_2 after the first event has taken place, then the probability of the first AND second event taking place one after other (in that order if probabilities of individual events are dependent on the order) is $p_1 \times p_2$.

In general, if a procedure can be broken down into a sequence of r events and the probability of the first event is p_1, the probability of the second event is p_2 after the first event has taken place, and so on till, the probability of the rth event is p_r, then the probability of the first AND second AND etc. rth event taking place one after other in that order (if probabilities of individual events are dependent on the order) is,

$$p_1 \times p_2 \times .... \times p_r.$$

Notes In this case the events must take place one after other. The order in which the events are performed could change the answer if the values of p_1 and p_2 etc. are affected by the order of precedence.

11.3.2 The Sum Rule (Disjunctive Rule or OR Logic)

If the probability of the first event is p_1, and the probability of the second event is p_2 and if these events cannot take place at the same time (mutually exclusive); then the probability of either first or second event taking place is $(p_1 + p_2)$.

📝 *Example:*

What is the probability that 13-card bridge hand contains:

(a)   All 13 hearts.

(b)   13 cards of the same suit.

(c)   7 spade cards and 6 club cards.

(d)   7 cards of one suit and 6 cards of another.

(e)   4 diamonds, 6 hearts, 2 spades and 1 club.

(f)   4 cards of one suit, 6 cards of second suit, 2 cards of third suit and 1 card of fourth suit.

Solution:

(a)
$$P(13\ H) = \frac{\binom{13}{13}}{\binom{52}{13}} = \frac{1}{\binom{52}{13}} = 0.1575 \times 10^{-12}$$

(b)
$$P(13) = \frac{\binom{4}{1}\binom{13}{13}}{\binom{52}{13}} = \frac{4}{\binom{52}{13}} = 6.229 \times 10^{-12}$$

(c)
$$P(7\ S, 6\ C) = \frac{\binom{13}{7}\binom{13}{6}}{\binom{52}{13}} = 4.637 \times 10^{-6}$$

(d)
$$P(7, 6) = \frac{\binom{4}{1}\binom{13}{7}\binom{3}{1}\binom{13}{6}}{\binom{52}{13}} = 0.5564 \times 10^{-4}$$

(e)
$$P(4\ D, 6\ H, 2\ S, 1\ C) = \frac{\binom{13}{4}\binom{13}{6}\binom{13}{2}\binom{13}{1}}{\binom{52}{13}} = 0.1959 \times 10^{-2}$$

(f)
$$P(4, 6, 2, 1) = \frac{\binom{4}{1}\binom{13}{4}\binom{3}{1}\binom{13}{6}\binom{2}{1}\binom{13}{2}\binom{1}{1}\binom{13}{1}}{\binom{52}{13}} = 0.0470$$

11.4 Summary

● Probability of an event can be explained or defined in four different ways.

● There are two important basic principles, of probability of multiple events. These are the 'product rule' and the 'sum rule'.

● As a measure of uncertainty, probability depends on the information available. Quite often when we know some information, probability of the event gets modified as compared to the probability of that event without such information.

● Probability of an event can be explained or defined in four different ways.

● The A-priori Probability (Mathematical or Classical Probability is also called as mathematical probability or objective probability or classical probability.

● An Experiment is a procedure that yields one of a given set of possible outcomes.

● The subjective probability of an event is the degree of confidence placed in the occurrence of an event by an individual based on all evidence available to him.

● There are two important basic principles, of probability of multiple events. These are the 'product rule' and the 'sum rule'. These are very similar to the basic principles of counting. Using these two rules and other combinatorial processes, we can derive probability of any combinations of the events.

11.5 Keywords

Classical Probability: The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible.

Event: An Event is a subset of the sample space

Experiment: An Experiment is a procedure that yields one of a given set of possible outcomes.

Sample: The Sample space of the experiment is the set of all possible outcomes.

11.6 Self Assessment

1. In a triangular series, probability of the Indian team winning match against Pakistan is 0.7 and that against Australia is 0.4. If probability of India winning both matches is 0.3, what is the probability that India will win at least one match so that it can enter final?

2. What is the probability of a hand of 13 dealt from a shuffled pack of 52 cards, contains exactly 2 kings and 1 ace?

3. An urn contains 'b' blue balls and 'r' red balls. They are removed at random and not replaced.

   (a) Show that the probability that the first red ball drawn is the $(k+1)^{th}$ ball drawn is equal to,

   $$\frac{\dbinom{r+b-k-1}{r-1}}{\dbinom{r+b}{b}}$$

   (b) Find the probability that the last ball drawn is red.

4. In a management class of 100 students, three foreign languages are offered as an additional subject viz. Japanese, French and German. There are 28 students taking Japanese, 26 taking French and 16 taking German. There are 12 students taking both Japanese and French, 4 taking Japanese and German and 6 that are taking French and German. In addition, we know that 2 students are taking all the three languages.

   (a) If a student is chosen randomly, what is the probability that he/she is not taking any of the three languages?

   (b) If a student is chosen randomly, what is the probability that he/she is taking exactly one language?

   (c) If two students are chosen randomly, what is the probability that at least one is taking language/s?

5. A question paper is divided into 3 sections: A, B and C. Section A contains 2 questions, section B contains 3 questions and section C contains 4 questions. A student is to answer five questions, of which at least one should be from each section. Find probability that the student attempts 5 questions such that he solves only one question from section A.

6. Four people North, South, East and West are dealt with 13 cards each, from a shuffled deck of 52 cards.

   (a) If South has no ace, find the probability that his partner (North) has exactly 2 aces.

   (b) If South and North together have 9 hearts, find the probability of east and west having 2 hearts each.

7. An urn contains 7 red and 3 white marbles. Three marbles are drawn from the urn one after another. Find the probability that first two are red and third is white if,

   (a) Marbles drawn are replaced after every draw.

   (b) Marbles drawn are not replaced after every draw.

8. Total of n people are present in a room. Assume that a year has 365 days and all dates are equally likely birthdays.

   (a) What is the probability that no two of them celebrate their birthday on the same day of the year?

   (b) What is the highest value of n so that the probability of any two of them celebrating their birthday on the same day is less than half?

9. In a bridge game of 52 cards, all are dealt out equally among 4 players called East, West, North and South. If North and South have a total of 8 spades among them, what is the probability that East has 3 spades of the remaining 5 spades?

10. There are 10 pairs of shoes on a shelf. If eight shoes are chosen at random,

    (a) What is the probability that no complete pairs of shoes are chosen?

    (b) What is the probability that exactly one pair of shoes is chosen?

11. Consider an experiment of tossing a fair coin until two heads or two tails occur in succession.

    (a) Describe the sample space.

    (b) What is the probability that the game will end before the sixth toss?

    (c) Given that the experiment ends with two heads, what is the probability that the experiment ends before sixth toss?

12. An urn contains *M* white balls and *N* red balls. Balls are drawn one after another.

    (a) How many draws are required so that *k* white balls are drawn before *r* red balls?

    (b) Find probability that *k* white balls are drawn before *r* red balls.

11.7 Review Questions

1. Consider an experiment of rolling a fair dice. Let the event *A* be an even number appears on the upper face. The event *B* is the number on the upper face is greater than 3. Find the probability of the number appearing on the upper face in either event *A* or *B*.

2. Three balls are randomly selected without replacement from a bag containing 20 balls numbered 1, 2, through 20. If we bet that at least one of the balls has a number greater than or equal to 17, what is the probability that we will win the bet?

3. *A* hits on an average 1 shot out of 4 on the target. *B* hits on an average 2 shots out of 5 on the target. If both fire one round each at a terrorist, what is the probability that the terrorist will be hit?

4. Two digits are selected from digits 1 through 9. If sum is even, find the probability that both numbers are odd.

5. There are 10 adjacent parking places in a parking lot. When you arrive, there are already 7 cars in the parking lot. What is the probability that you can find 2 adjacent unoccupied places for your car?

6. Two cards are drawn from a well-shuffled deck of cards.

    (a) What is the probability that both are black cards?

    (b) If both cards are black, what is the probability that one is ace and the other is either ten or face card?

7. The passengers get in an elevator on the ground floor of 20 storeyed building. What is the probability that they will all get off at different floors?

8. A hand of 13 cards is dealt from a well-shuffled deck of cards.

    (a) What is the probability that the hand contains exactly two kings and one ace?

    (b) What is the probability that the hand contains one ace, given that it contains exactly two kings?

9. Three dice are rolled. What is the probability that the highest value is twice the lowest?

10. In a gambling game Chuck-a-Luck, player bets with ₹ 1 on any number from 1 to 6. Then three dice are rolled. If the number called is shown on the face of the dice *i* times, the player wins *i* ₹. Otherwise he looses the ₹ 1 he has staked.

    (a) What should be the expected value of winning per game?

    (b) Is the game fair?

11. What is the probability that a bridge hand is void in at least one suit?

12. What is the probability that a hand of 13 cards contain,

    (a) The ace and king of at least one suit.

    (b) All 4 of at least 1 of the 13 denominations.

Answers: Self Assessment

1. Now, given that pro bability of Indian team winning match against Pakistan, $P(A)$ = 0.7, against Australia, $P(B)$ = 0.4 and against both $P(A \cap B) = 0.3$

   Therefore, probability that India will win at least one match is,

   $$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.7 + 0.4 - 0.3 = 0.8$$

2. Out of 13 cards, 2 kings must come from 4 kings in $\binom{4}{2}$ ways, 1 ace must come from 4 aces in $\binom{4}{1}$ ways, and remaining 10 cards must come from 44 non-kings and non-ace cards in $\binom{44}{10}$ ways. Thus, by product rule, the required probability of hand of 13 containing exactly 2 kings and 1 ace is,

   $$\frac{\binom{4}{2}\binom{4}{1}\binom{44}{10}}{\binom{52}{13}} = 0.09378$$

3. (a) Since the first red ball is $(k+1)^{\text{th}}$ ball, remaining $(r + b - k - 1)$ balls must contain $(r-1)$ red balls and remaining blue balls. This can be done in $\binom{r+b-k-1}{r-1}$ ways. However, total number of ways of drawing balls, of which r are red b are blue is,

   $$\frac{(r+b)!}{r!b!} = \binom{r+b}{b} \text{ ways.}$$

   Thus, the required probability is,

   $$\frac{\binom{r+b-k-1}{r-1}}{\binom{r+b}{b}}$$

   (b) By symmetry, probability of last ball is red is same as probability of first ball drawn is red. This is obviously,

   $$\frac{r}{r+b}$$

4. Let J, F and G be the symbols indicating number of students taking languages: Japanese, French and German respectively. Let S denote total number of students in the class. Thus, the given information can be written as,

   $$P(J) = \frac{28}{100} = 0.28 , \qquad P(F) = \frac{26}{100} = 0.26 \qquad P(G) = \frac{16}{100} = 0.16 ,$$

$P(J \cap F) = \dfrac{12}{100} = 0.12$, $P(J \cap G) = \dfrac{4}{100} = 0.04$, $\qquad P(F \cap G) = \dfrac{6}{100} = 0.06$ and

$P(J \cap F \cap G) = \dfrac{2}{100} = 0.02$

(a) Probability that a randomly chosen student is not taking any of the three languages is,

$P(J' \cap F' \cap G') = 1 - P(J \cup F \cup G)$

$= 1 - [P(J) + P(F) + P(G) - P(J \cap F) - P(J \cap G) - P(F \cap G) + P(J \cap F \cap G)]$

$= 1 - [0.28 + 0.26 + 0.16 - 0.12 - 0.04 - 0.06 + 0.02] = 0.5$

(b) Probability that a randomly chosen student is taking exactly one language is,

$P(J \cap F' \cap G') + P(J' \cap F \cap G') + P(J' \cap F' \cap G)$

$= [P(J) - P(J \cap F) - P(J \cap G) + P(J \cap F \cap G)]$

$\qquad + [P(F) - P(J \cap F) - P(F \cap G) + P(J \cap F \cap G)]$

$\qquad\qquad + [P(G) - P(J \cap G) - P(F \cap G) + P(J \cap F \cap G)]$

$= [P(J) + P(F) + P(G)] - 2[P(J \cap F) + P(J \cap G) + P(F \cap G)] + 3[P(J \cap F \cap G)]$

$= [0.28 + 0.26 + 0.16] - 2 [0.12 + 0.04 + 0.06] + 3 [0.02] = 0.32$

(c) Probability that at least one student out of two randomly chosen is taking language/s is, probability of both taking language or one is taking language and the other is not. This is,

$$\frac{\binom{50}{2}}{\binom{100}{2}} + \frac{\binom{50}{1} \times \binom{50}{1}}{\binom{100}{2}} = 0.2474 + 0.505 = 0.7525$$

5. Now student solves one question from each section and then solves 2 questions from remaining 6 questions. Thus, the sample space is,

$$\binom{2}{1}\binom{3}{1}\binom{4}{1}\binom{6}{2} = 360$$

However, our event is that the student attempts only one question from section A. Thus, he has to solve one question from each section and then solves 2 questions from remaining 5 questions from section B or C. Thus, the event space is,

$$\binom{2}{1}\binom{3}{1}\binom{4}{1}\binom{5}{2} = 240$$

Hence, the required probability of the event is,

$$P = \frac{240}{360} = \frac{2}{3}$$

6. (a) Since South hasn't got ace in his 13 cards, remaining 39 cards must have 4 aces. Out of these 39 cards, North gets 13 cards. Of these, the number of cases when he gets exactly 2 aces is, $\binom{4}{2}\binom{35}{11}$. Hence, the probability that North has exactly 2 aces is,

$$P = \frac{\binom{4}{2}\binom{35}{11}}{\binom{39}{13}} = 0.308$$

(b) South and North together have 26 cards, of which 9 are hearts. Thus, any of their opponents, say East can get his card from the remaining 26 cards, of which 4 are hearts. Number of ways East can be dealt cards is $\binom{26}{13}$. Of these number of cases with 2 hearts and remaining other cards is, $\binom{4}{2}\binom{22}{11}$. Hence, the probability of East having exactly 2 hearts is,

$$\frac{\binom{4}{2}\binom{22}{11}}{\binom{26}{13}} = 0.407$$

7. (a) $P = \dfrac{7}{10} \times \dfrac{7}{10} \times \dfrac{3}{10} = 0.147$

   (b) $P = \dfrac{7}{10} \times \dfrac{6}{9} \times \dfrac{3}{8} = 0.175$

8. (a) $P = \dfrac{365 \times 364 \times .... \times (365 - n + 1)}{365^n} = \dfrac{_{365}P_{(365-n+1)}}{365^n} = \dfrac{365!}{(365 - n)! \times 365^n}$

   (b) In the above probability formula by substituting various values of n (trial and error) we find that the probability of any two of them celebrating their birthday on the same day is less than half for the values $n \leq 22$.

9. Now North and South together have 26 cards with 8 spades out of these. Thus, for East and West there are 26 cards, out these 5 are spades. East is given 13 cards out of the remaining 26 cards, which can be in $\binom{26}{13}$ ways. The event that we are interested in is East getting 3 spades out of 5 in $\binom{5}{3}$ ways and remaining 10 non-spade cards out of 21 non-spades cards in $\binom{21}{10}$ ways. Hence, required probability is, $P = \dfrac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} = 0.339$

10. (a) Totally there are 10 pairs, i.e. 20 shoes, of which 8 are chosen. This can be done in

$\binom{20}{8}$ ways. Now our event is no complete pair is chosen. This can be achieved by

first selecting 8 pairs out of 10 in $\binom{10}{8}$ ways and then choosing one shoe each from

these in 2^8 ways. Thus, probability of the required event is,

$$P = \frac{\binom{10}{8} \times 2^8}{\binom{20}{8}} = 0.0914$$

(b) In this case, to achieve the event of exactly one pair is chosen, we first choose 1 pair

out of 10 in $\binom{10}{1}$ ways. Then we select 6 unpaired shoes out of remaining 9 pairs as,

$\binom{9}{6} \times 2^6$. Thus, probability of the required event is,

$$P = \frac{\binom{10}{1} \times \binom{9}{6} \times 2^6}{\binom{20}{8}} 0.4268$$

11. (a) Sample space is series of H and T terminated by HH or TT. Thus, it can be listed as,

HH, TT, HTT, THH, THTT, HTHH, THTHH, HTHTT, etc.

(b) Now each toss has probability of T as $\frac{1}{2}$ and probability of H also as $\frac{1}{2}$.

Therefore, using sum and product rule, probability that the game will end before

the sixth toss is, $\frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} + \frac{1}{32} + \frac{1}{32} = \frac{15}{16}$

(c) By symmetry of the problem, given that the experiment ends with two heads,

probability that the experiment ends before sixth toss is, $\frac{15}{32}$

Note: This problem can be solved by using geometric probability distribution that we will study later.

12. (a) Number of draws required, so that *k* white balls are drawn before *r* red balls is, $k + i$ –1 with $1 \leq i \leq r$ because minimum draws required are *k* and maximum draws as $k + r - 1$

    (b) Since *k* white balls are selected from *M* white balls and *i* red balls are selected from *N* red balls, the required probability is,

$$P(White = k, red < i) = \frac{\binom{M}{k-1}\binom{N}{i-1}}{\binom{M+N}{k-1+i-1}} \times \frac{M-k+1}{M+N-k-i+2}$$

Where $i = 1, 2, 3, r \leq N$

11.8 Further Readings

Books   Béla Bollobás, *Modern graph theory*, Springer

Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph theory, Combinatorics, and algorithms*, Birkhäuser

Online links   http://www.stat.psu.edu/~jiali/course/stat416/notes/chap1.pdf

http://www.mathgoodies.com/lessons/vol6/conditional.html

http://www.mathgoodies.com/lessons/vol6/independent_events.html

Unit 12: Conditional Probability

CONTENTS

Objectives

Introduction

Objectives

After studying this unit, you will be able to:

- Calculate conditional probability
- Discuss independent events
- Understand independent repeated trials

Introduction

In last unit you have studied about probability theory. The probability theory tries to measure the possibility of an outcome in numerical terms. As you know that many problems are concerned with values associated with outcomes of random experiments. For example, we select ten items randomly from the production lot with known proportion of defectives and put them in a packet. Now we want to know the probability that the packet contains more than one defective item. The study of such problems requires a concept of a random variable. Random Variable is a function that associates numerical values to the outcomes of experiments. This unit provides you brief description of conditional probability and random variable.

12.1 Conditional Probability

As a measure of uncertainty, probability depends on the information available. Quite often when we know some information, probability of the event gets modified as compared to the probability of that event without such information. If we know occurrence of say event *F*, probability of event *E* happening may be different as compared to original probability of *E* when we had no knowledge of the event *F* happening. If we let *E* and *F* denote two events, then the conditional probability that *E* occurs given that *F* has occurred is denoted by $P(E|F)$. If the event *F* occurs, then in order for *E* to occur it is necessary that the actual occurrence be a point in both *E* and in *F*. That is, it must be in $(E \cap F)$ or (*EF*). As we know that *F* has occurred, it follows that *F* becomes our new or reduced sample space. Hence, the probability that event $E \cap F$ occurs is equal to the probability of relative to the probability of $E \cap F$ relative to the probability of *F*. That is,

$$P(E|F) \quad = \quad \frac{P(E \cap F)}{P(F)}$$

Or,

$$P(E \cap F) \quad = \quad P(E|F) \ P(F) = P(F|E) P(E)$$

Or,

$$P(EF) \quad = \quad P(E|F) \ P(F) = P(F|E) P(E)$$

These are called as conditional probability formulae.

To explain this idea, let us consider an example: you go to your friend who tells you that he has two children. So you think about probability of at least one of them being a boy *P(B)*. Now the sample space is all possible combinations of two children. Considering elder child and younger child, all possible outcomes are *BB*, *BG*, *GB* and *GG*, where *B* represents boy and *G* represents girl. Also first letter represents the elder child and second letter the younger one. All these have equal probability of $\frac{1}{4}$ each (by using product rule and assuming the probability at birth of Boy or Girl is $\frac{1}{2}$). Of these, four equally likely occurrences, in three cases we have boy (at least one).

Thus, $P(B) = \frac{3}{4}$

Now as you enter the house one girl greets you. Your friend tells you that she is his daughter, now you quickly recalculate *P(B)* as follows. Of the original sample space event *BB* is now impossible. Hence, the sample space is reduced to three. Also the favorable events are now two viz. *BG* and *GB*. Hence, probability of one boy is,

$$P(B) = \frac{2}{3} \qquad\qquad \text{which is now modified.}$$

Further, friend also informs you that the girl you have seen is elder child. This additional information further changes the probability. Now sample space is reduced to two viz. GG and GB. Out of these, only one case includes boy. Hence,

$$P(B) = \frac{1}{2} \qquad\qquad \text{which is again modified.}$$

Suppose at this stage, a boy enters the room and your friend tells you that the boy is his son, obviously you know with certainty that your friend has a boy. Thus,

$$P(B) = 1 \qquad \text{which is again modified.}$$

So we have seen that with every additional piece of information, probability got modified. It could increase or decrease or remain unchanged depending upon the kind of information you receive and its impact on sample space/event. We will see later that in certain cases additional information does not change probability. Such events are called independent events.

Task   Analyze some examples of conditional probability in your daily life.

12.2 Multiplication Rule of Conditional Probability

Generalized expression for the probability of intersection of an arbitrary number of events can be obtained by repeated application of conditional probability formula as follows. Suppose E_1, E_2, E_3, ...,E_n are events in the sample space S, probability of intersection of events E_1, E_2, E_3,...,E_n is given by,

$$P(E_1 E_2 E_3 ... E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2)...P(E_n|E_1 E_2 E_3 ... E_{n-1})$$

Proof:

We apply conditional probability formula to all terms on R.H.S.

$$\text{R.H.S.} = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2)...P(E_n|E_1 E_2 E_3 ... E_{n-1})$$

$$= P(E_1)\frac{P(E_1 E_2)}{P(E_1)}\frac{P(E_1 E_2 E_3)}{P(E_1 E_2)}......\frac{P(E_1 E_2 E_3 ... E_n)}{P(E_1 E_2 E_3 ... E_{n-1})}$$

$$= P(E_1 E_2 E_3 ... E_n) = \text{L.H.S.} \qquad \text{Thus, the result is proved.}$$

Example: A deck of 52 playing cards is distributed randomly to 4 players. Find the probability that each player gets exactly one ace.

Solution:

Method I

Let the events E_1, E_2, E_3 and E_4 be defined as follows.

E_1 = {the ace of spades is with any one player}

E_2 = {the ace of hearts is with any player other than the one who has ace of spades}

E_3 = {the ace of diamonds is with any player other than those who have ace of spades or ace of hearts}

E_4 = {the ace of clubs is with the remaining player}

The required probability is a joint probability of the events E_1, E_2, E_3 and E_4. That is,

$$\text{Probability} = P(E_1 E_2 E_3 E_4)$$

$$= P(E_1)P(E_2|E_1)P(E_3|E_1 E_2)P(E_4|E_1 E_2 E_3) \text{ Using multiplication rule.}$$

Now, it is certain that one of the players will get ace of spades. Hence,

$$P(E_1) \quad = \quad 1$$

Since ace of spades is already identified with one of the players who also gets 12 other non-spade-ace cards. Other three players can get cards out of remaining 39. One of the remaining players gets ace of hearts, if this ace is in these 39 remaining cards out of 51 non-spade-ace cards. Probability of this event is,

$$P(E_2|E_1) \quad = \quad \frac{39}{51}$$

Similarly, after ace of spades and ace of hearts is identified with two players who also have got another 24 non-spade-ace and non-heart-ace cards (12 each), probability of one of the remaining players getting ace of diamond is,

$$P(E_3|E_1E_2) \quad = \quad \frac{26}{50}$$

Similarly, probability that the last player will get ace of clubs is,

$$P(E_4|E_1E_2E_3) \quad = \quad \frac{13}{49}$$

Thus, $\quad P(E_1E_2E_3E_4) \quad = \quad \dfrac{1 \times 39 \times 26 \times 13}{1 \times 51 \times 50 \times 49} \approx 0.105$

Method II

Probability of one of the players getting exactly one ace $= \dfrac{\dbinom{4}{1}\dbinom{48}{12}}{\dbinom{52}{13}}$

Having done that, probability of another player getting exactly one ace $= \dfrac{\dbinom{3}{1}\dbinom{36}{12}}{\dbinom{39}{13}}$

Similarly, probability of third player also getting exactly one ace $= \dfrac{\dbinom{2}{1}\dbinom{24}{12}}{\dbinom{26}{13}}$

Once three players get exactly one ace, the remaining player will automatically get exactly one ace with probability 1. Now, by product rule, the required probability is,

$$\frac{\dbinom{4}{1}\dbinom{48}{12}}{\dbinom{52}{13}} \times \frac{\dbinom{3}{1}\dbinom{36}{12}}{\dbinom{39}{13}} \times \frac{\dbinom{2}{1}\dbinom{24}{12}}{\dbinom{26}{13}}$$

$$= \frac{4 \times 48! \times 13! \times 39! \times 3 \times 36! \times 26! \times 13! \times 2 \times 24! \times 13! \times 13!}{12! \times 36! \times 52! \times 12! \times 24! \times 39! \times 12! \times 12! \times 26!}$$

$$= \frac{3 \times 2 \times 13 \times 13 \times 13}{51 \times 50 \times 49}$$

$$\approx 0.105$$

12.3 Law of Total Probability

Consider two events, *E* and *F*. Whatever may be the relation between the two events, we can always say that the probability of *E* is equal to the probability of intersection of *E* and *F*, plus the probability of the intersection of *E* and complement of *F*. That is,

$$P(E) = P(E\ F) + P(E\ F^C)$$

12.4 Random Variable

Let S be the sample space corresponding to the outcomes of a random experiment. A function $X : S \rightarrow R$ (where R is a set of real numbers) is called a random variable.

Random variable is a real valued mapping. Thus, the function has to be one-one or many-one correspondence. Thus, a random variable assigns a real number to each possible outcome of an experiment. Note that a random variable is a function, it is neither a variable nor random. A random variable is a function from the sample space of a random experiment (domain) to the set of real numbers (co-domain). Note that *X* is not a random variable if the mapping is one-many or any of the outcomes of the experiment is not mapped at all by the function.

Probability distribution of random variable must satisfy all Axioms of probability.

Example: Suppose two coins are tossed simultaneously. What are the values a random variable X would take, if it were defined as number of heads?

Solution:

Now sample space (possible outcomes) of the experiment is,

S = {TT, TH, HT, HH}

Since the random variable X is number heads, it takes three distinct values {0, 1, 2}

Example: Let X be a random variable defined as difference of the numbers that appear when a pair of dice is rolled. What are the distinct values of the random variable?

Solution:

When a pair of dice is rolled, there are 36 possible outcomes. This sample set is,

S = {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

Now the random variable X is the difference of the numbers. Its values are,

X = {0, 1, 2, 3, 4, 5}

📝

Example

X(1,1) = X(2,2) = X(3,3) = X(4,4) = X(5,5) = X(6,6) = 0

X(1,2) =X(2,1) =X(2,3) =X(3,2) =X(3,4) =X(4,3) =X(4,5) =X(5,4) =X(5,6) =X(6,5) = 1

12.5 Independent Events

As we have seen earlier, in general, knowledge of some information may change probability of an event. For example, if *E* and *F* are two events, knowledge that event *F* has occurred, generally changes the probability of the event *E*. That is,

$$P(E|F) \neq P(E)$$

However, if in special cases, knowledge that *F* has occurred does not change probability of *E*, we say that *E* is independent of *F*. In such case,

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

12.5.1 Condition of Independence of Two Events

Two events *E* and *F* are independent, if and only if,

$$P(E \cap F) = P(E) \times P(F)$$

Proof

Part I

If event *E* is independent of *F*,

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

But, we know that by conditional probability formula, $P(E|F) = \dfrac{P(E \cap F)}{P(F)}$

Substituting and simplifying we get,

$$P(E \cap F) = P(E) \times P(F)$$

Part II

If $P(E \cap F) = P(E) \times P(F)$ using conditional probability formula we get,

$$P(E|F) \times P(F) = P(E) \times P(F)$$

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

Hence, the result is proved.

Note: Events that are not independent are said to be dependent.

📋

Notes   Events that are not independent are said to be dependent.

12.5.2 Properties of Independent Events

1. If *E* and *F* are independent, then so are *E* and *FC*.

    Proof:

    Let *E* and *F* are independent. Now, $E = EF \cup EF^C$. Also *EF* and *EFC* are obviously mutually exclusive events. Therefore, by Axiom III,

    $$P(E) = P(EF) + P(EF^C)$$
    $$= [P(E) \times P(F)] + P(EF^C) \qquad \text{Because } E \text{ and } F \text{ are independent.}$$
    $$P(EF^C) = P(E) - [P(E) \times P(F)] = P(E)[1 - P(F)]$$
    $$= P(E) \times P(F^C)$$

    Therefore, *E* and F^C are independent.

2. If *E*, *F*, and *G* are independent, if and only if,

    $$P(EFG) = P(E)P(F)P(G), \qquad P(EF) = P(E)P(F)$$

    $$P(EG) = P(E)P(G) \qquad P(FG) = P(F)P(G)$$

3. If *E*, *F*, and *G* are independent, then *E* will be in dependent of any event formed from *F* and *G*.

12.6 Finite Probability Spaces

Finite probability space consists of two things (a) Finite space and (b) Probability

For e.g. Toss of a coin is a 'finite probability space' since we have 'finite space' -> (Head,Tail) . The Head and Tail are referred to as events.

Next there is probability of each individual event and the sum of probabilities for events is 1.

Let us take an example. Let the probability of a head be 2/3 and probability of a tail be 1/3.

Probability space of three tosses is {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} and each individual event is referred to as w.

The probability of HHH above is = 2/3* 2/3 * 2/3 = 8/27. On the same lines the probability of each individual event in the above three toss is {8/27, 4/27, 4/27, 2/27, 4/27, 2/27, 2/27, 1/27}. This individual probabilities is also called Probability measure and is shown as P.

Random variable is an experiment designed on Finite probability space. An example would be the number of heads in three tosses of coin. The value of this random variable can be 0,1,,2,3.

Probability of 0 heads = P(TTT) = 1/27

Probability of 1 heads = P(HTT) + P (THT) + P(TTH) = 6 / 27

Probability of 2 heads = P(HHT) + P (HTH) + P(THH) = 12 / 27 and

Probability of 3 heads = P(HHH) = 8/27

This is called distribution of a random variable. It is the probability of the random variable taking the values 0,1,2,3.

⚠

Caution There seems to be some underlying assumption in the above example that probability of each event in 3 toss is same for e.g. P(HHHH) = P(HTT)=P(HTH) and so on.

12.7 Independent Repeated Trials

Independent repeated trials of an experiment with two outcomes only are called Bernoulli trials. Call one of the outcomes success and the other outcome failure. Let p be the probability of success in a Bernoulli trial. Then the probability of failure q is given by q = 1 – p.

A binomial experiment consisting of a fixed number n of trials, each with a probability of success p, is denoted by B(n,p). The probability of k success in the experiment B(n,p) is given by:

$$P(k) = \binom{n}{k} p^k q^{n-k}$$

The function P(k) for $k = 0, 1, ..., n$ for B(n,p) is called a binomial distribution.

Bernoulli trials may also lead to negative binomial, geometric, and other distributions as well.

Case Study

A town has 3 doctors *A*, *B* and *C* operating independently. The probability that the doctors *A*, *B* and *C* would be available is 0.9, 0.6 and 0.7 respectively. What is the probability that at least one doctor is available when needed?

Solution:

Given is $P(A) = 0.9$, $P(B) = 0.6$ and $P(C) = 0.7$

Method I

Probability that at least one doctor is available is $P(A \cup B \cup C)$. Now using inclusion-exclusion principle,

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

Since *A*, *B* and *C* are independent,

$P(A \cap B) = P(A)P(B) = 0.54$ $\qquad$ $P(A \cap C) = P(A)P(C) = 0.63$

$P(B \cap C) = P(B)P(C) = 0.42$ $\qquad$ $P(A \cap B \cap C) = P(A)P(B)P(C) = 0.378$

Therefore, probability that at least one doctor is available is,

$P(A \cup B \cup C) = 0.9 + 0.6 + 0.7 - 0.54 - 0.63 - 0.42 + 0.378 = 0.988$

Method II

Here we will use product and sum rule.

Probability that at least one doctor is available is sum of one, two or three doctors are available. Thus, the required probability is,

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

Since all these events are independent (note that if A and B are independent, A and B', A' and B as well as A' and B' are all independent) we use product rule. Thus,

$P(A \cup B \cup C)$
$= 0.9 \times 0.4 \times 0.3 + 0.1 \times 0.6 \times 0.3 + 0.1 \times 0.4 \times 0.7$
$+ 0.9 \times 0.6 \times 0.3 + 0.9 \times 0.4 \times 0.7 + 0.1 \times 0.6 \times 0.7 + 0.9 \times 0.6 \times 0.7$
$= 0.108 + 0.018 + 0.028 + 0.162 + 0.252 + 0.042 + 0.378$
$= 0.988$

12.8 Summary

- Generalized expression for the probability of intersection of an arbitrary number of events can be obtained by repeated application of conditional probability formula.

- Finite probability space consists of two things (a) Finite space and (b) Probability.

- Generalized expression for the probability of intersection of an arbitrary number of events can be obtained by repeated application of conditional probability formula.

- Consider two events, E and F. Whatever may be the relation between the two events, we can always say that the probability of E is equal to the probability of intersection of E and F, plus the probability of the intersection of E and complement of F. That is, $P(E) = P(E F) + P(E F^C)$

- As we have seen earlier, in general, knowledge of some information may change probability of an event. For example, if E and F are two events, knowledge that event F has occurred, generally changes the probability of the event E. That is, $P(E|F) \neq P(E)$

- Independent repeated trials of an experiment with two outcomes only are called Bernoulli trials.

12.9 Keywords

Conditional Probability: The conditional probability of A given B is denoted by P (A|B) and defined by the formula

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

provided P (B) > 0. (If P (B) = 0, the conditional probability is not defined.) (Recall that AB is a shorthand notation for the intersection A $\cap$ B.)

Rules for complement conditional probabilities: Complement rule for conditional probabilities: P (A^c|B) = 1 – P (A|B)

Multiplication rule (or chain rule) for conditional probabilities: P(AB) = P(A|B)P (B). More generally, for any events $A_1,...,A_n$,

P ($A_1 A_2 ...A_n$)= P(A_1)P (A_2|A_1)P (A_3|$A_1 A_2$) ...P(A_n|$A_1 A_2 ...A_n$ – 1).

12.10 Self Assessment

1. The probability that a new product will be successful if a competitor does not launch a similar product is 0.67. The probability that a new product will be successful in the presence of a competitor's new product is 0.42. The probability that the competitor will launch a new product is 0.35. What is the probability that the product will be success?

2. A bin contains 3 different types of lamps. The probability that a type 1 lamp will give over 100 hours of use is 0.7, with the corresponding probabilities for type 2 and 3 lamps being 0.4 and 0.3 respectively. Suppose that 20 percent of the lamps in the bin are of type 1, 30 percent are of type 2 and 50 percent are of type 3.

    (a) What is the probability that a randomly selected lamp will last more than 100 hours?

    (b) Given that a selected lamp lasted more than 100 hours, what are the conditional probabilities that it is of type 1, type 2, and type 3?

3. A certain firm has plants A, B and C producing 35%, 15% and 50% respectively of the total output. The probabilities of non-defective product from these plants are 0.75, 0.95 and 0.85 respectively. The products from these plants are mixed together and dispatched randomly to the customer. A customer receives a defective product. What is the probability that it came from plant C?

4. In a college, 4 percent boys have work experience and 1 percent girls have work experience. Out of total students 60 percent are girls.

    (a) If we select a student randomly and he has work experience, what is the probability that she is a girl?

    (b) If we select a student randomly, what is the probability that he/she has work experience?

5. A computer system consists of 6 subsystems. Each subsystem might fail independently with a probability of 0.2. The failure of any subsystem will lead to a failure of a whole computer system. Given that the computer system has failed, what is the probability that system 1 and only system 1 has failed?

6. Two urns contain respectively 3 white and 4 black balls; and 2 white and 5 black balls. What is the probability that,

    (a) Ball drawn randomly from first urn is black?

    (b) One urn is selected at random, and a ball drawn from it is black?

    (c) One urn is selected at random, and a ball drawn from it is black; if probability of first urn selection is twice that of second?

    (d) Ball is drawn from first urn and put in the second and then one ball drawn from the second urn is black?

7. A product is produced on three different machines M_1, M_2 and M_3 with proportion of production from these machines as 50%, 30% and 20% respectively. The past experience shows percentage defectives from these machines as 3%, 4% and 5% respectively. At the end of the day's production, one unit of production is selected at random and it is found to be defective. What is the chance that it is manufactured by machine M_2.

8. A bag contains 4 tickets numbered 112, 121, 211 and 222. One ticket is drawn randomly. Let A_i be the event that i^{th} digit of the number on the ticket is 1 with i = 1, 2, 3. Comment on pair-wise and mutual independence of A_1, A_2 and A_3.

9. A highway has three recovery vans namely I, II and III. The probability of their availability at any time is 0.9, 0.7 and 0.8 and is independent of each other. What is the probability that at least one recovery van will be available at any time to attend the breakdown?

10. A husband and wife both appear for two vacancies in the same post. The probability of selection for husband and wife is 0.4 and 0.7 respectively. What is the probability that:

    (a) Both of them will be selected.

    (b) Only one of them will be selected.

    (c) None of them will be selected.

11. In a certain examination, results show that 20% students failed in P & C, 10% failed in Data Structure while 5% failed in both P & C and Data Structure. Are the two events 'failing in P & C' and 'failing in Data Structure' independent?

12. In a certain examination, results show that 20% students failed in P & C, 10% failed in Data Structure while 2% failed in both P & C and Data Structure. Are the two events 'failing in P & C' and 'failing in Data Structure' independent?

12.11 Review Questions

1. A pair of dice is rolled. Find the probabilities of the following events,

    (a) The sum of two numbers is even.

    (b) The sum of the two numbers is at least 8.

    (c) The product of two numbers is less than or equal to 9.

2. A bag contains 4 white and 2 black balls. Another bag contains 3 white and 5 red balls. One ball is drawn from each bag. What is the probability that they are of different colours?

3. A factory has three delivery vans A, B and C. The probability of their availability is 0.6, 0.75 and 0.8 respectively. Probability of availability of A and B both is 0.5, A and C both is 0.4 and B and C both is 0.7. The probability of all the three vans being available is 0.25. What is the probability that on a given day, you will get a van to deliver your product?

4. An integer between 1 and 100 (both inclusive) is selected at random. Find the probability of selecting a perfect square if,

    (a) All integers are equally likely.

    (b) An integer between 1 and 50 is twice as likely to occur as an integer between 51 and 100.

5. A ball is drawn from an urn containing 3 white and 3 black balls. After the ball is drawn, it is replaced and another ball is drawn. This goes on indefinitely. What is the probability that of first four balls drawn, exactly two are white?

6. A fair coin is tossed four times. Find the probability that we will get two heads and two tails.

7. An investment consultant predicts that the odds against price of a certain stock to go up during next week are 2:1 and odds in favour of the price to remain same during next week are 1:3. What is the probability that the stock will go down during the next week?

8. An office has three Xerox machines X_1, X_2 and X_3. The probability that on a given day machines X_1, X_2 and X_3 would work is 0.60, 0.75 and 0.80 respectively; both X_1 and X_2 work

is 0.50; both X_1 and X_3 work is 0.40; both X_2 and X_3 work is 0.70. The probability that all of them work is 0.25. Find the probability that on a given day at least one of the three machines works.

9. A bag contains 4 white and 2 black balls. Another bag contains 3 white and 3 black balls. One ball is drawn from each bag at random. What is the probability that they are of different colours?

10. Three books of science and two books of mathematics are arranged on a shelf. Find the probability that

    (a) Two mathematics books are side by side.

    (b) Two mathematics books are not side by side.

11. Two cards are drawn from a shuffled deck of cards. Determine the probability that

    (a) Both are aces.

    (b) Both are spades.

    (c) Both belong to same suit.

12. Ten students are seated at random in a row. Find the probability that two particular students are not seated side by side.

13. The first 12 letters of English alphabet are written at random. What is the probability that

    (a) There are 4 letters between A and B.

    (b) A and B are written side by side.

14. A factory has 65% male workers. 70% of the total workers are married. 47% of the male workers are married. Find the probability that a worker chosen randomly is,

    (a) A married female.

    (b) A male or married or both.

15. A dice is rolled. What is the probability that odd number shows up on the upper face if,

    (a) Dice is fair.

    (b) Dice is crooked with probability of getting number 6 is $\dfrac{1}{3}$. All other numbers are equally likely.

16. What is the maximum number of randomly selected people assembled at a party such that the chance of any two of them having birthday on the same date is less than 50-50?

17. In one of the cities of India out of 100000 people, 51500 are male and 48500 are female. Among the males, 9000 use cosmetics. Among the women, 30200 use cosmetics. If a person is selected at random, what is the probability that,

    (a) He or she uses cosmetics.

    (b) A male or a person using cosmetics is chosen.

    (c) A male not using cosmetics or a female using cosmetics, is chosen.

    (d) A male not using cosmetics is chosen.

18. An article manufactured by a company consists of two parts A and B. In the manufacturing process of part A, 9 out of 100 are likely to be defective. Similarly 5 out of 100 are likely to be defective in the process of part B. Calculate the probability that the assembled parts will not be defective.

19. A certain item is manufactured by three factories F_1, F_2 and F_3. It is known that factory F_1 turns out twice as many items as F_2 and that F_2 and F_3 turn out same number of items in one day. It is also known that defective percent of items from these factories is 2%, 2% and 4% respectively. One item is chosen at random from the mixed lot of items from these factories. The chosen item is found to be defective. Find the probability that this item is produced at the factory F_1.

20. There are 4 boys and 2 girls in room no.1 and 5 boys and 3 girls in room no. 2. A girl from one of the room laughed loudly. What is the probability that the girl who laughed loudly was from room no. 2?

21. In a community, 5% of the population suffers from cancer. The probability that a doctor is able to correctly diagnose a person with cancer, as suffering from cancer is 0.9. The doctor wrongly diagnoses a person without cancer as having cancer with probability 0.1. What is the probability that a randomly selected person diagnosed as having cancer is really suffering from cancer?

22. The occurrence of two independent events is known to be each greater than $\dfrac{1}{2}$. It is given that the probability that the first event will happen simultaneously with the second not occurring is $\dfrac{3}{25}$. Also the probability of the second event occurring simultaneously with the first not occurring is $\dfrac{8}{25}$. Find the probabilities of the respective events.

23. A problem in probability is given to three students, whose chances of solving are $\dfrac{1}{2}, \dfrac{1}{3}, \dfrac{1}{4}$ respectively. If all the three students solve the problem independently, what is the probability that the problem will be solved?

24. Explain the concept of independence of two events.

25. Does independence of two events imply that the events are mutually exclusive? Justify your answer.

26. Does mutually exclusiveness of two events imply independence? Justify with example.

27. When are two events independent as well as mutually exclusive?

28. Bag I contains 6 blue and 4 red balls. Bag II contains 2 blue and 6 red balls. Bag III contains 1 blue and 8 red balls.

    (a) A bag is chosen at random; a ball is drawn randomly from this bag. It turned out to be blue. Find the probability that bag I was chosen.

    (b) A bag is chosen at random; two balls are drawn together. Both balls were blue. Find the probability that bag II was chosen.

29. An explosion in a bomb dump of army unit can occur due to (i) Short circuit (ii) Defect in fusing equipment (iii) Human Error (iv) Sabotage. The probabilities of these four causes are known to be 0.3, 0.2, 0.4 and 0.1 respectively. The engineer feels that an explosion can occur with probability (i) 0.3, if there is a short circuit, (ii) 0.2, if there is a defective fusing equipment (iii) 0.25, if the worker makes an error (iv) 0.8, if there is a sabotage. Given that an explosion had occurred, determine the most likely cause of it.

30. Ten men went to a temple. They kept their sandals at shoe stand. When they returned, the attendant returned the sandals at random (possibly he mixed up the tokens). What is the probability that no one gets his own sandals?

31. A can hit a target 3 times in 5 shots, B can hit 2 times in 4 shots and C can hit 3 times in 4 shots. Effectiveness of the weapon is that it takes two hits to kill and one hit to injure a person. They fire one shot each on an escaping terrorist.

     (a)   What is the probability that the terrorist would die?

     (b)   What is the probability that terrorist will escape unhurt?

32. A committee of 4 people is formed randomly of managers from various departments; 3 from production, 2 from sales, 4 from logistics and 1 from finance. Find the probability that the committee has,

     (a)   One manager from each department.

     (b)   At least one manager from logistic department.

     (c)   The manager from finance.

33. A and B are two independent witnesses in a case. The probability of A speaking truth is x and the probability of B speaking truth is y. A and B agree on a particular statement. Show that if the statement is correct, then its probability is, $\dfrac{xy}{1-x-y+2xy}$.

34. Urn I contains 5 red and 5 black balls. Urn II contains 4 red and 8 black balls. Urn III contains 3 red and 6 black balls. A ball is drawn randomly from urn I and put in urn II without seeing the colour. Again a ball is drawn randomly from urn II and put in urn III without seeing the colour. What is the probability that the ball now drawn from urn III is red?

35. Suppose a company hires both MCA and non-MCA for same kind of tasks. After a period of employment some of each category is promoted. Table below gives the past data.

Status	Academic Qualification		Total
	MCA	Non-MCA	
Promoted	0.42	0.18	0.60
Not Promoted	0.28	0.12	0.40
Total	0.70	0.30	1

     (a)   If a person drawn is promoted, what is the probability that he is MCA?

     (b)   If a person drawn is MCA, what is the probability that he is promoted?

     (c)   Whether qualification of MCA and promotion is independent?

36. What is the conditional probability that a randomly generated bit string of length 4 contains at least two consecutive 0_s, given that first bit is 1?

37. Let E and F be the events that a student committee of n students has representatives of both sexes and at the most one girl respectively. Are E and F independent if,

     (i) $n = 2$?   (ii) $n = 4$?   (iii) $n = 5$?

38. A group of 6 people play the game "Odd person out". Each person flips a fair coin. If there is a person whose outcome is not the same as that of any other member of the group, this person is out. What is the probability that one person will be out in the first round itself?

Answers: Self Assessment

1. Let S denote that the product is successful, L denote competitor will launch a product and L^C denotes competitor will not launch the product. Now, from given data,

    $P(S|L^C) = 0.67$, $P(S|L) = 0.42$, $P(L) = 0.35$

    Hence, $P(L^C) = 1 - P(L) = 1 - 0.35 = 0.65$

    Now, using conditional probability formula, probability that the product will be success $P(S)$ is,

    $P(S) = P(S|L)P(L) + P(S|L^C)P(L^C)$

    $= 0.42 \times 0.35 + 0.67 \times 0.65 = 0.5825$

2. Let type 1, type 2 and type 3 lamps be denoted by $T1$, $T2$ and $T3$ respectively. Also, we denote S if a lamp lasts more than 100 hours and S^C if it does not. Now as per given data,

    $P(S|T1) = 0.7$, $P(S|T2) = 0.4$, $P(S|T3) = 0.3$

    $P(T1) = 0.2$, $P(T2) = 0.3$, $P(T3) = 0.5$

    (a) Now, using conditional probability formula,

    $P(S) = P(S|T1)P(T1) + P(S|T2)P(T2) + P(S|T3)P(T3)$

    $= 0.7 \text{ X } 0.2 + 0.4 \text{ X } 0.3 + 0.3 \text{ X } 0.5 = 0.41$

    (b) Now, using Bayes' formula,

    $$P(T1|S) = \frac{P(S|T1)P(T1)}{P(S)} = \frac{0.7 \times 0.2}{0.41} = 0.341$$

    $$P(T2|S) = \frac{P(S|T2)P(T2)}{P(S)} = \frac{0.4 \times 0.3}{0.41} = 0.293$$

    $$P(T3|S) = \frac{P(S|T3)P(T3)}{P(S)} = \frac{0.3 \times 0.5}{0.41} = 0.366$$

3. Let us use symbols D for defective and ND for non-defective.

    Given data can be written as,

    $P(ND|A) = 0.75 \Rightarrow P(D|A) = 0.25$

    $P(ND|B) = 0.95 \Rightarrow P(D|A) = 0.05$

    $P(ND|C) = 0.85 \Rightarrow P(D|A) = 0.15$

    Now we need to find probability of the item has come from C when we know that it is defective, i.e. $P(C|D)$, Using Bayes' formula,

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

$$= \frac{0.15 \times 0.5}{0.25 \times 0.35 + 0.05 \times 0.15 + 0.15 \times 0.5}$$

$$= \frac{0.075}{0.17} = 0.44$$

4. Let us use the symbols as G – Girl, B – Boy, W – Work Experience and NW – No work experience. Now given is,

    $$P(W|B) = 0.04 \Rightarrow P(NW|B) = 0.96$$

    $$P(W|G) = 0.01 \Rightarrow P(NW|G) = 0.99$$

    $$P(G) = 0.6 \Rightarrow P(B) = 0.4$$

    (a) Using Bayes' formula,

    $$P(G|W) = \frac{P(W|G)P(G)}{P(W|G)P(G) + P(W|B)P(B)} = \frac{0.6 \times 0.01}{0.6 \times 0.01 + 0.4 \times 0.04} = 0.2727$$

    (b) $P(W) = P(W|G)P(G) + P(W|B)P(B) = 0.6 \times 0.01 + 0.4 \times 0.04 = 0.022$

5. Let us denote subsystems as A_i and computer system as B. Given,

    $$P(A_i) = 0.2 \qquad i = 1, 2, 3, 4, 5, 6$$

    $$P(B|A_i) = 1 \qquad i = 1, 2, 3, 4, 5, 6$$

    Now, probability that system 1 and only system 1 has failed, given that the computer system has failed is,

    $$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{6} P(A_i)P(B|A_i)} = \frac{1}{6}$$

6. Let event A_1 : First urn is selected; event A_2 : Second urn is selected; event B: Ball drawn is black.

    (a) $P(B|A_1) = \dfrac{4}{7}$

    (b) $P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) = \dfrac{4}{7} \times \dfrac{1}{2} + \dfrac{5}{7} \times \dfrac{1}{2} = \dfrac{9}{14}$

    (c) $P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) = \dfrac{4}{7} \times \dfrac{2}{3} + \dfrac{5}{7} \times \dfrac{1}{3} = \dfrac{13}{21}$

    (d) Now it is possible that the first drawn ball is black or white. Accordingly the probability of the second draw will change. Let event *FB*: first drawn ball is black; event *FB^C*: first drawn ball is not black i.e. it is white.

    $$P(B) = P(FB) \times P(B|FB) + P(FB^C) \times P(B|FB^C) = \frac{4}{7} \times \frac{6}{8} + \frac{3}{7} \times \frac{5}{8} = \frac{39}{56}$$

Note: We can use tree diagram as a help in solving such problems.

7. Let M_1, M_2 and M_3 be the events that the product is manufactured on machines M_1, M_2 and M_3 respectively. Let D be the event that the item is defective. The given information can be written as,

    $P(M_1) = 0.5, \quad P(M_1) = 0.3, \quad P(M_1) = 0.2$

    $P(D|M_1) = 0.03 \quad P(D|M_2) = 0.04$ and $\quad P(D|M_3) = 0.05$

    We know that the selected item is defective. Therefore, by Bayes' theorem probability that the item is produced on machine M_2 is,

    $$P(M_2 \Big| D) = \frac{P(M_2)P(D|M_2)}{P(M_1)P(D|M_1) + P(M_2)P(D|M_2) + P(M_3)P(D|M_3)}$$

    $$= \frac{0.3 \times 0.04}{0.5 \times 0.03 + 0.3 \times 0.04 + 0.2 \times 0.05} = 0.324$$

8. Probability of first digit as 1 is, $P(A_1) = \dfrac{2}{4} = \dfrac{1}{2}$

    Probability of second digit as 1 is, $P(A_2) = \dfrac{2}{4} = \dfrac{1}{2}$

    Probability of third digit as 1 is, $P(A_3) = \dfrac{2}{4} = \dfrac{1}{2}$

    Now, $P(A_1 \cap A_2 \cap A_3) = 0 \qquad$ Also, $P(A_1)P(A_2)P(A_3) = \dfrac{1}{8}$

    Since $P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3)$ hence, A_1, A_2 and A_3 are not mutually independent. (They are dependent).

    Now, $P(A_2 \Big| A_1) = \dfrac{1}{2}$

Since $P(A_2|A_1) = P(A_1)$, A_1, and A_2 are pair-wise independent.

Similarly, $P(A_3|A_1) = P(A_1)$ and $P(A_2|A_3) = P(A_2)$. Hence, A_1 and A_3 as well as A_2 and A_3 are pair-wise independent.

Note that $P(A_3|A_1A_2) = 0 \neq P(A_3)$. Hence, A_1, A_2 and A_3 together are not mutually independent.

9. Let *I*, *II*, and *III* be the three events that the vans I, II and III are available. Probability that at least one recovery van will be available *P* is the union of these probabilities. Further, since probabilities of availability of vans are independent, their joint probability is the product of individual probabilities. Thus,

$$P(I \cup II \cup III) = P(I) + P(II) + P(III) - P(I \cap II) - P(I \cap III) - P(II \cap III) + P(I \cap II \cap III)$$

$$= P(I) + P(II) + P(III) - P(I) \times P(II) - P(I) \times P(III) - P(II) \times P(III) + P(I) \times P(II) \times P(III)$$

$$= 0.9 + 0.7 + 0.8 - 0.63 - 0.72 - 0.56 + 0.504 = 0.994$$

10. Let *H* and *W* be the two events that husband and wife get selected respectively. Both these events are independent. Hence, their joint probability is the product of their respective probability.

    (a) Probability that both of them selected,

    $$P(H \cap W) = P(H) \times P(W) = 0.4 \times 0.7 = 0.28$$

    (b) Probability that only one of them will be selected is,

    $$P = P(H \cup W) - P(H \cap W) = P(H) + P(W) - 2 \times P(H \cap W) = 0.4 + 0.7 - 0.56 = 0.54$$

    (c) Probability that only one them will be selected is,

    $$P[(H \cup W)'] = 1 - P(H \cup W) = 1 - P(H) - P(W) + P(H \cap W) = 1 - 0.4 - 0.7 + 0.28 = 0.18$$

    Note: This is same as [1- (result of i plus result of ii)]

11. Let '*A*' denote failing in P & C and '*B*' denote failing in Data Structure. The given is,

    $$P(A) = 0.2, \quad P(B) = 0.1, \quad P(A \cap B) = 0.05$$

    Now, $\qquad P(A \cap B) \neq P(A) \times P(B)$

    Hence, two events 'failing in P & C' and 'failing in Data Structure' are **not independent.**

12. Let '*A*' denote failing in P & C and '*B*' denote failing in Data Structure. The given is,

    $$P(A) = 0.2, \quad P(B) = 0.1, \quad P(A \cap B) = 0.02$$

    Now, $\qquad P(A \cap B) = P(A) \times P(B)$

    Hence, two events 'failing in P & C' and 'failing in Data Structure' are **independent.**

12.12 Further Readings

Books   Béla Bollobás, *Modern graph theory*, Springer

Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph theory, Combinatorics, and algorithms*, Birkhäuser

Online links   http://www.stat.psu.edu/~jiali/course/stat416/notes/chap1.pdf

http://www.mathgoodies.com/lessons/vol6/conditional.html

http://www.mathgoodies.com/lessons/vol6/independent_events.html

Unit 13: Random Variables

CONTENTS

Objectives

Introduction

Objectives

After studying this unit, you will be able to:

- Know about random variables

- Explain Baye's theorm

- Understand Probability Mass Functions

Introduction

Many problems are concerned with values associated with outcomes of random experiments. For example, we select ten items randomly from the production lot with known proportion of defectives and put them in a packet. Now we want to know the probability that the packet contains more than one defective item. The study of such problems requires a concept of a random variable. Random Variable is a function that associates numerical values to the outcomes of experiments.

13.1 Probability Distribution

Each outcome i of an experiment has a probability $P(i)$ associated with it. Similarly, every value of random variable $X = x_i$ is related to the outcome i of an experiment. Hence, for every value of random variable x_i, we have a unique real value $P(i)$ associated. Thus, every random variable X has probability P associated with it. This function $P(X=x_i)$ from the set of all events of the sample space S is called a probability distribution of the random variable.

The probability distribution (or simply distribution) of a random variable X on a sample space S is set of pairs $(X=x_i, P(X = x_i))$ for all $x_i: \in x(S)$, where $P(X = x_i)$ is the probability that X takes the value x_i.

Did u know? What is probability distribution?

Example: A random variable is number of tails when a coin is flipped thrice. Find probability distribution of the random variable.

Solution:

Sample space is HHH, THH, HTH, HHT, TTH, THT, HTT, TTT

The required probability distribution is,

Value of Random Variable	$X = x_i$	0	1	2	3
Probability	$P(X = x_i)$	$\dfrac{1}{8}$	$\dfrac{3}{8}$	$\dfrac{3}{8}$	$\dfrac{1}{8}$

Example: A random variable is sum of the numbers that appear when a pair of dice is rolled. Find probability distribution of the random variable.

Solution:

$X(1, 1) = 2$, $X (1, 2) = X (2, 1) = 3, X (1, 3) = X (2, 2) = X (3, 1) = 4$ etc.

Thus, probability distribution is,

$X = x_i$	2	3	4	5	6	7	8	9	10	11	12
$P(X=x_i)$	$\dfrac{1}{36}$	$\dfrac{2}{36}$	$\dfrac{3}{36}$	$\dfrac{4}{36}$	$\dfrac{5}{36}$	$\dfrac{6}{36}$	$\dfrac{5}{36}$	$\dfrac{4}{36}$	$\dfrac{3}{36}$	$\dfrac{2}{36}$	$\dfrac{1}{36}$

13.2 Discrete Random Variable

A random variable X is said to be discrete if it takes finite or countably infinite number of possible values. Thus, discrete random variable takes only isolated values. Random variables mentioned in the previous examples are discrete random variables:

Some of the practical examples of discrete random variable are,

1.   Number of accidents on an expressway.

2.   Number of cars arriving at a petrol pump.

3.   Number of students attending class.

4.   Number of customers arriving at a shop.

5.   Number of neutrons emitted by a radioactive isotope.

Task   Analyse two more practical examples of discrete random variables in your home.

Probability Mass Function (PMF)

Let X be a discrete random variable defined on a sample space S. Suppose $\{x_1, x_2, …, x_n\}$ is the range set of X. With each of x_i, we assign a number $P(x_i) = P(X = x_i)$ called the probability mass function (PMF) such that,

$$P(x_i) \geq 0 \text{ for } i = 1, 2, …, n \text{ and}$$

$$\sum_{i=1}^{n} P(x_i) = 1$$

The table containing the value of X along with the probabilities given by probability mass function (PMF) is called probability distribution of the random variable X.

Example: Let X represent the difference between the number of heads and the number of tails obtained when a coin is tossed n times. What are the possible values of X?

Solution:

When a coin is tossed n times, number of heads that can be obtained are $n, n – 1, n – 2, …., 2, 1, 0$. Corresponding number of tails are $0, 1, 2,….., n - 2, n - 1, n$. Thus, the sum of number of heads and number of tails must be equal to number of trials n.

Hence, values of X are from n to $– n$ as $n, n – 2, n – 4, ------ , n – 2r$ Where $r = 0, 1, 2,…., n$

Notes   If *n* is even X has one of its value as zero also. However, if *n* is odd X has values –1, 1 but not zero.

Example: Let X represent the difference between the number of heads and the number of tails obtained when a fair coin is tossed 3 times. What are the possible values of X and its PMF?

Solution:

Coin is fair. Therefore, probability of heads in each toss is $P_H = \dfrac{1}{2}$. Similarly, probability of tails in each toss is $P_T = \dfrac{1}{2}$. X can take values *n* – 2*r* where *n* = 3 and *r* = 0, 1, 2, 3.

e.g. X = 3 (HHH), X = 1 (HHT, HTH, THH), X = -1 (HTT, THT, HTT) and X = -3 (TTT)

Thus, the probability distribution of X (possible values of X and PMF) is,

X = x_i	-3	-1	1	3	Total $= \sum\limits_{i=1}^{4} P(x_i)$
p.m.f. P(X_i)	$\dfrac{1}{8}$	$\dfrac{3}{8}$	$\dfrac{3}{8}$	$\dfrac{1}{8}$	1

Example: Suppose a fair dice is rolled twice. Find the possible values of random variable X and its associated PMF, if:

1.   X is the maximum of the values appearing in two rolls.

2.   X is the minimum of the values appearing in two rolls.

3.   X is the sum of the values appearing in two rolls.

4.   X is value appearing in the first roll minus the value appearing in the second roll.

Solution:

Since the dice is fair, probability of any of the outcome of pair of numbers (1, 1), (1, 2), (1, 2), etc. up to (6, 5), (6, 6) appearing in two rolls is $\dfrac{1}{36}$.

1.   Now, X can take values from 1 to 6. The probabilities of X taking these values can be calculated by adding the probabilities of various outcomes that give the particular value of X.

    For example, X = 4 i.e. 'maximum of the two values appearing on dice is four' can be obtained with outcomes {(1, 4), (4, 1), (2, 4), (4, 2), (3, 4), (4, 3), (4, 4)} i.e. 7 ways. Hence,

    P{X = 4} = $\dfrac{7}{36}$.

The values and probability distribution of the random variable X is:

X = x_i	1	2	3	4	5	6	$\sum P(x_i)$
p.m.f. P(x_i)	$\dfrac{1}{36}$	$\dfrac{3}{36}$	$\dfrac{5}{36}$	$\dfrac{7}{36}$	$\dfrac{9}{36}$	$\dfrac{11}{36}$	1

2. Now, X can take values from 1 to 6. Their probabilities can be calculated by adding the probabilities of various outcomes that give the value of X.

    For example, X = 4 i.e. 'minimum of the two values appearing on dice is four' can be obtained with outcomes {(4, 4), (4, 5), (5, 4), (4, 6), (6, 4)} i.e. 5 ways.

    Hence, P{X = 4} = $\dfrac{5}{36}$

    The values and probability distribution of the random variable X is:

X = x_i	1	2	3	4	5	6	$\sum P(x_i)$
p.m.f. P(x_i)	$\dfrac{11}{36}$	$\dfrac{9}{36}$	$\dfrac{7}{36}$	$\dfrac{5}{36}$	$\dfrac{3}{36}$	$\dfrac{1}{36}$	1

3. Now, X can take values from 2 to 12. Their probabilities can be calculated by adding the probabilities of various outcomes that give the value of X.

    For example, X = 6 i.e. 'sum of the two values appearing on dice is six' can be obtained with outcomes {(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)} i.e. 5 ways.

    Hence, P{X = 6} = $\dfrac{5}{36}$

    The values and probability distribution of the random variable X is:

X = x_i	2	3	4	5	6	7	8	9	10	11	12	$\sum P(x_i)$
p.m.f. P(X = x_i)	$\dfrac{1}{36}$	$\dfrac{2}{36}$	$\dfrac{3}{36}$	$\dfrac{4}{36}$	$\dfrac{5}{36}$	$\dfrac{6}{36}$	$\dfrac{5}{36}$	$\dfrac{4}{36}$	$\dfrac{3}{36}$	$\dfrac{2}{36}$	$\dfrac{1}{36}$	1

4. Now, X can take values from –5 to 5. Their probabilities can be calculated by adding the probabilities of various outcomes that give the value of X.

    For example, X = 4 i.e. 'value appearing in first roll minus value appearing in second roll on dice is four' can be obtained with outcomes {(5, 1), (6, 2)} i.e. 2 ways.

    Hence, P{X = 4} = $\dfrac{2}{36}$

    The values and probability distribution of the random variable X is:

X = x_i	–5	–4	–3	–2	–1	0	1	2	3	4	5	$\sum P(x_i)$
p.m.f. P(X = x_i)	$\dfrac{1}{36}$	$\dfrac{2}{36}$	$\dfrac{3}{36}$	$\dfrac{4}{36}$	$\dfrac{5}{36}$	$\dfrac{6}{36}$	$\dfrac{5}{36}$	$\dfrac{4}{36}$	$\dfrac{3}{36}$	$\dfrac{2}{36}$	$\dfrac{1}{36}$	1

13.3 Continuous Random Variable

Random variables could also be such that their set of possible values is uncountable. Examples of such random variables are time between arrivals of two vehicles at petrol pump or time taken for an angioplasty operation at a hospital or lifetime of a component.

Probability Density Function (PDF)

Like we have PMF for discrete random variable, we define PDF for continuous random variable. Let X be a continuous random variable. Function defined for all real $x \in (-\infty, \infty)$ is called probability density function (PDF) if for any set B of real numbers, we get probability,

$$P\{X \in B\} \quad = \quad \int_B f(x)dx$$

All probability statements about X can be answered in terms of $f(x)$. Thus,

$$P\{a \leq X \leq b\} \quad = \quad \int_a^b f(x)dx$$

Notes   Probability of a continuous random variable at any particular value is zero, since

$$P\{X = a\} = P\{a \leq X \leq a\} = \int_a^a f(x)dx = 0$$

13.4 Properties of Random Variables and their Probability Distributions

Properties of a random variable can be studied only in terms of its PMF or PDF We need not refer to the underlying sample space, once we have probability distribution of a random variable. Since the random variable is a function relating all outcomes of a random experiment, the probability distribution of random variable must satisfy the axioms of probability. These in case of discrete and continuous random variable are stated as,

Axiom-I: Any probability must be between zero and one.

For discrete random variable, this can be stated as,

$$0 \leq p(x_i) \leq 1$$

For continuous random variable, this axiom can be stated for any real numbers a and b as,

$$0 \leq P\{a \leq x \leq b\} \leq 1 \qquad \text{or}$$

$$0 \leq \int_a^b f(x)dx \leq 1$$

Axiom II: Total probability of sample space must be one.

For discrete random variable, this can be stated as,

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

For continuous random variable, this axiom can be stated as,

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

Axiom III: For any sequence of mutually exclusive events E_1, E_2, ...etc. i.e. $E_i \cap E_j = \phi$ for $i \neq j$, probability of a union set of events is sum of their individual probabilities. This axiom can also be written as $P(E_1 E_2) = P(E_1) + P(E_2)$, where E_1 and E_2 are mutually exclusive events.

For discrete random variable, this can be stated as,

$$P\left(\bigcup_{i=1}^{\infty} E_i \right) = \sum_{i=1}^{\infty} P(E_i)$$

For continuous random variable, this axiom can be stated as,

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx \qquad \text{also,}$$

$$P(a \leq x \leq b \cup c \leq x \leq d) = \int_a^b f(x)dx + \int_c^d f(x)dx$$

Example: Find the probability between $X = 1$ and 2 i.e. $P(1 \leq X \leq 2)$ for a continuous random variable whose PDF is given as,

$$f(x) = \frac{1}{6}x + k \qquad \text{for } 0 \leq x \leq 3$$

Solution:

Now, PDF must satisfy probability Axiom II. Thus,

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

Or,
$$\int_0^3 (\frac{1}{6}x + k)dx = \left[\frac{x^2}{12} + kx\right]_0^3 = \frac{3}{4} + 3k = 1$$

$$k = \frac{1}{12}$$

Now,
$$P(1 \leq X \leq 2) = \int_1^2 f(x)dx = \int_1^2 (\frac{1}{6}x + \frac{1}{12})dx = \frac{1}{3}$$

📝 *Example:* PDF of a continuous random variable is given by,

$$f(x) \;=\; \begin{cases} kx(2-x) & 0 < x < 2 \\ 0 & otherwise \end{cases}$$

(a)   Find k

(b)   Find $P(x < \dfrac{1}{2})$

Solution:

(a)   According to the Axiom II of probability,

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

Or, $$\int_{0}^{2} kx(2-x)dx \;=\; \left[kx^2 - \frac{kx^3}{3} \right]_{0}^{2} = 1$$

Or, $$k \;=\; \frac{3}{4}$$

(b) $$P\left(x < \frac{1}{2}\right) \;=\; \int_{-\infty}^{\frac{1}{2}} f(x)dx = \int_{0}^{\frac{1}{2}} \frac{3}{4}x(2-x)dx = \left[\frac{3}{4}x^2 - \frac{x^3}{4} \right]_{0}^{2} = \frac{5}{32}$$

13.5 Cumulative Distribution Function (CDF)

Another important concept in probability distribution of random variable is the cumulative distribution function (CDF) or just a distribution function. It is the accumulated value of the probability up to a given value of the random variable. Let X be a random variable, then the cumulative distribution function (CDF) is defined as a function $F(a)$ such that,

$$F(a) = P\{X \le a\}$$

13.5.1 Cumulative Distribution Function (CDF) for Discrete Random Variable

Let X be a discrete random variable defined on a sample space S taking values $\{x_1, x_2 \ldots, x_n\}$ with probabilities $p(x_1), p(x_2), \ldots p(x_n)$ respectively. Then, cumulative distribution function (CDF) denoted as $F(a)$ and expressed in terms of PMF as,

$$F(a) \;=\; \sum_{x_i \le a} p(x_i)$$

Notes

1. The CDF is defined for all values of $x_i \in R$. However, since the random variable takes only isolated values, the CDF is constant between two successive values of X and has steps at the points x_i, i =1, 2, …,n. Thus, the CDF for a discrete random variable is a step function.

2. $F(\infty) = 1$ and $F(-\infty) = 0$

3. Properties of a random variable can be studied only in terms of its CDF We need not refer to the underlying sample space or PMF, once we have CDF of a random variable.

13.5.2 Cumulative Distribution Function (CDF) for Continuous Random Variable

Let X be a continuous random variable defined on a sample space S with PDF $f(x)$. Then cumulative distribution function (CDF) denoted as $F(a)$ and expressed in terms of PDF as,

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

Also, differentiating both sides we get,

$$\frac{d}{da}F(a) = f(a)$$

Thus, the density is the derivative of the cumulative distribution function.

13.5.3 Properties of CDF

1. $F(a)$ is non-decreasing function (monotonously increasing function) i.e.

   For, $a < b$, $F(a) \le F(b)$

2. $\lim_{a \to \infty} F(a) = 1$

3. $\lim_{a \to -\infty} F(a) = 0$

4. F is a right continuous.

5. $P(X = a) = F(a) - \lim_{n \to \infty} F(a - \frac{1}{n})$

6. $P(a < X \le b) = F(b) - F(a)$

7. For pure continuous random variable, $P(X = a) = \int_{a}^{a} f(x)dx = 0$

8. $P(a \le X \le b) = F(b) - F(a) + P(a)$ Note that in case of continuous random variable, $P(a) = 0$

9. $P(a \le X < b) = F(b) - F(a) - P(b) + P(a)$ Note that in case of continuous random variable, $P(a) = 0, P(b) = 0$

10. $P(a < X < b) = F(b) - F(a) - P(b)$ Note that in case of continuous random variable, $P(b) = 0$

11. $P(X > a) = 1 - P(X \le a) = 1 - F(a)$

12. $P(X \ge a) = 1 - P(X \le a) + P(X = a) = 1 - F(a) + P(a)$ Note that in case of continuous random variable, $P(a) = 0$

Example: A random variable is number of tails when a coin is flipped thrice. Find PMF and CDF of the random variable.

Solution:

Sample space for the given random variable is HHH, THH, HTH, HHT, TTH, THT, HTT, TTT.

First we calculate the PMF of the random variable. Then CDF is calculated by taking partial sums of the probabilities of values of random variable 'less than or equal to' a particular value. The CDF is shown below:

Random Variable	$X = x_i$	0	1	2	3
p.m.f.	$P(X = x_i)$	$\dfrac{1}{8}$	$\dfrac{3}{8}$	$\dfrac{3}{8}$	$\dfrac{1}{8}$
c.d.f.	$F(a) = P\{X = x_i \le a\}$	$\dfrac{1}{8}$	$\dfrac{4}{8} = \dfrac{1}{2}$	$\dfrac{7}{8}$	1

While describing the CDF, we write as follows:

CDF $\qquad F(a) \;=\; 0 \qquad\qquad\qquad\qquad -\infty < a < 0$

$\qquad\qquad\qquad = \;\dfrac{1}{8} \qquad\qquad\qquad\qquad 0 \le a < 1$

$\qquad\qquad\qquad = \;\dfrac{4}{8} \qquad\qquad\qquad\qquad 1 \le a < 2$

$\qquad\qquad\qquad = \;\dfrac{7}{8} \qquad\qquad\qquad\qquad 2 \le a < 3$

$\qquad\qquad\qquad = \; 1 \qquad\qquad\qquad\qquad 3 \le a < \infty$

Example: CDF of a random variable is given as,

$\qquad\qquad F(a) \;=\; 0 \qquad\qquad\qquad\qquad -\infty < a < 0$

$\qquad\qquad\qquad = \;\dfrac{1}{2} \qquad\qquad\qquad\qquad 0 \le a < 1$

$\qquad\qquad\qquad = \;\dfrac{2}{3} \qquad\qquad\qquad\qquad 1 \le a < 2$

13.6.3 Properties of Expectation

Effect of Change of Origin and Scale on $E(X)$

$$E[aX + b] = aE(X) + b$$

Proof

Let a random variable $Y = aX + b$ where a and b are constants and X be a random variable. Then Y has the same probability distribution that of X.

For discrete random variable, by definition,

$$E(Y) = \sum_{i=1}^{n} y_i \times P(y_i)$$

But $y_i = ax_i + b$ all i. Therefore,

$$E(aX + b) = \sum_{i=1}^{n}(ax_i + b)P(x_i) = a\sum_{i=1}^{n} x_i \times P(x_i) + b\sum_{i=1}^{n} P(x_i)$$

$$= aE(X) + b$$

$$\therefore \qquad \sum_{i=1}^{n} P(x_i) = 1 \text{ By Axiom II}$$

For continuous random variable, by definition,

$$E[Y] = \int_{-\infty}^{\infty} y \, f(y) \, dy$$

But, $y = ax + b$ for all real x and y. Therefore,

$$E[aX + b] = \int_{-\infty}^{\infty} (ax + b) f(x) dx$$

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

$$= aE(X) + b$$

Notes

1.   Put $a = 0$ and we get $E(b) = b$ where b is a constant.

2.   Put $b = 0$ and we get $E(aX) = aE(X)$ where a is a constant.

3.   Put $a = 1$ and we get $E(X + b) = E(X) + b$.

Expected Value of Constant is the Constant Itself

Thus, $\qquad E(C) \;=\; C$

Proof

For discrete random variable, by definition,

$$E(X) \;=\; \sum_{i=1}^{n} x_i \times P(x_i)$$

But $x_i = C$ for all i. Therefore,

$$E(C) \;=\; \sum_{i=1}^{n} C \times P(x_i) = C \sum_{i=1}^{n} P(x_i) = C \qquad \because \sum_{i=1}^{n} p(x_i) = 1 \qquad \text{By Axiom II}$$

For continuous random variable, by definition,

$$E(X) \;=\; \int_{-\infty}^{\infty} x f(x) dx$$

But $x = C$. Therefore,

$$E(C) \;=\; \int_{-\infty}^{\infty} C f(x) dx = C \int_{-\infty}^{\infty} f(x) dx = C \qquad \because \int_{-\infty}^{\infty} f(x) dx = C \ \text{By Axiom II}$$

Expectation of a Function

Let, $Y = g(X)$ is a function of a random variable X, then Y is also a random variable with the same probability distribution of X

For discrete random variable X and Y, probability distribution of Y is also $P(x_i)$. Thus, the expectation of Y is,

$$E(Y) \;=\; E[g(x_i)] = \sum_{i=1}^{\infty} g(x_i) P(x_i)$$

For continuous random variable X, and Y probability distribution of Y is also $f(x)$. Thus, the expectation of Y is,

$$E(Y) \;=\; E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Example: A random variable is number of tails when a coin is flipped thrice. Find expectation (mean) of the random variable.

Solution:

The required probability distribution is,

Random Variable	$X = x_i$	0	1	2	3
p.m.f.	$P(X = x_i)$	$\dfrac{1}{8}$	$\dfrac{3}{8}$	$\dfrac{3}{8}$	$\dfrac{2}{8}$
	$x_i \times P(x_i)$	0	$\dfrac{3}{8}$	$\dfrac{6}{8}$	$\dfrac{3}{8}$

Now the expectation of the random variable is,

$$E(X) \;=\; \sum_{i=1}^{4} x_i \times P(x_i) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{12}{8} = \frac{3}{2}$$

Example: X is a random variable with probability distribution

$X = x_i$	0	1	2
$P(X = x_i)$	0.3	0.3	0.4

$$Y \;=\; g(X) = 2X + 3$$

Find expected value or mean of Y that is E(y).

Solution:

Now for $X = 0, 1, 2$ $Y = 3, 5, 7$ respectively. Hence, the distribution of Y is,

$X = x_i$	0	1	2
$Y = y_i$	3	5	7
$p(Y = y_i)$	0.3	0.3	0.4

Hence,

$$E(Y) \;=\; E[g(x_i)] = \sum_{i=1}^{n} g(x_i)P(x_i) = \sum_{i=1}^{n} y_i P(x_i)$$

$$= \; 3 \times 0.3 + 5 \times 0.3 + 7 \times 0.4$$

$$= \; 5.2$$

13.7 Variance of a Random Variable

The expected value of a random variable X, namely $E(X)$ provides a measure of central tendency of the probability distribution. However, it does not provide any idea regarding the spread of the distribution.

Suppose we have two coffee packet filling machines that fill 200 gms packets. You promise the customers that you would give one packet free as penalty if the coffee is short of the specified weight of 200 gms by say 5 gms. Due to random process weight of coffee in each packet follows

a random distribution. Let *X* be a random variable denoting the weight of the coffee with distribution for two machines as follows:

Machine A

X = x_i	190	195	200	205	210
P(X = x_i)	0.1	0.2	0.4	0.2	0.1

Machine B

X = x_i	198	199	200	201	202
P(X = x_i)	0.1	0.2	0.4	0.2	0.1

Mean or expected value of weight of the coffee in a packet is 200 gms for both the machines. However, the spread of the distribution for machine 1 is greater. Since the machine 2 gives much lesser spread, we consider it more accurate and prefer it. Obviously, we will have to spend less as penalty and also at the same time less coffee will be filled as excess in over-weight packets. Thus, in this case lesser the spread is preferable. (Of course such a machine may cost more in initial investment).

Hence, besides the central tendency or mean, we need a measure of a spread of a probability distribution. As we expect *X* to take values around its mean *E*(*X*), it would appear that a reasonable way to measure a spread is to find an average value of the deviation i.e. how far each value is from the mean. This quantity is called mean deviation and can be mathematically expressed as $E(X - \mu)$ where $\mu = E(X)$. However, this quantity is always zero for any distribution. Hence, we either take the deviation from the median or take absolute value of deviation as $E(|X - \mu|)$. Although this quantity gives a good measure of dispersion or spread, it is not amenable to further mathematical treatment. Hence, more tractable quantity of mean of a squared deviation is considered. It is called variance of a random variable.

13.7.1 Definition of a Variance of Random Variable

Let X be a discrete random variable on a sample space S. The variance of X denoted by $Var(X)$ or σ^2 is defined as,

$$Var(X) \quad = \quad E[(X - \mu)^2] = E[(X - E(X))^2] = \sum (x_i - \mu)^2 P(x_i)$$

Further, it can be shown that,

$$Var(X) \quad = \quad E\left[X^2\right] - \left(E\left[X\right]\right)^2$$

Proof:

Let X is random variable with mean $\mu = E[X]$, then the variances of X is,

$$Var(X) \quad = \quad E\left[\left(X - \mu\right)^2\right]$$

For discrete random variable,

$$Var(X) \quad = \quad \sum_i (x_i - \mu)^2 P(x_i) = \sum_i \left(x_i^2 - 2\mu x_i + \mu^2\right) P(x_i)$$

$$= \quad \sum_i x_i^2 P(x_i) - 2\mu \sum_i x_i P(x_i) + \mu^2 \sum_i P(x_i)$$

$$= E\left[X^2\right] - 2\mu^2 + \mu^2$$

$$= E\left[X^2\right] - \left(E[X]\right)^2$$

For continuous random variable.

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx = \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)dx$$

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

$$= E(X^2) - 2\mu E(X) - [E(X)]^2]$$

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

Since dimensions of variance are square of dimensions of *X*, for comparison, it is better to take a square root of variance. It is known as standard deviation and denoted by S.D.(*X*), or σ that is,

$$S.D. = \sigma = \sqrt{Var(x)}$$

Notes

1. Analogous to the mean is the center of gravity of a distribution of mass.

2. Analogous to the variance is the moment of inertia.

13.7.2 Properties of Variance

$$Var(g(X)) = E([g(X)]^2) - [E(g(X))]^2$$

Effect of Change of Origin and Scale on Variance

$$Var(aX + b) = a^2 Var(X) = a^2 \sigma^2$$

Proof:

By definition of variance,

$$Var(aX + b) = E[(aX + b) - E(aX + b)]^2$$

$$= E[(aX + b) - (aE(X) + b)]^2 = E[a(X - E(X))]^2$$

$$= E[a^2(X - E(X)^2] = a^2 E[(X - E(X)^2] = a^2 Var(X)$$

$$= a^2 \sigma^2$$

Notes

1. Put $a = 0$ and we get $Var(b) = 0$ where b is a constant.

2. Put $b = 0$ and we get $Var(aX) = a^2 Var(X)$ where a is a constant.

3. Put $a = 1$ and we get $Var(X + b) = Var(X)$

Standardised Random Variable

If X is a random variable with mean m and standard deviation σ and random variable Y is defined as $Y = \dfrac{X - \mu}{\sigma}$ then, mean and variance of distribution of Y are zero and one respectively.

Proof:

$$E(Y) = E\left[\frac{x - \mu}{\sigma}\right] = \frac{1}{\sigma}[E(X) - \mu] = 0$$

and,

$$Var(Y) = \frac{1}{\sigma^2} Var(X) = \frac{\sigma^2}{\sigma^2} = 1$$

Example: Suppose we have two coffee packet filling machines that fill 200 gms packets. You promise the customers that you would give one packet free as a penalty if the coffee is short of the specified weight of 200 gms by 5 gms. Due to random process weight of coffee in each packet follows a random distribution. Let X be a random variable denoting the weight of the coffee with distribution for two machines as follows:

Machine A

$X = x_i$	190	195	200	205	210
$P(X = x_i)$	0.1	0.2	0.4	0.2	0.1

Machine B

$X = x_i$	198	199	200	201	202
$P(X = x_i)$	0.1	0.2	0.4	0.2	0.1

Find the mean and variance of the weight these coffee packs will have. Which of the machine will you prefer?

Solution:

Machine A

$X = x_i$	190	195	200	205	210	**Total**
$P(X = x_i)$	0.1	0.2	0.4	0.2	0.1	**1**
$x_i P(x_i)$	19	39	80	41	21	200
$x_i^2 P(x_i)$	3610	7605	16000	8405	4410	40030

Thus, the mean is, $\mu = E(X) = \sum_{all} x_i P(x_i) = 200$ **Ans.**

Also, $E(X^2) = \sum_{all} x_i^2 P(x_i) = 40030$

Hence, $Variance = E(X^2) - [E(X)]^2 = 40030 - 40000 = 30$ **Ans.**

Now, $S.D. = \sigma = \sqrt{Variance} = \sqrt{30} = 5.48$

Machine B

$X = x_i$	198	199	200	201	202	**Total**
$P(X = x_i)$	0.1	0.2	0.4	0.2	0.1	**1**
$x_i P(x_i)$	19.8	39.8	80	40.2	20.2	200
$x_i^2 P(x_i)$	3920.4	7920.2	16000	8080.2	4080.4	40001.2

Thus, the mean is, $\mu = E(X) = \sum_{all} x_i P(x_i) = 200$ **Ans.**

Also, $E(X^2) = \sum_{all} x_i^2 P(x_i) = 40001.2$

Hence, $Variance = E(X^2) - [E(X)]^2 = 40001.2 - 40000 = 1.2$ **Ans.**

Now, $S.D. = \sigma = \sqrt{Variance} = \sqrt{1.2} = 1.1$

From the above result, it can be seen that machine B is preferable since it has very small variance as compared to the machine A. In fact, we could roughly say that in case of machine A we will have to give free packets as a penalty for about 27% of the customers. In case of machine B, not even 1% customers will get coffee pack that is under weight by 5 gms. Also, the coffee in over weight packs from machine B will also be very small quantity as compared to machine A and hence, less costly.

13.8 Moments of Random Variable

So far we have studied mean and variance of a random variable. The mean is a measure of central tendency and variance measures dispersion. However, to get complete information about the probability distribution, we also have to study the shape in terms of symmetry (skewness) and peakedness (Kurtosis). Moments of random variable serve this purpose.

13.8.1 Moments about the Origin (Raw Moments)

E (X) is also referred to as the First moment of X about the origin. The rth raw moment of X i.e. r^{th} moment about the origin is defined as,

For discrete random variable,

$$\mu_r' = E(X^r) = \sum_{i=1}^{n} x_i^r P(x_i) \qquad\qquad r = 1, 2, 3, \dots..$$

For continuous random variable,

$$\mu_r' = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx$$

In particular,

$$\mu_1' = E(X) = mean$$

$$\mu_2' = E(X^2) = \sum_{i=1}^{n} x_i^2 P(x_i)$$

or,

$$\mu_2' = E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$$

$$\mu_3' = E(X^3) = \sum_{i=1}^{n} x_i^3 P(x_i)$$

or,

$$\mu_3' = E(X^3) = \int_{-\infty}^{\infty} x^3 f(x)dx$$

And so on, $E(X^n)$ is called n^{th} moment of X about origin. First raw moment is like center of gravity, second raw moment is like moment of inertia, and so on.

13.8.2 Moments about any Arbitrary Point '*a*'

The rth moment of random variable X about a point '*a*' is defined as:

For discrete random variable,

$$\mu_r'(a) = E[(X-a)^r] = \sum_{i=1}^{n} (x_i - a)^r P(x_i) \qquad\qquad r = 1, 2, 3, ...$$

For continuous random variable,

$$\mu_r'(a) = E[(X-a)^r] = \int_{-\infty}^{\infty} (x-a)^r f(x)dx$$

In particular, first moment about *a*, is first moment about origin minus *a*. Or

$$\mu_1'(a) = E[(X-a)] = E[X] - a$$

Also,

$$\mu_2'(a) = E[(X-a)^2] = \mu_2' - 2a\mu_1' + a^2$$

13.8.3 Moments about the Arithmetic Mean (Central Moment)

The rth central moment or r^{th} moment about the arithmetic mean of random variable X is defined as,

For discrete random variable,

$$\mu_r = E[(X-\mu)^r] = \sum_{i=1}^{n} (x_i - \mu)^r P(x_i) \qquad\qquad r = 1, 2, 3, ...$$

For continuous random variable,

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

First Moment about the Arithmetic Mean (First Central Moment)

In particular with $r = 1$,

$$\mu_1 = E[(X-\mu)] = E(X) - \mu = \mu - \mu = 0$$

First Moment about the Arithmetic Mean (First Central Moment) is always Zero.

Second Central Moment

With $r = 2$,

$$\mu_2 = E[(X-\mu)^2] = Var(X) \qquad \text{By definition of variance.}$$

Second central moment is variance.

Effect of Change of Origin and Scale on Central Moments

If X and Y are random variables such that $Y = a + bX$, where a and b are constants,

$$\mu_r(Y) = E[(Y-\mu_Y)^r] = E[a + bX - E(a + bX)]^r = E[a + bX - a - bE(X)]^r$$

$$= E(b^r[X - E(X)]^r) = b^r E[(X-\mu_X)^r]$$

$$= b^r \mu_r(X)$$

Thus, central moments are invariant to the change of origin but not to the change of scale.

13.9 Moment Generating Function (MGF)

Moment generating function (MGF) is a very powerful concept that helps knowing completely about the probability distribution of a random variable through a single equation. Moreover, it is much easy to calculate mean and variance using MGF for many random variables than calculating directly.

Moment generating function (MGF) denoted as $M(t)$ of the random variable X is defined for all real values of t by,

$$M(t) = E\left[e^{tx}\right]$$

Thus, for discrete random variable,

$$M(t) = \sum_i e^{tx_i} p(x_i)$$

and for continuous random variable,

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

$M(t)$ is called the moment generating function because all the moments of X can be obtained by successively differencing $M(t)$ and then evaluating the result at $t = 0$.

For example, first moment of X, is

$$E[X] \quad = \quad M'(t)_{t=0}$$

Second moment of X, is

$$E[X^2] \quad = \quad M''(t)_{t=0}$$

Proof:

$$M'(t) \quad = \quad \frac{d}{dt}E\left[e^{tx}\right] = E\left[\frac{d}{dt}e^{tx}\right] = E\left[xe^{tx}\right]$$

Here, we have used the rule of differentiation under summation sign where summation limits are independent of the variable of differentiation.

Now putting t = 0 we get,

$$M'(0) \quad = \quad E[X]$$

Also,

$$M''(t) \quad = \quad \frac{d}{dt}M'(t) = \frac{d}{dt}E\left[xe^{tx}\right] = E\left[\frac{d}{dt}xe^{tx}\right] = E\left[x^2e^{tx}\right]$$

Now putting t = 0 we get,

$$M''(0) \quad = \quad E\left[X^2\right]$$

Thus, in general,

$$M^n(0) \quad = \quad E\left[X^n\right]$$

If we know MGF of a random variable probability distribution, we can calculate all the moments of the random variable. Thus, we can get the entire information about the discrete or continuous random variable from any one of the PMF or PDF or CDF or MGF Random variables are uniquely explained by PMF or PDF or .CDF or MGF Random variables are often classified according to their PMF or PDF

Did u know? What is Successive Differencing?

13.10 Bayes' Theorem

Let E and F be two events in a sample space S. We can express *F* as,

$$F = FE \cap FE^C$$

General form of Bayes' Formula

Suppose events E_1, E_2, …,E_n form a partition of a sample space S of random experiment. In other words, E_1, E_2…E_n are mutually exclusive and collectively exhaustive events of sample space S. Suppose F is any other event with $P(F) > 0$, defined on S. Then,

$$P(E_i \,|\, F) \;=\; \frac{P(F/E_i)P(E_i)}{\displaystyle\sum_{i=1}^{n} P(F/E_i)P(E_i)}$$

For an outcome to be in F, it must either be in both F and E or be in F but not in E. Now, FE and FE^C are mutually exclusive, (elements in FE must be in E and elements in FE^C must not be in E). Therefore,

$$P(F) \;=\; P(FE) + P(FE^C) = P(F|E) \times P(E) + P(F|E^c) \times P(E^C)$$

$$\;=\; P(F|E) \times P(E) + P(F|E^c) \times [1 - P(E)] \qquad\qquad …(1)$$

Now, using (1) and conditional probability formula,

$$P(E|F) \;=\; \frac{P(E \cap F)}{P(F)}$$

$$\;=\; \frac{P(EF)}{P(F/E)P(E) + P(F/E^C)P(E^C)}$$

$$\;=\; \frac{P(F/E)P(E)}{P(F/E)P(E) + P(F/E^C)P(E^C)}$$

This is called Bayes' formula.

Case Study

The following is the density function of a continuous random variable X given as;

$$f(x) = \begin{cases} ax & \text{for} & 0 \le x \le 1 \\ a & \text{for} & 1 \le x \le 2 \\ 3a - ax & \text{for} & 2 \le x \le 3 \\ 0 & \text{Otherwise} \end{cases}$$

Find (i) a (ii) Cumulative distribution function

Solution:

(i) The PDF must satisfy the axioms of probability. By Axiom II $\int_{-\infty}^{\infty} f(x)dx = 1$. Hence,

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} axdx + \int_{1}^{2} adx + \int_{2}^{3} (3a - ax)dx = a\left[\frac{x^2}{2}\right]_{0}^{1} + a[x]_{1}^{2} + a\left[3x - \frac{x^2}{2}\right]_{2}^{3} = 1$$

Contd…

---

or,

$$a\left[\frac{1}{2}\right] + a[1] + a\left[\frac{1}{2}\right] = 2a = 1 \quad \text{or,}$$

$$a = \frac{1}{2}$$

(ii) CDF is given by $F(b) = \int_{-\infty}^{b} f(x)dx$

For x < 0

$$F(b) = \int_{-\infty}^{b} (0)dx = 0$$

For, $0 \le x \le 1$

$$F(b) = \int_{-\infty}^{0} (0)dx + \int_{0}^{b} \frac{1}{2}xdx = 0 + \frac{1}{2}\left[\frac{x^2}{2}\right]_{0}^{b} = \frac{b^2}{4}$$

For, $1 \le x \le 2$

$$F(b) = \int_{-\infty}^{0} (0)dx + \int_{0}^{1} \frac{1}{2}xdx + \int_{1}^{b} \frac{1}{2}dx = 0 + \frac{1}{4} + \frac{1}{2}[x]_{1}^{b} = \frac{1}{4} + \frac{1}{2}[b-1] = \frac{b}{2} - \frac{1}{4}$$

For, $2 \le x \le 3$

$$F(b) = \int_{-\infty}^{0} (0)dx + \int_{0}^{1} \frac{1}{2}xdx + \int_{1}^{2} \frac{1}{2}dx + \int_{2}^{b} \frac{1}{2}[3-x]dx = 0 + \frac{1}{4} + \frac{1}{2} + \frac{1}{2}\left[3x - \frac{x^2}{2}\right]_{2}^{b}$$

$$= \frac{3}{4} + \frac{1}{2}\left[\left(3b - \frac{b^2}{2}\right) - (6-2)\right] = \frac{1}{4}(6b - b^2 - 5)$$

Thus, the CDF is,

$$F(b) = \begin{cases} 0 & \text{for} \quad -\infty < x \le 0 \\ \dfrac{b^2}{4} & \text{for} \quad 0 \le x \le 1 \\ \dfrac{b}{2} - \dfrac{1}{4} & \text{for} \quad 1 \le x \le 2 \\ \dfrac{1}{4}(6b - b^2 - 5) & \text{for} \quad 2 \le x \le 3 \\ 1 & \text{for} \quad 3 \le x < \infty \end{cases}$$

13.11 Summary

- Random variable is a real valued mapping. Thus, the function has to be one-one or many-one correspondence. Thus, a random variable assigns a real number to each possible outcome of an experiment.

- Each outcome *i* of an experiment has a probability $P(i)$ associated with it. Similarly, every value of random variable $X = x_i$ is related to the outcome *i* of an experiment.

- A random variable *X* is said to be discrete if it takes finite or countably infinite number of possible values.

- Random variables could also be such that their set of possible values is uncountable.

- Properties of a random variable can be studied only in terms of its PMF or PDF

- Another important concept in probability distribution of random variable is the cumulative distribution (CDF)

- One of the most important concepts in probability theory is that of expectation of a random variable.

- The expected value of a random variable *X*, namely $E(X)$ provides a measure of central tendency of the probability distribution.

- The mean is a measure of central tendency and variance measures dispersion.

- Moment generating function (MGF) is a very powerful concept that helps knowing completely about the probability distribution of a random variable through a single equation.

13.12 Keywords

Arithmetic Mean: The sum of the values of a random variable divided by the number of values.

Continuous Random Variable: In probability theory, a continuous probability distribution is a probability distribution which possesses a probability density function.

Cumulative Distribution Function: In probability theory and statistics, the Cumulative Distribution Function (CDF), or just distribution function, describes the probability that a real-valued random variable *X* with a given probability distribution will be found at a value less than or equal to *x*.

13.13 Self Assessment

1. A random variable *X* has the following PMF

$X = x_i$	0	1	2	3	4	5	6
$P(X = x_i)$	k	$3\,k$	$5\,k$	$7\,k$	$9\,k$	$11\,k$	$13\,k$

Find

(a) k (b) $P(X \geq 2)$ (c) $P(0 < X < 5)$

(d) What is the minimum value of *C* for which $P(X \leq C) > 0.5$?

(e) What is distribution function of X?

2. Determine k such that the following functions are PMFs

   (a) $p(x) = kx \quad for \quad x = 1, 2, 3, ..., 10$

   (b) $p(x) = k\dfrac{2^x}{x!} \quad for \quad x = 0, 1, 2, 3$

   (c) $p(x) = k(2x^2 + 3x + 1) \quad for \quad x = 0, 1, 2, 3$

3. In the game of Two-Finger Morra, 2 players show one or two fingers and simultaneously guess the number of fingers their opponent will show. If only one of the players guesses correctly, he wins an amount (in ₹) equal to the sum of the fingers shown by him and his opponent. If both players guess correctly or if neither guesses correctly, then no money is exchanged. Consider a specified player and denote by X the amount of money he wins in a single game.

   (a) If each player acts independently of the other, and if a player makes his choice of the number of fingers he will hold up and he will guess that his opponent will hold up in such way that each of the 4 possibilities is equally likely, what are the possible values of X and what are their associated probabilities?

   (b) Suppose that each player acts independently of each other, and if each player decides to hold up the same number of fingers that he guesses his opponent will hold up, and if each player is equally likely to hold up 1 or 2 fingers, what are the possible values of X and their associated probabilities?

4. Find the value of C for which random variables have following PMF s. for $i = 1, 2, 3, \ldots\ldots$

   (a) $P(X = i) = C \times 2^{-i}$ (b) $P(X = i) = C \times \dfrac{2^{-i}}{i}$

   (c) $P(X = i) = C \times \dfrac{2^i}{i!}$ (d) $P(X = i) = C \times i^{-2}$

5. Two balls are chosen randomly from an urn containing 8 white, 4 black and 2 orange balls. Suppose that we win Rs. 2 each for black ball selected and we lose Rs. 1 for each white ball selected. Let X be the amount we win.

   (a) What are the possible values of X and the probabilities associated with each value?

   (b) If we play the game 100 times, and to play every time we have to pay Rs. 2 as table money, what is the amount we are expected to get?

   (c) Is the game fair?

6. 5 men and 5 women are ranked according to their scores on an examination. Assume that no two scores are alike and all 10! possible ranking are equally likely. Let X denote the highest ranking achieved by a woman. (For example $X = 2$ if the top ranking woman has the rank 2). Find the probability distribution (PMF) of X.

7. A random variable X taking 4 values with probabilities, $\dfrac{1 + 3x}{4}, \dfrac{1 - x}{4}, \dfrac{1 + 2x}{4}$ and $\dfrac{1 - 4x}{4}$.

   Find the condition on X so that these values represent the PMF of X.

8. PDF of a continuous random variable is given as,

$$f(x) = c(4x - 2x^2) \qquad \text{for } 0 < x < 2$$

$$= 0 \qquad\qquad \text{Otherwise}$$

Find (i) c (ii) $P(X > 1)$

9. X is a continuous random variable with distribution given as (Note: Distribution means CDF)

$$F(x) = \begin{cases} 0 & if & x < 0 \\ \dfrac{x}{2} & if & 0 \le x \le 2 \\ 1 & if & 2 < x \end{cases}$$

Find (i) $P(\dfrac{1}{2} \le X \le \dfrac{3}{2})$ (ii) $P(1 \le X \le 2)$

10. The time in years X required to complete a software project has a PDF given by,

$$f(x) = \begin{cases} kx(1-x) & 0 \le x \le 1 \\ 0 & Otherwise \end{cases}$$

Compute the probability that the project will be completed in less than four months.

11. A random variable is sum of the numbers that appear when a pair of dice is rolled. Find the mean value of the sum we expect.

12. A random variable has MGF $M(t) = \dfrac{3}{3-t}$ Obtain standard deviation.

13. PMF of a random variable is given as $P(X = i) = 2^{-i}$ for i = 1, 2, 3, ……

(a) Verify $P(X = i) = 2^{-i}$ is a PMF

(b) Find mean and variance.

14. X is a continuous random variable with PDF given as,

$$f(x) = \begin{cases} \dfrac{x}{2} & if & 0 \le x \le 2 \\ 0 & Otherwise \end{cases}$$

Find

(a) $P(1 \le x \le 1.5)$ (b) Expectation of X or mean ì

(c) Variance of X (d) CDF

15. If Y is a random variable with PDF,

$$f(y) = \begin{cases} \dfrac{k}{\sqrt{y}} & if & 0 < y < 4 \\ 0 & Otherwise \end{cases}$$

Find

(a)   The value of k (b)   Distribution function of y

(c) $P\,(1 < y < 2)$

16.   The following is the PDF of a continuous random variable X,

$$f(x) = \begin{cases} 0 & for & x < 0 \\ x & for & 0 \le x < 1 \\ 2 - x & for & 1 \le x < 2 \\ 0 & for & 2 \le x \end{cases}$$

Find mean and variance of X.

17.   A random variable X with PDF

$$f(x) = \begin{cases} kx^2 e^{-x} & for & x \ge 0 \\ 0 & Otherwise \end{cases}$$

Find

(a) k (b)   Mean of X (c)   Variance of X

13.14 Review Questions

1.   PMF of a discrete random variable is defined as,

$$P(X = i) = \frac{1}{2^i} \qquad \text{for } i = 1, 2, 3, \ldots\ldots$$

Find

(a)   Mean

(b)   Variance

(c)   P (X even)

(d) $P(X \ge 5)$

(e) $P(X \quad divisible \quad by \quad 3)$

2.   Verify whether the following can be PMF for the given values of X.

(a) $P(x) = \dfrac{1}{4}$ for $x = 0, 1, 2, 3, 4$

(b) $P(x) = \dfrac{x+1}{10}$ for $x = 0, 1, 2, 3$

(c) $P(x) = \dfrac{x^2}{30}$ for $x = 0, 1, 2, 3, 4$

(d) $P(x) = \dfrac{x-2}{5}$ for $x = 0, 1, 2, 3, 4, 5$

3. Continuous random variable X assumes values between 2 and 5 with PDF **Notes**

$$f(x) = \begin{cases} k(1+x) & for \quad 2 \le x \le 5 \\ 0 & otherwise \end{cases}$$

Find :

(a) k

(b) $P(x < 4)$

4. A box contains 5 red and 5 blue balls. Two balls are drawn at random. If they are of the same colour, then you win Rs. 110; if they are of different colours, then you lose ₹ 100. Find,

(a)   The expected value of the amount you win.

(b)   Variance of the amount you win.

5. Following is the CDF of a discrete random variable X.

X	1	2	3	4	5	6	7	8
$F(X \le a)$	0.08	0.12	0.23	0.37	0.48	0.62	0.85	1

(a)   Find PMF of X.

(b)   Find $P(X \le 4)$ and $P(2 \le X \le 6)$

(c)   Find $P(X = 5 | X \ge 3)$

(d)   Find $P(X = 6 | X \ge 4)$

6. A random variable X has the following probability distribution (PMF)

$X = x_i$	0	1	2	3
$P(X = x_i)$	$\dfrac{1}{5}$	$\dfrac{2}{5}$	$\dfrac{1}{5}$	$\dfrac{1}{5}$

Find probability distribution of

(a) $W = X - 1$

(b) $Y = \dfrac{3X + 2}{2}$

(c) $Z = X^2 + 2$

7. A box of 20 mangoes contains 4 bad mangoes. Two mangoes are drawn at random without replacement from this box. Obtain the probability distribution of the number of bad mangoes in the sample.

8. A baker sells on an average 1000 pastries of 5 types every day. Profit from these five types is ₹ 1, 1.5, 0.75, 0.25 and 1.60 respectively. Proportion of his sales for these pastries is 10%, 5%, 20%, 50% and 15% respectively. What are his expected daily profits from sale of pastries?

9. A man wishes to open a door of his house in dark. He has a bunch of 7 keys. He tries keys one by one discarding the unsuccessful keys. Find the expected number of keys he tries before the door is opened?

10. The probability distribution of weekly sales of DVD in a shop is,

Demand (X)	10	11	12	13	14	15
Probability $P(X)$	0.05	0.10	0.25	0.40	0.15	0.05

The shop earns a profit of ₹ 700 per set. If it is not sold, his loss in terms of inventory is ₹ 300 per set. How many sets should be stocked so that he maximizes profit?

11. Consider the following game. A person flips a coin repeatedly until a head comes up. This person receives a payment of 2^n Rupees if the first head comes up at nth flip.

    (a) Let X be a random variable equal to the amount of money the person wins. Show that the expected value of X does not exist (i.e. it is infinite). Show that the person should be willing to wager any amount of money to play this game. (This is known as the St. Petersburg Paradox. Why is it called paradox?).

    (b) Suppose that the person receives 2^n if first head comes up before the 8th flip. What is the expected value of the amount of money the person wins? How much money should a person be willing to pay to play this game?

12. Suppose that A and B are the two events with probabilities, $P(A) = \dfrac{3}{4}$ and $P(B) = \dfrac{1}{3}$

    (a) What is the largest value $P(A \cap B)$ can be? What is the smallest it can be?

    (b) What is the largest value $P(A \cup B)$ can be? What is the smallest it can be?

13. There are three cards in a box. Both sides of one card are black. Both sides of another card are red. The third card has one side black and one side red. We pick up one card at random and observe only one side,

    (a) If the side is black, what is the probability that the opposite side is also black?

    (b) What is the probability that the opposite side is the same colour as the one we observe?

14. What is the expected sum of the numbers that appear on two dice, each biased so that a 3 comes up twice as often as each other number?

15. Suppose that we roll a dice until six comes up or we have rolled it ten times. What is the expected number of times we roll the dice?

16. Let X is the number appearing on the first dice when two dice are rolled and let Y be the sum of the numbers appearing on the two dice. Show that $E(X\,Y) \neq E(X)\,E(Y)$ also comment.

17. What is the variance of the number of heads that come up when a fair coin is flipped 10 times?

18. Let $S = \{a, b, c\}$ be sample space associated with a certain experiment. If $P(a) = k$, $P(b) = 2k^2$ and $P(c) = k^2 + k$.

    (a) Find k

    (b) Are $A = \{a, b\}$ and $B = \{b, c\}$ independent events ?

19. A continuous random variable X has PDF given by,

$$f(x) = \begin{cases} Ax^2 & for \quad 0 \le x \le 1 \\ 0 & Otherwise \end{cases}$$

(a) Find A

(b) Find $P\left(\dfrac{1}{5} \le x \le \dfrac{1}{2}\right)$

(c) Find $P(x \le 0.3)$

(d) Find $P\left(x > \dfrac{3}{4} \Big| x > \dfrac{1}{2}\right)$

20. A continuous random variable X has PDF given by,

$$f(x) = \begin{cases} 12x^3 - 21x^2 + 10x & for \quad 0 \le x \le 1 \\ 0 & Otherwise \end{cases}$$

(a) Find $P\left(x \le \dfrac{1}{2}\right)$

(b) Find $P\left(x > \dfrac{1}{2}\right)$

(c) Find k such that $P(x \le k) = \dfrac{1}{2}$

21. A continuous random variable X has PDF given by,

$$f(x) = \begin{cases} \dfrac{1}{4} & for \quad -2 \le x \le 2 \\ 0 & Otherwise \end{cases}$$

(a) Find $P(x \le 1)$

(b) Find $P(|x| > 1)$

(c) Find $P(2x + 3 > 5)$

22. A continuous random variable with range (–3, 3) has a PDF given as,

$$f(x) = \begin{cases} \dfrac{1}{16}(3+x)^2 & for \quad -3 \le x < -1 \\ \dfrac{1}{16}(6 - 2x^2) & for \quad -1 \le x < 1 \\ \dfrac{1}{16}(3-x)^2 & for \quad 1 \le x \le 3 \end{cases}$$

(a) Verify that the distribution satisfies the Axiom I and Axiom II of probability.

(b) Find mean

(c) Find variance

23. A continuous random variable X represents time in months an electronic part functions before failure. X has PDF given by,

$$f(x) = \begin{cases} Cxe^{\frac{-x}{2}} & for \quad x \geq 0 \\ 0 & Otherwise \end{cases}$$

(a) Find C

(b) Find cumulative density $F(a = 5) = P(x \leq 5)$

(c) What is the probability that the part will function at least 5 months?

24. Consider a function,

$$f(x) = \begin{cases} C(2x - x^3) & for \quad 0 \leq x \leq \dfrac{5}{2} \\ 0 & Otherwise \end{cases}$$

Could $f(x)$ be a PDF?

25. A continuous random variable X has PDF given by,

$$f(x) = \begin{cases} a + bx^2 & for \quad 0 \leq x \leq 1 \\ 0 & Otherwise \end{cases}$$

If expected value $E(X) = \dfrac{3}{5}$, find a and b.

Answers: Self Assessment

1. (a) In order that the $P(X = x_i)$ be a PMF, it should satisfy axioms of probability. To satisfy Axiom II,

$$\sum_{All} P(x_i) = 1 \Rightarrow 49k = 1 \Rightarrow k = \frac{1}{49}$$

(b) $P(X \geq 2) = 1 - P(0) - P(1) = 1 - 4k = \dfrac{45}{49}$

(c) $P(0 < X < 5) = P(1) + P(2) + P(3) + P(4) = 24k = \dfrac{24}{49}$

(d) $P(X \leq 3) = \dfrac{16}{49} < 0.5$ and $P(X \leq 4) = \dfrac{25}{49} > 0.5$

Therefore, the minimum value of C for which $P(X \leq C) > 0.5$ is C = 4

(e) Probability mass function is:

$X = x_i$	0	1	2	3	4	5	6
$P(X = x_i)$	$\dfrac{1}{49}$	$\dfrac{3}{49}$	$\dfrac{5}{49}$	$\dfrac{7}{49}$	$\dfrac{9}{49}$	$\dfrac{11}{49}$	$\dfrac{13}{49}$

Distribution Function or CDF is given as $F(a) = P(X \le a)$

$X = x_i$	0	1	2	3	4	5	6
$F(a)$	$\dfrac{1}{49}$	$\dfrac{4}{49}$	$\dfrac{9}{49}$	$\dfrac{16}{49}$	$\dfrac{25}{49}$	$\dfrac{36}{49}$	1

2. Now, for $p(x)$ to be the PMF, it must satisfy the axioms of probability. Using Axiom II,

   (a) $\sum_{All} P(x_i) = 1 \Rightarrow (k + 2k + 3k + ... + 10k) = 1 \Rightarrow k(1 + 2 + 3 + ... + 10) = 1$

   $\Rightarrow 55k = 1 \Rightarrow k = \dfrac{1}{55}$

   (b) $\sum_{All} P(x_i) = 1 \Rightarrow k + 2k + 2k + \dfrac{4}{3}k = 1 \Rightarrow \dfrac{19}{3}k = 1 \Rightarrow k = \dfrac{3}{19}$

   (c) $\sum_{All} P(x_i) = 1 \Rightarrow k + 6k + 15k + 28k = 1 \Rightarrow 50k = 1 \Rightarrow k = \dfrac{1}{50}$

3. Let random variable X be the amount won by player A in each game.

   (a) There are 2 possibilities for holding and two possibilities for guessing fingers for each player. Thus, there are $2^4 = 16$ possible equally likely outcomes, with probability of each equal to $\dfrac{1}{16}$. These outcomes and corresponding value of X is shown below:

A	Hold	1	1	1	1	2	1	1	1	2	2	2	1	2	2	2	2
	Guess	1	1	1	2	1	1	2	2	1	1	2	2	1	2	2	2
B	Hold	1	1	2	1	1	2	1	2	1	2	1	2	2	1	2	2
	Guess	1	2	1	1	1	2	2	1	2	1	1	2	2	2	1	2
$X = x_i$		0	2	–3	–2	3	0	0	0	0	0	0	3	–4	–3	4	0

   Thus, the values of X and their associated probabilities are:

$X = x_i$	–4	–3	–2	0	2	3	4	$\sum P(x_i)$
PMF $P(X = x_i)$	$\dfrac{1}{16}$	$\dfrac{2}{16}$	$\dfrac{1}{16}$	$\dfrac{8}{16}$	$\dfrac{1}{16}$	$\dfrac{2}{16}$	$\dfrac{1}{16}$	1

   (b) There are 2 possibilities for holding and guessing fingers for each player. Thus, there are $2^2 = 4$ possible equally likely outcomes, with probability of each equal to $\dfrac{1}{4}$. These outcomes and corresponding value of X is shown below:

A	Hold	1	1	2	2
	Guess	1	1	2	2
B	Hold	1	2	1	2
	Guess	1	2	1	2
$X = x_i$		0	0	0	0

   Thus, X can take only one value of 0 with associated probability of 1.

4.   (a)   To be a PMF it must satisfy the Axiom II of probability, i.e. $\sum_{i=1}^{\infty} P(X=i)=1$

Now, for $P(X=i)=C\times 2^{-i}$

$$\sum_{i=1}^{\infty} P(X=i)=\sum_{i=1}^{\infty} C\times 2^{-i}=\frac{C}{2}[\sum_{i=0}^{\infty} 2^{-i}]=\frac{C}{2}[\sum_{i=0}^{\infty}\left(\frac{1}{2}\right)^{i}]=\frac{C}{2}\times\frac{1}{\left(1-\frac{1}{2}\right)}=C=1$$

Therefore, $C=1$

Note: We have used the sum of in infinite series $1+r+r^2+r^3+..........=\dfrac{1}{1-r}$

(b)   To be a PMF it must satisfy the Axiom II of probability, i.e. $\sum_{i=1}^{\infty} P(X=i)=1$

Now, for $P(X=i)=C\times\dfrac{2^{-i}}{i}$.

$$\sum_{i=1}^{\infty} P(X=i)=\sum_{i=1}^{\infty} C\times\frac{2^{-i}}{i}=C\left[\frac{\left(\frac{1}{2}\right)}{1}+\frac{\left(\frac{1}{2}\right)^2}{2}+\frac{\left(\frac{1}{2}\right)^3}{3}+.......\right]=C[-\log_e\left(1-\frac{1}{2}\right)]=C\log_e(2)=1$$

Therefore, $C=\dfrac{1}{\log_e 2}$

Note: We have used the sum of infinite series $-\dfrac{x^1}{1}-\dfrac{x^2}{2}-\dfrac{x^3}{3}-.......=\log_e(1-x)$

(c)   To be a PMF it must satisfy the Axiom II of probability, i.e. $\sum_{i=1}^{\infty} P(X=i)=1$.

Now, for $P(X=i)=C\times\dfrac{2^i}{i!}$

$$\sum_{i=1}^{\infty} P(X=i)=\sum_{i=1}^{\infty} C\times\frac{2^i}{i!}=C\left[\frac{2^1}{1!}+\frac{2^2}{2!}+\frac{2^3}{3!}+.......\right]=C[e^2-1]=1$$

Therefore, $C=\dfrac{1}{e^2-1}$

Note: We have used the sum of infinite series $\dfrac{x^0}{0!}+\dfrac{x^1}{1!}+\dfrac{x^2}{2!}+\dfrac{x^3}{3!}-.......=e^x$

(d)   To be a PMF it must satisfy the Axiom II of probability, i.e. $\sum_{i=1}^{\infty} P(X=i)=1$

Now, for $P(X=i)=C\times i^{-2}$

$$\sum_{i=1}^{\infty} P(X=i) = \sum_{i=1}^{\infty} C \times i^{-2} = C\sum_{i=1}^{\infty}\left[\frac{1}{i}\right]^2 = C\left[\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+......\right] = C[2] = 1$$

Therefore $\quad C = \dfrac{1}{2}$

Note: We have used the sum of infinite series $\left[\dfrac{1}{1^2}+\dfrac{1}{2^2}+\dfrac{1}{3^2}+......\right] = 2$

5   Let W, B and O denote the white, black and orange balls.

(a) The possible values of X and the probabilities associated with each value is given below as PMF (probability distribution).

Selection of Balls	W&W	W&O	W&B	O&O	O&B	B&B
X	4	2	1	0	–1	–2
Probability $P(X=x_i)$	$\dfrac{\binom{8}{2}}{\binom{14}{2}}=\dfrac{28}{91}$	$\dfrac{\binom{8}{1}\binom{2}{2}}{\binom{14}{2}}=\dfrac{16}{91}$	$\dfrac{\binom{8}{1}\binom{4}{1}}{\binom{14}{2}}=\dfrac{32}{91}$	$\dfrac{\binom{2}{2}}{\binom{14}{2}}=\dfrac{1}{91}$	$\dfrac{\binom{2}{2}\binom{4}{1}}{\binom{14}{2}}=\dfrac{8}{91}$	$\dfrac{\binom{4}{2}}{\binom{14}{2}}=\dfrac{6}{91}$

Note that, $\sum P(x_i) = 1$

(b) Expected value of winning per game is,

$$E(X) = \sum x_i P(x_i) = 4\times\frac{28}{91}+2\times\frac{16}{91}+1\times\frac{32}{91}+0\times\frac{1}{91}-1\times\frac{8}{91}-2\times\frac{6}{91}$$

$$=\frac{156}{91}=1.741$$

Now for playing 100 times, we need to pay table money as ₹ 100 × 2 = 200 and we are expected to win ₹ 1.741 × 100 = 174. Thus, we are expected to lose ₹ 200 – 174 = 26.

(c) No. The game is biased against us.

6. Since there are 5 women among total 10 people, possible values of *X* could be 1, 2, 3, 4, 5 and 6. Other values viz. 7, 8, 9 and 10 are not feasible. Further, total possible rankings are 10! ways. Now we can calculate the associated probabilities as follows:

X = 1 CASE:

First woman can be ranked in 5 ways and the remaining 9 people can be ranked in 9! ways. Thus, probability,

$$P(X=1) = \frac{5\times 9!}{10!} = \frac{1}{2}$$

Or,

There are 5 women out of 10. Hence, probability of selecting first woman is $\dfrac{5}{10}=\dfrac{1}{2}$.

X =2 CASE:

Now the top woman has second rank. Means a man has first rank. First man can be in 5 ways. Then second woman can be in 5 ways. Remaining 8 people can have any order. Thus, probability,

$$P(X = 2) = \frac{5 \times 5 \times 8!}{10!} = \frac{5}{18}$$

Or,

Probability of first man is $\frac{1}{2}$. . Having selected one man, there are 9 people available for second rank of which 5 are women. Hence, probability of woman for second rank is $\frac{5}{9}$. By product rule, probability of top woman having second rank, i.e. first rank of man and second woman is $\frac{1}{2} \times \frac{5}{9} = \frac{5}{18}$.

Continuing in this fashion, we get probability distribution as,

$X = x_i$	1	2	3	4	5	6	7	8	9	10
$P(X = x_i)$	$\frac{1}{2}$	$\frac{5}{18}$	$\frac{5}{36}$	$\frac{5}{84}$	$\frac{5}{252}$	$\frac{1}{252}$	0	0	0	0

Note that, $\sum P(x_i) = 1$

7   For given probabilities to be the PMF these must satisfy the axioms of probability.

Axiom I: $\sum\limits_{all} p(x_i) = 1$

Now, $\sum\limits_{all} p(x_i) = \frac{1+3x}{4} + \frac{1-x}{4} + \frac{1+2x}{4} + \frac{1-4x}{4} = 1$

Thus, the Axiom I is satisfied for all real values of x.

Axiom II: $0 \le p(x_i) \le 1$

Thus, $0 \le \frac{1+3x}{4} \le 1 \Rightarrow 0 \le 1 + 3x \le 4 \Rightarrow -\frac{1}{3} \le x \le 1$ ...(1)

Also, $0 \le \frac{1-x}{4} \le 1 \Rightarrow 0 \le 1 - x \le 4 \Rightarrow -3 \le x \le 1$ ...(2)

Also, $0 \le \frac{1+2x}{4} \le 1 \Rightarrow 0 \le 1 + 2x \le 4 \Rightarrow -\frac{1}{2} \le x \le \frac{3}{2}$ ...(3)

And, $0 \le \frac{1-4x}{4} \le 1 \Rightarrow 0 \le 1 - 4x \le 4 \Rightarrow -\frac{3}{4} \le x \le \frac{1}{4}$ ...(4)

The values of x that satisfy all the conditions (1), (2), (3) and (4) are,

$-\dfrac{1}{3} \le x \le \dfrac{1}{4}$. These values satisfy Axiom II. Hence, $-\dfrac{1}{3} \le X \le \dfrac{1}{4}$ is the required condition on

X so that given probabilities represent PMF of X.

8.   (a)   PDF must satisfy probability Axiom I. Hence, $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$. Hence,

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{2} f(x)dx + \int_{2}^{\infty} f(x)dx = 0 + \int_{0}^{2} c(4x - 2x^2)dx + 0$$

$$c\left[4\frac{x^2}{2} - 2\frac{x^3}{3}\right]_0^2 = c\left[8 - \frac{16}{3}\right] = 1$$

$$c = \frac{3}{8}$$

(b) $P(X > 1) = \displaystyle\int_{1}^{\infty} f(x)dx = \int_{1}^{2} \frac{3}{8}(4x - 2x^2)dx = \frac{1}{2}$

9.   We use definition of CDF as, $F(a) = P(x \le a)$

Thus, $P(a \le X \le b) = P(x \le b) - P(x \le a) = F(b) - F(a)$

(a)   Now, $P\left(\dfrac{1}{2} \le X \le \dfrac{3}{2}\right) = F\left(\dfrac{3}{2}\right) - F\left(\dfrac{1}{2}\right) = \dfrac{3}{4} - \dfrac{1}{4} = \dfrac{1}{2}$

(b)   Similarly, $P(1 \le X \le 2) = F(2) - F(1) = 1 - \dfrac{1}{2} = \dfrac{1}{2}$

10.   First we need to find the value of k. For this we use Axiom II i.e. $\displaystyle\int_{\infty}^{\infty} f(x)dx = 1$. Thus,

$$\int_{\infty}^{\infty} f(x)dx = \int_{0}^{1} kx(1 - x)dx = k\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \frac{k}{6} = 1$$

Therefore, $k = 6$

Now, the probability that the project will be completed in less than four months, i.e. $P(x \le 4)$ is,

$$P(x \le 4) = \int_{x=0}^{\frac{1}{3}} 6x(1 - x)dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^{\frac{1}{3}} = \frac{7}{27} = 0.259$$

11. The required probability distribution is,

$X = x_i$	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_i)$	$\dfrac{1}{36}$	$\dfrac{2}{36}$	$\dfrac{3}{36}$	$\dfrac{4}{36}$	$\dfrac{5}{36}$	$\dfrac{6}{36}$	$\dfrac{5}{36}$	$\dfrac{4}{36}$	$\dfrac{3}{36}$	$\dfrac{2}{36}$	$\dfrac{1}{36}$
$X_i \times p(x_i)$	$\dfrac{2}{36}$	$\dfrac{6}{36}$	$\dfrac{12}{36}$	$\dfrac{20}{36}$	$\dfrac{30}{36}$	$\dfrac{42}{36}$	$\dfrac{40}{36}$	$\dfrac{36}{36}$	$\dfrac{30}{36}$	$\dfrac{22}{36}$	$\dfrac{12}{36}$

Now the mean of the random variable is,

$$E(X) = \sum_{i=1}^{11} x_i \times P(x_i) \frac{2+6+12+20+30+42+56+72+90+110+132}{36} = \frac{572}{36} = 7$$

12 Given, $M(t) = \dfrac{3}{3-t}$

Therefore, $M'(t) = \dfrac{3}{(3-t)^2}$ and, $M''(t) = \dfrac{6}{(3-t)^3}$

We know that $E(X) = M'(0) = \dfrac{1}{3}$ and, $E(X^2) = M''(0) = \dfrac{2}{9}$

Note: $E(X) = M(0) = 1$

Now, $var(X) = E(X^2) - [E(X)]^2 = \dfrac{2}{9} - \dfrac{1}{9} = \dfrac{1}{9}$

$S.D. = \sqrt{var(X)} = \dfrac{1}{3}$

Note: This problem can be solved using exponential distribution, which is discussed later.

13. We will solve the problem by using fundamental definition. However, this can be solved very simply by using geometric distribution because $P(X = i) = 2^{-i}$ is a geometric random variable with parameter $p = \dfrac{1}{2}$. However, we will show this in next unit where geometric distribution and its properties are explained.

(a) Given $P(X = i) = 2^{-i}$ for *i* = 1, 2, 3, ……

Thus, we can see that $P(X = i) = 2^{-i} = \dfrac{1}{2^i} \geq 0$ for any value of *i*   ...(1)

Also, $\displaystyle\sum_{i=1}^{\infty} P(X = i) = \sum_{i=1}^{\infty} \frac{1}{2^i} = \frac{1}{2} \sum_{i=0}^{\infty} \frac{1}{2^i} = \frac{1}{2} \times \frac{1}{1-\dfrac{1}{2}} = 1$ ...(2)

Note: We have used sum of infinite G.P. $1 + r + r^2 + r^3 + \dots\dots = \dfrac{1}{1-r}$ if $r \leq 1$

From (2) we can see that Axiom II is satisfied by $P(X = i) = 2^{-i}$

Also from (1) and (2) we see that Axiom I is satisfied.

Thus, $P(X = i) = 2^{-i}$ is indeed a PMF

(b)   Now, MGF $= M(t) = \sum_{i=1}^{\infty} e^{ti} \left(\dfrac{1}{2}\right)^i = \sum_{i=1}^{\infty} \left(\dfrac{e^t}{2}\right)^i = \dfrac{\dfrac{e^t}{2}}{1 - \dfrac{e^t}{2}} = \dfrac{e^t}{2 - e^t}$

Differentiating w.r.t. t,

$M'(t) = \dfrac{e^t(2 - e^t) - e^t(-e^t)}{(2 - e^t)^2} = \dfrac{2e^t}{(2 - e^t)^2}$

And, $M''(t) = \dfrac{2[e^t(2 - e^t)^2 - e^t(2)(2 - e^t)(-e^t)]}{(2 - e^t)^4} = \dfrac{2e^t[(2 - e^t) + 2e^t]}{(2 - e^t)^3} = \dfrac{2e^t(2 + e^t)}{(2 - e^t)^3}$

Now, $E(X) = M'(0) = \dfrac{2}{(2 - e^0)^2} = 2$

And, $E(X^2) = M''(0) = \dfrac{2e^0(2 + e^0)}{(2 - e^0)^3} = 6$

Thus,

Mean $= E(X) = 2$

Variance $= Var(X) = E(X^2) - [E(X)]^2 = 6 - 4 = 2$

14.   Note that this is a Ramp distribution commonly used in electronics.

(a) $P(1 \le x \le 1.5) = \int_{x=1}^{1.5} f(x)dx = \int_{x=1}^{1.5} \dfrac{x}{2} dx = \dfrac{1}{2}\left[\dfrac{x^2}{2}\right]_1^{1.5} = \dfrac{5}{16}$

(b)   Mean $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{x=0}^{2} \dfrac{x^2}{2} dx = \dfrac{1}{2}\left[\dfrac{x^3}{3}\right]_0^2 = \dfrac{4}{3}$

(c)   Now, $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_{x=0}^{2} \dfrac{x^3}{2} dx = \dfrac{1}{2}\left[\dfrac{x^4}{4}\right]_0^2 = 2$

Now, $Var(X) = E(X^2) - [E(X)]^2 = 2 - \dfrac{16}{9} = \dfrac{2}{9}$

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

For $-\infty < x < 0$ $F(a) = \int_{-\infty}^{a} 0 dx = 0$

For $0 \le x \le 2$ $\qquad$ $F(a) = \int_0^a \frac{x}{2} dx = \frac{a^2}{4}$

For $2 \le x < \infty$ $\qquad$ $F(a) = 1$

15. (a) The PDF must satisfy the axioms of probability. By Axiom II

$$\int_{y=0}^4 f(y)dy = \int_{y=0}^4 \frac{k}{\sqrt{y}} dy = k \left[2\sqrt{y} \right]_0^4 = 1 \qquad \text{or,}$$

$k = 4$

(b) Distribution function or CDF is given as, $F(b) = \int_{-\infty}^b f(y)dy$. Hence, the CDF for the given PDF is,

$$F(b) = \int_0^b \frac{1}{4\sqrt{y}} dy = \frac{1}{4} \left[2\sqrt{y} \right]_0^b = \frac{1}{2}\sqrt{b}$$

Thus, the CDF is

$$F(b) = \begin{cases} 0 & for \quad -\infty < b \le 0 \\ \dfrac{1}{2}\sqrt{b} & for \quad o < b < 4 \\ 1 & for \quad 4 \le b < \infty \end{cases}$$

(c) $P(1 < y < 2) = \int_1^2 f(y)dy = F(2) - f(1) = \frac{1}{2}(\sqrt{2} - 1) = 0.207$

16. Now first we find the first and second moments about origin, $E(X)$ and $E(X^2)$.

$$E(X) = \int_{-\infty}^\infty xf(x)dx = \int_{-\infty}^0 (0)dx + \int_0^1 x^2 dx + \int_1^2 x(2-x)dx + \int_2^\infty (0)dx = \left[\frac{x^3}{3} \right]_0^1 + \left[x^2 + \frac{x^3}{3} \right]_1^2 = 1$$

$$E(X^2) = \int_{-\infty}^\infty x^2 f(x)dx = \int_{-\infty}^0 (0)dx + \int_0^1 x^3 dx + \int_1^2 x^2(2-x)dx + \int_2^\infty (0)dx = \left[\frac{x^4}{4} \right]_0^1 + \left[\frac{2x^3}{3} + \frac{x^4}{4} \right]_1^2 = \frac{7}{6}$$

(a) $mean = E(X) = 1$

(b) $Variance = E(X^2) - \left[E(X) \right]^2 = \frac{1}{6}$

17. (a) The PDF must satisfy the axioms of probability. By Axiom II $\int_{-\infty}^\infty f(x)dx = 1$. Hence,

$$\int_{x=0}^\infty f(x)dx = \int_{x=0}^\infty kx^2 e^{-x} dx = k \left[(x^2)(\frac{e^{-x}}{-1}) - (2x)(e^{-x}) + (2)(\frac{e^{-x}}{-1}) \right]_0^\infty = 1 \qquad \text{or,}$$

$k[(0-0+0)-(0-0-2)]=1 \qquad \text{or,}$

$k = \dfrac{1}{2}$

(b) $mean = E(X) = \int\limits_{-\infty}^{\infty} xf(x)dx = \int\limits_{0}^{\infty} x(kx^2 e^{-x})dx$

$= \dfrac{1}{2}\left[(x^3)(\dfrac{e^{-x}}{-1}) - (3x^2)(e^{-x}) + (6x)(\dfrac{e^{-x}}{-1}) - (6)(e^{-x})\right]_{0}^{\infty} = \dfrac{1}{2}[6] = 3$

(c)   Now, $E(X^2) = \int\limits_{-\infty}^{\infty} x^2 f(x)dx = \int\limits_{0}^{\infty} x^2 (kx^2 e^{-x})dx$

$= \dfrac{1}{2}\left[(x^4)(\dfrac{e^{-x}}{-1}) - (4x^3)(e^{-x}) + (12x^2)(\dfrac{e^{-x}}{-1}) - (24x)(e^{-x}) + (24)(\dfrac{e^{-x}}{-1})\right]_{0}^{\infty}$

$= \dfrac{1}{2}[24] = 12$

$Variance = E(X^2) - [E(X)]^2 = 12 - (3)^2 = 3$

13.15 Further Readings

Books   Béla Bollobás, Modern graph theory, Springer

Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph theory, Combinatorics, and Algorithms*, Birkhäuser

Online links   http://en.wikipedia.org/wiki/Random_variable

http://en.wikipedia.org/wiki/Probability_distribution

http://cnx.org/content/m16825/latest/

Unit 14: Probability Distributions

Objectives

After studying this unit, you will be able to:

- Understand binomial distribution

- Know poisson distribution

- Explain geometric distribution

- Discuss expectation

- Describe variance of a random variable

Introduction

For a discrete random variable, probability mass function (PMF) can be calculated using underlying probability structure on the sample space of the random experiment. However, in many practical situations, the random variable of interest follows a specific pattern, which can be described by a standard probability distribution. In these cases, PMF can be expressed in algebraic form and various characteristics of distribution like mean, variance, moments, etc. can be calculated using known closed formulae. These standard distributions are also called 'probability models'. We will study a few standard discrete probability distributions. Usually when discrete random variables can take only integer values, i rather than xi is used to represent these values.

14.1 Binomial Distribution

Binomial random variable is very useful in practice, which counts the number of successes when '*n*' Bernoulli trials are performed.

Suppose *n* independent Bernoulli trials are performed. Each trial results in a success with probability *p* (or a failure with probability q =1 - *p*). If *X* represents the number of successes that occur in *n* trials, then *X* is said to be a Binomial random variable and the probability distribution is known as Binomial distribution with parameters (*n, p*). It is denoted as $X \rightarrow B(n, p)$.

Thus, Bernoulli distribution is just a binomial distribution with *n* = 1, i.e., parameters (1, *p*).

Did u know? Is Bernoulli distribution is just a binomial distribution with *n* = 1, i.e. parameters (1, *p*)?

14.1.1 Probability Mass Function (PMF)

PMF of a binomial random variable having parameters (*n, p*) is given by,

$$p(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad\qquad i = 0, 1.........n$$

> *Notes*   If there are i successes out of n trials, for any one of the combinations of i successful trials, the probability is,
>
> $$p \times p \times ... \times p \ i \ \text{times} \times (1-p) \times (1-p) \times ... \times (1-p) \ (i-1) \ \text{times} = p^i (1-p)^{n-i}$$

However, these successful *i* trials could be any of the $\binom{n}{i}$ combinations. Hence,

$$p(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad\qquad i = 0, 1.............. n$$

> *Notes*   Because $\binom{n}{i} \geq 0$, P ≥ 0 and (1 - p) ≥ 0 we can say that, $P(X = i) \geq 0$

Also,

$$\sum_{i=1}^{n} p(X = i) \qquad\qquad = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

$$= [p + (1-p)]^n \qquad\qquad \text{Using binomial theorem.}$$

$$\therefore \quad \sum_{i=1}^{\infty} p(X = i) = 1 \qquad\qquad \text{Axiom II is satisfied.}$$

Hence, $0 \leq P(X = i) \leq 1$ \qquad\qquad Axiom I is satisfied.

14.1.2 Cumulative Distribution Function (CDF)

CDF of a binomial random variable having parameters (n, p) is given by,

$$F(a) = P(X \le a) = \sum_{i=1}^{a} \binom{n}{i} p^i (1-p)^{n-i}$$

14.1.3 Expectation or Mean (μ)

Expectation of a Binomial random variable having parameters (n, p) could be calculated as follows. For convenience, we first calculate a more generalized moment, $E(X^k)$.

$$E(X^k) = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$

$$= \sum_{i=1}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} \quad \because \text{At } i = 0, \text{ the term is zero and hence can be}$$

dropped.

$$= \sum_{i=1}^{n} i \binom{n}{i} i^{k-1} p^i (1-p)^{n-i}$$

$$= \sum_{i=1}^{n} n \binom{n-1}{i-1} \cdot i^{k-1} \cdot p \cdot p^{i-1} \cdot (1-p)^{n-i} \qquad \text{Using } i\binom{n}{i} = n\binom{n-1}{i-1}$$

$$= np \sum_{i=1}^{n} \binom{n-1}{i-1} \cdot i^{k-1} \cdot p^{(i-1)} \cdot (1-p)^{n-i}$$

Put $j = i - 1$ or, $i = j + 1$ Thus, we get,

$$E(X^k) = np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j}$$

Thus,

$$E(X^k) = npE\left[(Y+1)^{k-1}\right]$$

Where Y is a Binomial random variable with parameters $(n-1, p)$. This is by comparing with the PMF of Binomial random variable with parameters $(n-1, p)$.

Now, putting $k = 1$

$$E(X) = npE\left[(Y+1)^0\right] = np$$

Thus, Expected value of X or mean is,

$$\mu = E(X) = np$$

Task Calculate the expectation of a Binomial random variable having parameters (0,1)

14.1.4 Variance: Var(X)

Variance of a Binomial random variable having parameters (n, p) could be calculated as follows:

Now in previous result of $E\left[X^k\right] = npE\left[(Y+1)^{k-1}\right]$ we put $k = 2$

$$E\left[X^2\right] \;=\; npE\left[(Y+1)^1\right] = npE(Y+1) = np[E(Y)+1]$$

$$=\; np[(n-1)p+1]$$

Now,

$$var\,(X) \;=\; E(X^2) - [E(X)]^2$$

$$=\; np[(n-1)p+1] - (np)^2$$

$$=\; np[np - p + 1 - np]$$

$$=\; np(1-p) = npq \qquad\qquad \text{Where } q = (1-p)$$

14.1.5 Moment Generating Function (MGF)

Moment Generating Function (MGF) of Binomial random variable is,

$$M(t) \;=\; E(e^{tx}) = \sum_{i=0}^{n} e^{ti} \binom{n}{i} p^i (1-p)^{n-i}$$

$$=\; \sum_{i=0}^{n} \binom{n}{i} (pe^t)^i (1-p)^{n-i} = (pe^t + 1 - p)^n$$

This is by using Binomial theorem,

$$(a+b)^n \;=\; \sum_{i=0}^{n} \binom{n}{i} (a)^i (b)^{n-i} \;\text{ with } a = pe^t \text{ and } b = (1-p)$$

Notes

$M'(t) = n(pe^t + 1 - p)^{n-1}(pe^t)$

$M''(t) = n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 + n(pe^t + 1 - p)^{n-1}(pe^t)$

Now,

$$\mu \;=\; E[X] = M'(0) = np(p + 1 - p)^{n-1} = np$$

$$E[X^2] \;=\; n(n-1)p^2 + np = np[(n-1)p+1]$$

These are same as the earlier values.

14.1.6 Applications of Binomial Distribution

When to use binomial distribution is an important decision. Binomial distribution can be used when the following conditions are satisfied:

1. Trials are finite (and not very large), performed repeatedly for '*n*' times.

2. Each trial (random experiment) should be a Bernoulli trial, the one that results in either success or failure.

3. Probability of success in any trial is '*p*' and is constant for each trial.

4. All trials are independent.

These trials are usually the experiments of selection 'with replacement'. In cases where number of population is very large, drawing a small sample from it does not change probability of success significantly. Hence, we could consider the distribution as Bernoulli distribution.

Following are some of the real life examples of applications of Binomial distribution:

1. Number of defective items in a lot of *n* items produced by a machine.

2. Number of male births out of *n* births in a hospital.

3. Number of correct answers in a multiple-choice test.

4. Number of seeds germinated in a row of *n* planted seeds.

5. Number of re-captured fish in a sample of *n* fishes.

6. Number of missiles hitting the targets out of *n* fired.

Did u know? What is random experiment?

14.1.7 Calculation of PMF of Binomial Random Variable using Recurrence Relation

In order to calculate the probability of Binomial random variable, one must calculate $\binom{n}{i}$ which is quite tedious when *n* and *i* are large. However, if we write the PMF as a recurrence relation, the calculations become easy, particularly on computer. The same relationship can also be utilized for calculation CDF on computer.

Let *X* be a Binomial random variable with parameters (*n, p*). We observe that,

$$P(X=i) = \binom{n}{i}p^i(1-p)^{n-i} \qquad i=0,1.........n$$

and,

$$p(X=i+1) = \binom{n}{i+1}p^{i+1}(1-p)^{n-i-1} \qquad i=0,1..........,n-1$$

Substituting we get,

$$p(X = i+1) \quad = \quad \frac{n-i}{i+1} \times \frac{p}{1-p} \times p(X = i) \qquad\qquad i = 0,1\ldots\ldots,n-1$$

This is the required recurrence relation. We can calculate values of PMF using initial condition as,

$$p(X = 0) \quad = \quad \binom{n}{0} p^0 (1-p)^n = (1-p)^n$$

Example: If *X* is a binomial random variable with parameters (n, p) then as *k* increases from 0 to *n*, P{X = k} first increases and then decreases monotonically, reaching largest value when *k* is largest integer greater than or equal to $(n+1)p$.

Solution:

Since *X* is a Binomial random variable,

$$P(X = k) \quad = \quad \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

Hence,

$$\frac{P(X = k)}{P(X = k-1)} \quad = \quad \frac{\dfrac{n!}{(n-k)!k!} p^k (1-p)^{n-k}}{\dfrac{n!}{(n-k+1)!(k-1)!} p^{k-1} (1-p)^{n-k+1}}$$

$$= \quad \frac{(n-k+1)p}{k(1-p)}$$

Now, $P(X = k) \geq P(X = k-1)$, if and only if

$$(n - k + 1)\,p \geq k\,(1 - p)$$

or, $\quad (n + 1)\,p \geq k$

or, $\quad k \leq (n + 1)\,p$

Thus, X increases monotonically, till $k \leq (n + 1)\,p$

Similarly, $P\{X = k\} \leq P\{X = k-1\}$, if and only if $k \geq (n + 1)\,p$

Thus, *X* decreases monotonically, after $k \geq (n + 1)\,p$

Hence, the result is proved.

Example: A biased coin has probability of heads as $\dfrac{1}{3}$. This coin is tossed 6 times. Find the probability of getting:

(1)   4 heads   (2)   At least 2 heads   (3)   At the most 1 head

(4)   4 tails   (5)   At least 1 tail   (6)   All tails.

Solution:

Let X denote the number of heads obtained in 6 tosses. X is a Binomial random variable with parameters ($n = 6$, $p = \dfrac{1}{3}$). For Binomial random variable, probability of i successes is given by,

$$P(X = i) \quad = \quad \binom{n}{i} p^i (1-p)^{n-i} \qquad\qquad i = 0,1.........n$$

In this case,

$$= \quad \binom{6}{i}\left(\frac{1}{3}\right)^i \left(\frac{2}{3}\right)^{6-i} \qquad\qquad i = 0,1,2,3,4,5,6$$

1.   Probability of 4 heads.

$$P(X = 4) \quad = \quad \binom{6}{4}\left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^{6-4} = 0.0823$$

2.   Probability of at least 2 heads.

$$P(X \geq 2) \quad = \quad 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1)$$

$$= \quad 1 - \binom{6}{0}\left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^6 - \binom{6}{1}\left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^5 = 0.6488$$

3.   Probability of at most 1 head.

$$P(X \leq 1) \quad = \quad P(X = 0) + P(X = 1)$$

$$= \quad \binom{6}{0}\left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^6 + \binom{6}{1}\left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^5 = 0.3512$$

4.   Probability of 4 tails. This is same as probability of 2 heads.

$$P(X = 2) \quad = \quad \binom{6}{2}\left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{6-2} = 0.3292$$

5.   At least 1 tail. This means all events except all heads.

$$1 - P(X = 6) \quad = \quad 1 - P(X = 6) = 1 - \binom{6}{6}\left(\frac{1}{3}\right)^6 \left(\frac{2}{3}\right)^{6-6} = 0.9986$$

6.   All tails. This is same as no heads.

$$P(X = 0) \quad = \quad \binom{6}{0}\left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^6 = 0.08779$$

Example: Suppose a particular trait of person is classified on the basis of one pair of genes. 'd' represents dominant gene and 'r' represents recessive gene. 'dd' is pure dominant, 'rr' is pure recessive and 'rd' (or 'dr') is hybrid. Child receives one gene each from parents. If with respect to a particular trait, say colour of eyes as blue, 2 hybrid parents (both have blue eyes)

have four children. What is probability that 3 have outward appearance of dominant gene (blue eyes)?

Solution:

Probability of child inheriting gene 'd' or 'r' from his father is $\frac{1}{2}$ each. Similarly, probability of child inheriting gene 'd' or 'r' from his mother is $\frac{1}{2}$ each. Thus, by product rule, probabilities of child inheriting genes are,

$$P(dd) \;=\; \frac{1}{4}, P(rr) = \frac{1}{4}, P(rd) = \frac{1}{2}$$

Child has outward appearance of dominant gene if its gene pair is 'dd' or 'rd'. Thus, by sum rule the probability of child having outward appearance of dominant gene is, $\frac{3}{4}$. Parents have 4 children. Probability of each child having outward appearance of dominant gene is independent of each other. Thus, probability of each child having outward appearance of dominant gene follows Binomial distribution with parameters, $\left(n = 4, p = \frac{3}{4} \right)$

Hence, probability that 3 have outward appearance of dominant gene (blue eyes)

$$P(X = 3) \;=\; \binom{4}{3}\left(\frac{3}{4}\right)^{3}\left(\frac{1}{4}\right)^{1} = \frac{27}{64}$$

14.1.8 Fitting of Binomial Distribution

Usually, when we want to predict, interpolate or extrapolate the probabilities for a given probability distribution, it would be easier to get the results if the probability distribution is approximated to a standard probability distribution. In case the probability distribution (or a frequency distribution which is not necessarily a probability distribution) is concerned with a random variable X which takes finite integer values 0, 1, 2, …, n assumption of Binomial distribution may work as a model for the given data. This is known as fitting Binomial distribution to the given data. We first estimate the parameters of distribution (n, p) from the data and then compute probabilities and expected frequencies.

The parameter p is estimated by equating the mean of Binomial distribution $\mu = np$ with the data mean $\bar{x}$. Thus,

$$\hat{p} \;=\; \frac{\bar{x}}{n} \text{ and } \hat{q} = 1 - \hat{p} \text{ where } \hat{p} \text{ means } p \text{ estimate, and } \hat{q} \text{ means } q \text{ estimate.}$$

$$\bar{x} \;=\; \frac{\Sigma f_i x_i}{\Sigma f_i}$$

With the estimated parameters, we calculate all the probability values (frequencies) for the given data points. If the observed values are quite close to the estimates, the Binomial model under consideration is satisfactory. There is a more advanced test called as 'Chi-square test' to ascertain the correctness of the fit, which is beyond the scope of this book.

📑 *Example:* The following data gives number of seeds germinated in a row of 5 seeds each. Fit a Binomial distribution to the data and calculate expected frequency.

x_i	0	1	2	3	4	5
f_i	10	20	30	15	15	10

Solution:

Now,

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{235}{100} = 2.35 \qquad \text{Hence,}$$

$$\hat{p} = \frac{\bar{x}}{n} = \frac{2.35}{5} = 0.47 \qquad \hat{q} = 1 - \hat{p} = 0.53$$

$$N = \Sigma f_i = 100 \qquad \frac{\hat{p}}{\hat{q}} = 0.8868$$

Now, either by using PMF with $n = 5$ and $p = 0.47$ or by using recurrence relation, we can find probabilities and hence, expected frequencies. We demonstrate using recurrence relation.

$X = i$	0	1	2	3	4	5	Total
$\frac{(n-i)}{(i+1)}$	5	2	1	0.5	0.2	0	
$P(X = i)$	0.0418	0.1853	0.3287	0.2915	0.1293	0.0229	0.9995
$E_i = N \times P(X)$	4.18	18.53	32.87	29.15	12.93	2.29	99.95

We observe that fitting is reasonably good, except at both ends.

14.1.9 Mode of Binomial Distribution

Mode of a distribution is that value of the variable for which PMF attains its maximum. Thus, if M is a mode, then PMF increases till M and then decreases. Obviously, if PMF is increasing then the ratio,

$$\frac{P(X = i)}{P(X = i-1)} > 1$$

$$\frac{(n-i+1)p}{i(1-p)} > 1 \text{ or,}$$

$$\frac{(n+1)p - i}{i(1-p)} > 0 \text{ or,}$$

$$(n+1)p > i$$

CASE I: $(n+1)p$ is not an integer.

If $(n+1)p$ not an integer, then Mode M is an integer part of $(n+1)p$.

CASE II: $(n+1)\ p$ is an integer.

If $(n+1)\,p$ is an integer, then at Mode M we get $(n+1)p = i$. Thus, $P(X = i) = P(X = i-1)$. Hence, the mode is not unique. Its value is $(n+1)p$ and $(n+1)p - 1$. The distribution is called bimodal.

14.2 Poisson Distribution

When number of trials or sample space is infinitely large, and where each trial is Bernoulli trial, Poisson random variable is used. S.D. Poisson introduced it for application of probability theory to law suits. It has a tremendous range of applications in diverse areas because it may be used as an approximation for a Binomial random variable with parameters (n, p) when n is large and p is small enough to keep the product np moderate.

14.2.1 Probability Mass Function (PMF)

A random variable X, taking on one of the countable infinite values 0, 1, 2… is said to be Poisson random variable with parameters 'λ', if for some $\lambda > 0$, probability mass function, (PMF) is,

$$p(i) \quad = \quad P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!} \qquad\qquad i = 0, 1, 2,\ldots$$

Notes Because $e^{-\lambda} \geq 0, \lambda^i \geq 0$ and $i! \geq 0$ hence, $e^{-\lambda}\dfrac{\lambda^i}{i!} \geq 0$

Also, $\qquad\qquad \displaystyle\sum_{i=0}^{\infty} p(i) \quad = \quad e^{\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}\cdot e^{\lambda} \qquad\qquad$ Using series $\dfrac{x^0}{0!} + \dfrac{x^1}{1!} + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \ldots = e^x$

$$= \quad 1 \qquad\qquad\qquad\qquad \text{Axiom II is satisfied.}$$

Since Axiom II is proved and $e^{-\lambda}\dfrac{\lambda^i}{i!} \geq 0$ for all i, we get

$$0 \quad \leq \quad e^{-\lambda}\frac{\lambda^i}{i!} \leq 1 \qquad\qquad\qquad \text{Axiom I is satisfied.}$$

14.2.2 Cumulative Density Function (CDF)

Cumulative density function (CDF) of a Poisson random variable having parameter 'λ' ($\lambda > 0$), is given by,

$$F(a) \quad = \quad P\{X \leq a\} = e^{-\lambda} + \frac{\lambda}{1!}e^{-\lambda} + \frac{\lambda^2}{2!}e^{-\lambda} + \ldots + \frac{\lambda^a}{a!}e^{-\lambda}$$

$$= \quad e^{-\lambda}\left(\frac{\lambda^0}{0!} + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^a}{a!}\right)$$

14.2.3 Expectation or Mean (μ)

Expectation of a Poisson random variable having parameter λ is,

$$\mu \;=\; E[X]=\sum_{i=0}^{\infty} i e^{-\lambda}\frac{\lambda^{i}}{i!}=\sum_{i=1}^{\infty} i e^{-\lambda}\frac{\lambda^{i}}{i!} \text{ Because the first term is zero.}$$

$$=\; \lambda\sum_{i=1}^{\infty} e^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!} \qquad\qquad \text{Putting } j = i-1 \text{ we get,}$$

$$=\; \lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^{i}}{j!}=\lambda\cdot e^{-\lambda}\cdot e^{-\lambda}$$

$$=\; \lambda$$

14.2.4 Variance: Var(X)

Variance of a Poisson random variable having parameter λ is calculated as follows:

$$E[X^2] \;=\; \sum_{i=0}^{\infty} i^{2} e^{-\lambda}\frac{\lambda^{i}}{i!}=\sum_{i=1}^{\infty} i^{2} e^{-\lambda}\frac{\lambda^{i}}{i!} \text{ Because the first term is zero.}$$

$$=\; \lambda\sum_{i=1}^{\infty} i\frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!} \qquad\qquad \text{Putting } j = i-1 \text{ we get,}$$

$$=\; \lambda\sum_{i=0}^{\infty}(j+1)\frac{e^{-\lambda}\lambda^{j}}{j!}=\lambda\left[\sum_{i=0}^{\infty} j\frac{e^{-\lambda}\lambda^{j}}{j!}+\sum_{j=0}^{\infty}\frac{e^{-\lambda}\lambda^{j}}{j!}\right]$$

$$=\; \lambda(\lambda+1)$$

$$Var(X) \;=\; E\left[X^{2}\right]-\left(E[X]\right)^{2}=\lambda(\lambda+1)-\lambda^{2}$$

$$=\; \lambda$$

📋

Notes   Find the variance of a Poisson random variable having parameter 1.

14.2.5 Moment Generating Function (MGF)

Moment Generating Function (MGF) of Poisson random variable is,

$$M(t) \;=\; E\left[e^{tx}\right]=\sum_{i=0}^{\infty} e^{ti} e^{-\lambda}\frac{\lambda^{i}}{i!}$$

$$=\; e^{-\lambda}\sum_{i=0}^{\infty}\frac{(\lambda e^{t})^{i}}{i!}=e^{-\lambda}e^{\lambda e^{t}}$$

$$=\; e^{\lambda(e^{t}-1)}$$

Now,

$$M'(t) = e^{\lambda(e^t-1)} \cdot \lambda e^t$$

$$M''(t) = e^{\lambda(e^t-1)} \cdot \left(\lambda e^t\right)^2 + \lambda e^t e^{\lambda(e^t-1)}$$

Hence,

$$E[X] = M'(0) = \lambda$$

$$E[X^2] = M''(0) = \lambda^2 + \lambda$$

This is same as the earlier results.

14.2.6 Poisson Approximation of Binomial Distribution

Prove

Poisson distribution can be used as an approximation for a binomial random variable with parameter (n, p) when *n* is large and *p* is small enough so that *np* is of moderate size, In such cases. we take $\lambda = np$,

Proof

Suppose that *X* is a binomial random variable with parameters (*n, p*) and let $\lambda = np$

$$
\begin{aligned}
P\{X=i\} &= \frac{n!}{(n-i)!i!} \cdot p^i (1-p)^{n-i} \\[2mm]
&= \frac{n(n-1)(n-2)....(n-i+1)}{i!} \cdot \left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i} \\[2mm]
&= \frac{n(n-1)(n-2)....(n-i+1)}{n \times n \times n \times ... \times n} \begin{matrix} i & terms \\ i & terms \end{matrix} \frac{\lambda^i}{i!} \frac{\left(1-\dfrac{\lambda}{n}\right)^n}{\left(1-\dfrac{\lambda}{n}\right)^i} \\[2mm]
&= e^{-\lambda} \cdot \frac{\lambda^i}{i!}
\end{aligned}
$$

This is PMF of Poisson random variable.

Notes   Here we have used the following results:

1.   As $n \to \infty \dfrac{n(n-1)(n-2)....(n-i+1)}{n \times n \times ... \times n} \begin{matrix} i & terms \\ i & terms \end{matrix} = 1$

This is because as $n \to \infty$ the ratio is $\dfrac{\infty}{\infty}$. Then we use L'hospital's rule and differentiate numerator and denominator i times. Numerator and denominator both become i! and get cancelled. Thus, the value of limit is 1.

Contd...

---

2. $\lim_{n\to\infty}\left(1-\dfrac{\lambda}{n}\right)^{i} = 1^{i} = 1$

3. $\lim_{n\to\infty}\left(1-\dfrac{\lambda}{n}\right)^{n} = \lim_{n\to\infty}\left[\binom{n}{0}\left(\dfrac{-\lambda}{n}\right)^{0}1^{n} + \binom{n}{1}\left(\dfrac{-\lambda}{n}\right)^{1}1^{n-1} + \binom{n}{2}\left(\dfrac{-\lambda}{n}\right)^{2}1^{n-2} + ...\right]$

$= \lim_{n\to\infty}\left[\dfrac{1}{0!} - \dfrac{n}{n}\cdot\dfrac{\lambda^{1}}{1!} + \dfrac{n(n-1)}{n\times n}\cdot\dfrac{\lambda^{2}}{2!} - \dfrac{n(n-1)(n-2)}{n\times n\times n}\times\dfrac{\lambda^{3}}{3!} + \dfrac{n(n-1)(n-2)(n-3)}{n\times n\times n\times n}\times\dfrac{\lambda^{4}}{4!} + ...\right]$

$\left[\dfrac{1}{0!} - \dfrac{\lambda^{1}}{1!} + \dfrac{\lambda^{2}}{2!} - \dfrac{\lambda^{3}}{3!} + \dfrac{\lambda^{4}}{4!} + ...\right] = e^{-\lambda}$

Hence, in other words, if independent trials, each of which result in a success with probability *p*, are performed, further when *n* is large and *p* small enough to make *np* moderate, the number of success that occurs is approximately Poisson distribution with parameter λ = *np*.

14.2.7 Applications of Poisson Distribution

The Poisson random variable has a tremendous range of application in diverse areas. Some of the common applications where Poisson distribution is used are,

1. Number of accidents on the express way in one day.

2. Number of misprints on a page.

3. Number of vehicles arriving at a petrol pump in one hour.

4. Number of a particles discharged in one second by a radioactive material.

5. Number of earthquakes occurring in one year in a particular seismic zone.

6. Number of deaths of policy-holders in one year.

We can compute Poisson distribution function using recurrence relation as follows.

$$\dfrac{P(X=i+1)}{P(X=i)} \;=\; \dfrac{e^{-\lambda}\dfrac{\lambda^{i+1}}{(i+1)!}}{e^{-\lambda}\dfrac{\lambda^{i}}{i!}} = \dfrac{\lambda}{i+1}$$

$$P(X=i+1) \;=\; \dfrac{\lambda}{i+1}P(X=i)$$

Initial value is $P(X=0) \;=\; e^{-\lambda}$

Thus, $$P(X=1) \;=\; \dfrac{\lambda}{0+1}P(X=0) = \lambda e^{-\lambda}$$

$$P(X=2) \;=\; \dfrac{\lambda}{1+1}P(X=1) = \dfrac{\lambda^{2}}{2}e^{-\lambda}$$

And so on.

📑 *Example:* Number of errors on a single page has Poisson distribution with average number of errors of one per page. Calculate the probability that there is at least one error on a page.

Solution:

This is a case of Poisson distribution with $\lambda = 1$. Now,

$$P(X \geq 1) \quad = \quad 1 - P(X = 0) = 1 - e = 1 - e^{-1} = 0.632$$

📑 *Example:* Number of accidents on an express-way each day is a Poisson random variable with average of three accidents per day. What is the probability that no accident will occur today?

Solution:

This is a case of Poisson distribution with $\lambda = 3$. Now,

$$P(X = 0) \quad = \quad e^{-3} = 0.0498$$

14.3 Geometric Distribution

Suppose that independent trials are performed until a success occurs, and each trial has a probability of success as p ($0 < p < 1$). Now if a random variable X is number of trials required till success occurs, then X is called as Geometric random variable.

14.3.1 Probability Mass Function (PMF)

A random variable X, is said to be Geometric random variable with parameters 'p', if probability mass function, (PMF) is,

$$P(X = i) \quad = \quad (1 - p)^{i-1} p \qquad\qquad i = 1, 2.......$$

Proof:

Suppose that independent trials, each having a probability p, $0 < p < 1$ of being a success are performed until a success occurs. Show that probability of the number of trials required till the success occurs is given by,

$$P(X = i) \quad = \quad (1 - p)^{i-1} p \qquad\qquad i = 1, 2.......$$

Proof:

Now the probability of success in each trial is p. Hence, probability of failure in each trial is $(1 - p)$. If first time success occurs in i^{th}. trial, the first $(i - 1)$ trials are necessarily failures followed by a success. Probability of such a thing happening is,

$$\left[(1 - p)(1 - p)(1 - p).......(i - 1) \text{ times} \right] \times p$$

Or,

$$P(X = i) \quad = \quad (1 - p)^{i-1} p$$

1. $0 \le (1-p) \le 1$ and $0 \le p \le 1$

   Hence, $0 \le P(X = i) \le 1$ Axiom I is satisfied.

2. Also, $\sum_{n=1}^{\infty} p\{x = i\} = \sum_{i=1}^{\infty}(1-p)^{i-1}p = p\sum_{i=1}^{\infty}(1-p)^{i-1} = p\dfrac{1}{1-(1-p)} = 1$ Axiom II is satisfied.

   Using sum of infinite G.P. with first term 1 and common ratio $(1 - p)$ and noting the sum of G.P. as,

   $$a + ar + ar^2 + \dots + ar^n \quad = \quad \frac{a(1-r^{n+1})}{1-r} \text{ with } a = 1$$

   $$1 + r + r^2 + \dots + r^n \quad = \quad \frac{1-r^{n+1}}{1-r}$$

   Taking the limit $n \rightarrow \infty$ we get the result.

   $$\sum_{n=1}^{\infty} r^{n-1} \quad = \quad \frac{1}{1-r}$$

14.3.2 Cumulative Distribution Function (CDF)

Cumulative distribution function (CDF) of a Geometric random variable having parameters p, is given by,

$$F(a) \quad = \quad P\{X \le a\} = \sum_{i=1}^{a}(1-p)^{i-1}p = p \times \sum_{i=1}^{a}(1-p)^{i-1}$$

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

$$= \quad 1 - q^a \qquad \text{where } 1 - p = q$$

14.3.3 Expected Value or Mean (μ)

Expectation of a Geometric random variable having parameter p is,

$$\mu \quad = \quad E(X) = \sum_{i=1}^{\infty} i(1-p)^{i-1}p = \sum_{i=0}^{\infty} i(q)^{i-1}p$$

The first term of the summand is zero. Also $1 - p = q$. Therefore,

$$\mu \quad = \quad p\sum_{i=0}^{\infty}\frac{d}{dq}(q^i) = p \times \frac{d}{dq}\sum_{i=0}^{\infty} q^i$$

Using differentiation under summation sign we get,

$$\mu \;=\; p \times \frac{d}{dq}\left(\frac{1}{1-q}\right) = p\,\frac{1}{(1-q)^2} = p\,\frac{1}{p^2}$$

$$=\; \frac{1}{p}$$

Did u know? What is the second term of summand?

14.3.4 Variance: Var(X)

Variance of a Geometric random variable having parameters p is calculated as follows. Now,

$$E[X^2] \;=\; \sum_{i=1}^{\infty} i^2 q^{i-1} p = p\sum_{i=1}^{\infty}\frac{d}{dq}(iq^i) = p\,\frac{d}{dq}\sum_{i=1}^{\infty} iq^i$$

$$=\; p\,\frac{d}{dq}\sum_{i=1}^{\infty} iq^{i-1} p \times \frac{q}{p} = p\,\frac{d}{dq}\left[\frac{q}{1-q}\sum_{i=1}^{\infty} iq^{i-1} p\right]$$

$$=\; p\,\frac{d}{dq}\left[\frac{q}{1-q}\cdot E[X]\right] = p\cdot\frac{d}{dq}\left(\frac{q}{(1-q)^2}\right)$$

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

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

Thus,
$$Var(X) \;=\; E(X^2) - [E(X)]^2$$

$$=\; \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p}$$

$$=\; \frac{1-p}{p^2}$$

Task   Find the variance of a Geometric random variable having parameter 0.

14.3.5 Moment Generating Function (MGF)

Moment Generating Function (MGF) of a Geometric random variable with parameter p is,

$$M(t) \;=\; E[e^{tx}] = \sum_{i=1}^{\infty} e^{ti}(1-p)^{i-1} p$$

$$= p\sum_{i=1}^{\infty}\left(e^t\right)^{i-1}e^t(1-p)^{i-1}=pe^t\sum_{i=1}^{\infty}\left[e^t(1-p)\right]^{i-1}$$

The summation is of G.P. whose first term is 1 and common ratio is $e^t(1-p)$. Since t is going to be equated to zero, we consider t to be close to zero. Hence $0 < e^t(1-p) < 1$. Thus, the sum is

$\dfrac{1}{1-e^t(1-p)}$. Thus,

$$M(t) = \frac{pe^t}{1-(1-p)e^t} \qquad \text{This is the required MGF}$$

Now,
$$M'(t) = \frac{\left[1-(1-p)e^t\right]pe^t - pe^t\left[-(1-p)e^t\right]}{\left[1-(1-p)e^t\right]^2}$$

$$= \frac{pe^t - p(1-p)e^{2t} + p(1-p)e^{2t}}{\left[1-(1-p)e^t\right]^2}$$

$$= \frac{pe^t}{\left[1-(1-p)e^t\right]^2}$$

$$M''(t) = \frac{\left[1-(1-p)e^t\right]^2 pe^t - pe^t\times 2\left[1-(1-p)e^t\right]\left[-(1-p)e^t\right]}{\left[1-(1-p)e^t\right]^4}$$

Now,
$$E[X] = M'(0)=\frac{p}{p^2}=\frac{1}{p}$$

and

$$E[X^2] = M''(0)=\frac{p^3+2p^2(1-p)}{p^4}=\frac{2p^2-p^3}{p^4}$$

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

This is same as the earlier result.

14.3.6 Applications of Geometric Distribution

Geometric distributions are used in many applications, since they give probability distribution of time before a particular event happens. Some of the examples are:

1. Number of attempts required before a certain experiment succeeds.
2. Number of times a product can be used before it fails.
3. Number of times we have to draw an item from a mixed lot till the item with the required property is drawn.
4. Number of missiles required to be fired to hit the target with certain assurance.

14.3.7 Memory-less Property of Geometric Distribution

One of the interesting properties of geometric random variable is, it is memory-less i.e. it does not remember the number of trials conducted before the present trial. Just to imagine in practice, probability of accident taking place in the next one day does not change whether accident has taken place today or not.

Definition of Memory-less Property

We say that a nonnegative random variable X is memory-less if the probability that t year old item will survive for at least $(s + t)$ hours, is same as the initial probability that it survives for at least s hours. Mathematically, memory-less property can be expressed as,

$$P\{X > s+t | X > t\} \;=\; P\{X > s\} \qquad\qquad \text{for all } s, t \geq 0$$

Using conditional probability theorem, $P(E|F) = \dfrac{P(E,F)}{P(F)}$ we get,

$$\frac{P\{X > s+t, X > t\}}{P\{X > t\}} \;=\; P\{X > s\}$$

Or, the memory-less property reduces to,

$$P\{X > s+t\} \;=\; P\{X > s\}\, P\{X > t\}$$

Geometric Distribution is Memory-less

We know that for geometric distribution,

$$P\{X > a\} \;=\; 1 - P\{X \leq a\} = 1 - F(a) = 1 - (1 - q^a) = q^a$$

Therefore,

$$P\{X > s+t\} \;=\; q^{(s+t)}$$

$$P\{X > s\} \;=\; q^s$$

$$P\{X > t\} \;=\; q^t$$

Now,

$$P\{X > s+t\} \;=\; q^{(s+t)} = q^s q^t$$

$$\;=\; P\{X > s\}\, P\{X > t\}$$

Thus, geometric distribution is memory-less.

Example: PMF of a random variable is,

$$P(X = i) = 2^{-i} \quad for \quad i = 1, 2, ...$$

(i)   Verify it is PMF   (ii)   Find mean.   (iii)   Find variance.

Solution:

(i)   PMF must satisfy the axioms of probability. Now if we will first prove the second axiom.

That is,
$$\sum_{i=1}^{n} P(X = i) = 1$$

Now,
$$\sum_{i=1}^{n} P(X = i) = \sum_{i=1}^{\infty} 2^{-i} = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{-i} = \left(\frac{1}{2}\right)^{1} + \left(\frac{1}{2}\right)^{2} + \left(\frac{1}{2}\right)^{3} + ...$$

$$= \frac{\frac{1}{2}}{1 - \left(\frac{1}{2}\right)} = 1 \quad \text{Axiom II is satisfied.}$$

We have used infinite geometric series result stated as,

$ar^{0} + ar^{1} + ar^{2} + ... = \dfrac{a}{1 - r}$ Where a is the first term and r is the common ratio. In this example, $a = \dfrac{1}{2}$ and $r = \dfrac{1}{2}$

It is clear that P for any value of i is positive, since it is a power of a positive number. Also, sum of all P is 1. Hence,

$0 \le P(X = i) \le 1$ Axiom I is satisfied.

Hence, given $P(X = i)$ is PMF

(ii) Now, we can see that the given random variable is a geometric random variable with parameter $p = \dfrac{1}{2}$. Thus, the mean is given by,

$$\mu = E(X) = \frac{1}{p} = 2$$

(iii) Variance of geometric random variable with parameter $p = \dfrac{1}{2}$ is,

$$\text{var}(X) = \frac{1 - p}{p^{2}} = 2$$

14.4 Negative Binomial Distribution

Suppose that independent trials, each having probability of being successful as 'p' ($0 < p < 1$), are performed until a total of r successes is accumulated, then number of trials required to achieve r successes follow negative binomial distribution.

14.4.1 Probability Mass Function (PMF)

A random variable X, taking on one of the countable infinite values $r, r + 1, r + 2$............., is said to be Negative Binomial random variable with parameters (r, p), if for some $0 < p < 1$, and $r > 0$, probability mass function, (PMF) is,

$$P(X = n) \;=\; \binom{n-1}{r-1} p^r (1-p)^{n-r} \qquad\qquad n = r, r+1, r+2, ...$$

Note that with $r = 1$, the distribution reduces to a Geometric distribution.

Proof:

Suppose that independent trials, each having probability of being successful as 'p' $(0 < p < 1)$, are performed until a total of r successes is accumulated, then the probability of number of trials required to achieve r successes is given by,

$$P(X = n) \;=\; \binom{n-1}{r-1} p^r (1-p)^{n-r} \qquad\qquad n = r, r+1, r+2, ...$$

Proof:

Suppose that independent trials, each having probability of being successful as 'p' $(0 < p < 1)$, are performed until a total of r successes is accumulated. Let the rth success be achieved in nth trial. Hence, $(r - 1)$ successes must have been achieved in $(n - 1)$ trials. Probability of achieving $(r - 1)$ successes and hence, $[(n - 1) - (r - 1)] = (n - r)$ failures in total of $(n - 1)$ trials in any particular sequence is $p^{r-1}(1-p)^{n-r}$. However, these $(r - 1)$ successes could be any of the $(n - 1)$ trials, which can be in $p^{r-1}(1-p)^{n-r}$ different ways. Hence, probability of achieving $(r - 1)$ successes in any of the $\binom{n-1}{r-1}$ trials is:

$$\binom{n-1}{r-1} p^{r-1} (1-p)^{n-r}$$

Now the nth trial must be a success, which has a probability of p. Therefore, by product rule, probability of number of trials required to achieve r successes is, $\binom{n-1}{r-1} p^r (1-p)^{n-r}$. Further, we can achieve r^{th} success in any of the trials from $n = r, r + 1, r + 2$....... Thus, the result is proved.

Notes

1. Since p and $(1 - p)$ are positive, their powers must be positive. $\binom{n-1}{r-1}$ is number of ways of selecting items and hence, positive. Therefore, $P(X = n) \geq 0$. Now, to prove Axiom II we use Binomial theorem for negative powers as follows:

    Using Binomial theorem,

    $$(a + b)^n \;=\; \binom{n}{0} a^0 b^n + \binom{n}{1} a^1 b^{n-1} + \binom{n}{2} a^2 b^{n-2} + ...$$

Contd...

$$= \frac{a^0}{0!}b^n + n\frac{a^1}{1!}b^{n-1} + n(n-1)\frac{a^2}{2!}b^{n-2} + n(n-1)(n-2)\frac{a^3}{3!}b^{n-3} + ...$$

Now putting, $n = -r$ and $b = 1$, we get

$$(a+1)^{-r} = \frac{a^0}{0!} - r\frac{a^1}{1!} + r(r+1)\frac{a^2}{2!} - r(r+1)(r+2)\frac{a^3}{3!} + ...$$

Now putting, $a = -x$

$$(1-x)^{-r} = \frac{x^0}{0!} + r\frac{x^1}{1!} + r(r+1)\frac{x^2}{2!} + r(r+1)(r+2)\frac{x^3}{3!} + ...$$

This is one of the forms of negative binomial theorem.

Now,

$$\sum_{n=r}^{\infty} P(X=n) = \sum_{n=r}^{\infty} \binom{n-1}{r-1}p^r(1-p)^{n-r}$$

$$= \binom{r-1}{r-1}p^r(1-p)^{r-r} + \binom{r}{r-1}p^r(1-p)^1 + \binom{r+2}{r-1}p^r(1-p)^3 + ...$$

$$= p^r\left[\frac{(1-p)^0}{0!} + r\frac{(1-p)}{1!} + (r+1)(r)\frac{(1-p)^2}{2!} + (r+2)(r+1)(r)\frac{(1-p)^3}{3!} + ...\right]$$

Comparing with the result of negative binomial expansion,

$$\sum_{n=r}^{\infty} p(X=n) = p^r\left[1-(1-p)\right]^{-r}$$

$$= p^r \cdot p^{-r} = 1 \qquad\qquad \text{Axiom II is satisfied.}$$

2. Since, $\sum_{n=r}^{\infty} p(X=n) = 1$ and also $P(X=n) \geq 0$, this implies that

$$0 \leq P(X=n) \leq 1 \qquad\qquad \text{Axiom I is satisfied.}$$

14.4.2 Expected Value or Mean (μ)

Expectation of a Negative Binomial random variable having parameters (r, p) is calculated as follows:

$$E(X^k) = \sum_{n=r}^{\infty} n^k\binom{n-1}{r-1}p^r(1-p)^{n-r} = \sum_{n=r}^{\infty} n^k\frac{(n-1)!}{(n-r)!(r-1)!}p^r(1-p)^{n-r}$$

$$= \sum_{n=r}^{\infty} \frac{n^{k-1} \times n!}{(n-r)!r!} \times rp^r(1-p)^{n-r} = \frac{r}{p}\sum_{n=r}^{\infty} n^{k-1}\binom{n}{r}p^{r+1}(1-p)^{n-r}$$

Or   we could have used identity $n\binom{n-1}{r-1} = r\binom{n}{r}$

Now, put $n = m-1$. Thus,

$$E(X^k) \;=\; \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{k-1} \binom{m-1}{r} p^{r+1} (1-p)^{m-(r+1)}$$

Now if we compare the terms, we can find that our new PMF is for negative binomial random variable with parameters $(r+1, p)$ instead of (r, p).

Thus,

$$E(X^k) \;=\; \frac{r}{p} E\left[(y-1)^{k-1} \right]$$

Where X is a negative binomial random variable with parameters (r, p) and Y is a negative binomial random variable with parameters $(r + 1, p)$. Now putting $k = 1$ we get,

$$E(X) \;=\; \frac{r}{p} E\left[(y-1)^0 \right] = \frac{r}{p} E[1]$$

$\therefore$ $\qquad\qquad\qquad \mu \;=\; E[X] = \dfrac{r}{p}$

Task   Calculate the expectation of a Binomial random variable having parameters (0, 1).

14.4.3 Variance: Var(X)

Variance of a negative binomial random variable having parameters (r, p) could be calculated as follows. Putting $k = 2$ in the expression for $E(X^k)$ above, we get,

$$E(X^2) \;=\; \frac{r}{p} E(y-1) = \frac{r}{p}\left[E(Y) - E(1) \right] = \frac{r}{p}\left[\frac{r+1}{p} - 1 \right]$$

Because $\qquad\qquad E[Y] \;=\; \dfrac{r+1}{p}$

Now,

$$\mathrm{var}(X) \;=\; E(X^2) - [E(X)]^2$$

$$=\; \frac{r}{p}\left[\frac{r+1}{p} - 1 \right] - \left(\frac{r}{p} \right)^2 = \frac{r}{p}\left(\frac{r}{p} + \frac{1}{p} + -1 - \frac{r}{p} \right)$$

$$=\; \frac{r(1-p)}{p^2}$$

14.4.4 Moment Generating Function (MGF)

Moment Generating Function (MGF) of a negative binomial random variable having parameter (r, p) is,

$$
\begin{aligned}
M(t) &= E[e^{tx}] = \sum_{n=r}^{\infty} e^{tn} \binom{n-1}{r-1} p^r (1-p)^{(n-r)} \\[2ex]
&= p^r e^{tr} \sum_{n=r}^{\infty} e^{(tn-tr)} \binom{n-1}{r-1} (1-p)^{(n-r)} \\[2ex]
&= p^r e^{tr} \sum_{n=r}^{\infty} \binom{n-1}{r-1} \left[(1-p)e^t\right]^{(n-r)}
\end{aligned}
$$

Now, using negative binomial theorem (binomial expansion for negative powers) that we have proved, we get

$$
M(t) = (pe^t)^r \left[1-(1-p)e^t\right]^{-r} = \left[\frac{pe^t}{1-(1-p)e^t}\right]^r
$$

This is the required MGF Now,

$$
\begin{aligned}
M'(t) &= r\left[\frac{pe^t}{1-(1-p)e^t}\right]^{r-1} \left[\frac{[1-(1-p)e^t]pe^t - pe^t(1-p)e^t(-1)}{(1-(1-p)e^t)^2}\right] \\[2ex]
&= r\left[\frac{pe^t}{1-(1-p)e^t}\right]^{r-1} \left[\frac{pe^t - p(1-p)e^{2t} + p(1-p)e^{2t}}{(1-(1-p)e^t)^2}\right] \\[2ex]
&= r\left[\frac{pe^t}{1-(1-p)e^t}\right]^{r-1} \frac{pe^t}{(1-(1-p)e^t)^2} \\[2ex]
&= r\left[\frac{pe^t}{1-(1-p)e^t}\right]^{r} \frac{1}{1-(1-p)e^t}
\end{aligned}
$$

$$
M''(t) = r\left[\frac{(1-(1-p)e^t\phi'(t) - \phi(t)(1-p)e^t(-1)}{(1-(1-p)e^t)^2}\right]
$$

Now, putting t =0, in these expressions, we get,

$$
M(t) = 1
$$

$$
E(X) = \phi'(0) = \frac{r}{p}
$$

$$
E(X^2) = M''(0) = r\frac{p \cdot \dfrac{r}{p} + 1(1-p)}{p^2} = \frac{r}{p}\left[\frac{r+1-p}{p}\right] = \frac{r}{p}\left[\frac{r+1}{p} - 1\right]
$$

These are same as the earlier results.

14.4.5 Applications of Negative Binomial Distribution

Negative binomial distributions are used in many applications, since they give probability distribution of number of trials required before a particular event happens for *r* times. Some of the examples are:

1. Number of attempts required before a certain experiment succeeds *r* times.

2. Number of times a product can be used if its number of usage cycles is *r*.

3. Number of times we have to draw items from a mixed lot, till *r* items with the required property are drawn.

4. Number of missiles required to be fired to achieve multiple hits on the target to get the required certain assurance of destruction.

Caution The geometric random variable is a negative Binomial random variable with parameters (1, *p*). Hence, if we substitute *r* = 1 in all results of negative random distribution, we get the results of geometric random distribution.

14.5 Hyper-geometric Distribution

Binomial distribution is applied whenever we draw a random sample with replacement. This is because in such case probability of success *p* remains constant in every draw. Also, the successive draws remain independent. However, in case of random samples without replacement, probability of success *p* does not remain constant. Also, such draws are not independent. In such case, the Hyper-geometric distribution is used.

To explain this, let us consider a bag containing 8 balls; 3 being red and 5 being white. Suppose we draw two balls with replacement, that is after the first draw we put the drawn ball back in the bag. For both the draws, probability of drawing red ball is $\frac{3}{8}$. On the other hand, if we draw two balls without replacement, the probability of drawing a red ball in the first draw is $\frac{3}{8}$. But in the second draw, the probability of drawing a red ball is $\frac{2}{7}$ or $\frac{3}{7}$ depending on whether the ball drawn in the first draw is red or white respectively. Thus, the probability of success *p* changes from draw to draw. Also, it is dependent on the result of the previous draw.

14.5.1 Probability Mass Function (PMF)

Suppose that a sample size *n* is to be chosen randomly without replacement, from the population of *N* items of which *m* are of a special type (say defective items) and remaining *N* – *m* are of other type (say serviceable items). The number of special type items in the drawn sample follows

Hyper-geometric distribution. If we let *X* denote the number of special items out of selected sample of size *n* then, PMF of random variable *X* is given by,

$$P(X=i) \ = \ \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \qquad\qquad i = 0, 1,\ldots\ldots n$$

A random variable *X*, whose probability mass function is given by the above equation for some values of (*n, N, m*) is said to be a hyper-geometric random variable and denoted as $X \rightarrow H(n,N,m)$.

Notes   Although the value of *i* should between 0 and n (both inclusive), it must satisfy

$n - (N - m) \leq i \leq minimum\ (n, m)$

Prove:

Suppose that a sample size *n* is to be chosen randomly without replacement, from the population of *N* items, of which *m* are of a special type (say defective items) and remaining *N – m* are of other type (say serviceable items); probability that *i* items are special items out of selected sample of size *n* is given by,

$$P(i) \ = \ \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

Proof:

If the sample of size *n* has *i* items of special type, they must have come from the *m* special items in the population. This selection is possible in $\binom{m}{i}$ ways. Remaining (*n – i*) non-special items in the sample must have come from (*N – m*) non-special items in the population. This selection is possible in $\binom{N-m}{n-i}$ ways. Thus, by the product rule number of ways of selecting a sample of size *n* that contains *i* special items is $\binom{m}{i}\binom{N-m}{n-i}$. Total number of ways of selecting the sample

of size n is $\binom{N}{n}$ Hence, by definition of probability, the probability that i items are special items out of selected sample of size n is,

$$P(i) \;=\; \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

Proved.

Notes

1. Proving that PMF satisfies axioms of probability is very simple. First PMF is ratio of number of way of selections and hence, must be greater than zero i.e. $P(i) \geq 0$.

   Also, $\displaystyle\sum_{i=0}^{n} p(X=i) \;=\; \sum_{i=0}^{n} \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$

   $$=\; \frac{\binom{m}{0}\binom{N-m}{n}+\binom{m}{1}\binom{N-m}{n-1}+\binom{m}{2}\binom{N-m}{n-2}+.....+\binom{n}{n}\binom{N-m}{0}}{\binom{N}{n}}$$

   $$=\; \frac{\binom{m+N-m}{n}}{\binom{N}{n}} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1 \qquad \text{Axiom II is satisfied.}$$

   Here we used Vander monde identity or Convolution Theorem,

   $$\binom{m+n}{r} \;=\; \binom{m}{0}\binom{n}{r}+\binom{m}{1}\binom{n}{r-1}+......+\binom{m}{r}\binom{n}{0}$$

2. Since $P(i) \geq 0$ and $\displaystyle\sum_{i=0}^{n} P(i)=1$, obviously,

   $0 \leq P(i) \leq 1$ \qquad Axiom I is satisfied.

14.5.2 Expected Value or Mean (μ)

Expectation of a hyper-geometric random variable having parameters (n, N, m) is calculated as follows,

$$\mu \ = \ E[X^k] = \sum_{i=0}^{n} i^k p(i) \qquad\qquad = \sum_{i=0}^{n} i^k \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

Using $\qquad i\binom{m}{i} \ = \ m\binom{m-1}{i-1} \qquad$ and $\qquad n\binom{N}{n} = N\binom{N-1}{n-1}$

$$E[X^k] \ = \ \sum_{i=0}^{n} i^k m \frac{\binom{m-1}{i-1}\binom{N-m}{n-1}}{\frac{N}{n}\binom{N-1}{n-1}} = \frac{nm}{N} \sum_{i=1}^{n} i^{k-1} \frac{\binom{m-1}{i-1}\binom{N-m}{n-i}}{\binom{N-1}{n-1}}$$

We have removed term $i = 0$ since its value is zero.

Now, we substitute $i = j + 1$

Thus, for $i = 1$ we have $j = 0$ *and* for $i = n$ we get $j = n$-1. Therefore,

$$E[X^k] \ = \ \frac{nm}{N} \sum_{j=0}^{n-1} \frac{(j+1)^{k-1}\binom{m-1}{j}\binom{N-m}{n-1-j}}{\binom{N-1}{n-1}}$$

Comparing terms we see that n is replaced by $(n - 1)$, m is replaced by $(m - 1)$ and N is replaced by $(N - 1)$. Thus, considering Y as a Hyper-geometric random variable with parameters $(n -1, N -1, m -1)$ and its dummy variable indicated by j, we can write,

$$E[X^k] \ = \ \frac{nm}{N} E\left[(Y+1)^{k-1}\right]$$

Now, putting $k =1$ we get,

$$\mu \ = \ E[X] = \frac{nm}{N} E[1] = \frac{nm}{N}$$

Notes   If we call $\dfrac{m}{N} = p$ that is probability of defective, then expected value is np as in case of Binomial distribution.

14.5.3 Variance: Var(X)

Variance of a hyper-geometric random variable having parameters (n, N, m) could be calculated as follows. Putting $k = 2$ in the expression for $E(X^k)$ above, we get,

$$E\left[X^2\right] \;=\; \frac{nm}{N}E(Y+1) = \frac{nm}{N}[E(Y)+E(1)]$$

$$=\; \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1}+1\right]$$

$$Var(X) \;=\; E(X^2)-[E(X)]^2 = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1}+1\right]-\left(\frac{nm}{N}\right)^2$$

$$=\; \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1}+1-\frac{nm}{N}\right] = \frac{nm}{N}\left[\frac{nm-n-m+1+N-1}{N-1}-\frac{nm}{N}\right]$$

$$=\; \frac{nm}{N}\left[\frac{Nnm-nN-mN+N^2-Nnm+nm}{N(N-1)}\right] = \frac{nm}{N}\left[\frac{N(N-n)-m(N-n)}{N(N-1)}\right]$$

$$=\; \frac{nm}{N}\cdot\frac{(N-m)(N-n)}{N(N-1)}$$

putting

$$p \;=\; \frac{m}{N}$$

$$Var(X) \;=\; n\times p\times(1-p)\times\frac{N-n}{N-1} = \frac{N-n}{N-1}\times np(1-p)$$

Notes	Suppose N is large in relation to n, $\dfrac{N-n}{N-1}\approx 1$

Then, $Var(X) \;=\; np(1-p)$

This is same as Binomial distribution. Thus, if N is large as compared to n, that is, population is large as compared to the sample, we can approximate the process of drawing sample without replacement to that with replacement. Hence, we can use Binomial random variable (and hence Binomial distribution) in place of hyper-geometric random variable (or Hyper-geometric distribution). The factor $\dfrac{N-n}{N-1}$ is sometimes known as finite population factor.

14.5.4 Binomial Approximation to Hyper-geometric Distribution

The hyper-geometric random variable approaches binomial random variable if N is large in relation to n i.e. ($N \to \infty$) and m is large in relation to i.

Proof:

Let $X \rightarrow H(n, N, m)$. Hence,

$$P(i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} = \frac{m!}{(m-i)!i!} \cdot \frac{(N-m)!}{(N-m-n+i)!(n-i)!} \cdot \frac{(N-n)!n!}{N!}$$

$$= \binom{n}{i} \cdot \frac{m}{N} \cdot \frac{m-1}{N-1} \cdots \cdots \frac{m-i+1}{N-i+1} \cdot \frac{N-m}{N-i} \cdot \frac{N-m-1}{N-i-1} \cdots \cdots \frac{N-m-(n-i-1)}{N-i-(n-i-1)}$$

$$\approx \binom{n}{i} p^i (1-p)^{n-i}$$

Because,

$$\frac{m-x}{N-x} = \frac{\frac{m}{N} - \frac{x}{N}}{1 - \frac{x}{N}} \rightarrow p \text{ as } N \rightarrow \infty \text{ for all finite values } x = 1, 2, \ldots i-1 \text{ and } \frac{m}{N} = p.$$

Also, $\dfrac{N-m}{N-x} = \dfrac{1 - \frac{m}{N}}{1 - \frac{x}{N}} = \dfrac{1-p}{1} = 1 - p$ as $N \rightarrow \infty$ for all finite values x and $\dfrac{m}{N} = p$.

14.5.5 Applications of Hyper-geometric Distribution

Hyper-geometric distribution is used when a random sample, taken without replacement from the population, consists of two classes (e.g. defective and not defective items). Some of the common applications are,

1. Sample testing in quality control department. In this case, sample is usually not put back in stock from which the sample is drawn and tested.

2. A lake contains N fish. A sample of fish is taken from the lake, marked and released back in the lake. Next time, another sample of fish is selected and the marked fish are counted.

3. Opinion surveys, when the persons have to give answers of 'yes' or 'no' type.

The following conditions must be satisfied for the application of the hyper-geometric distribution.

1. The population is divided into two mutually exclusive categories.

2. The successive outcomes are dependent.

3. The probability of success changes from trial to trial.

4. The number of draws is fixed.

Did u know? What is Sample Testing?

14.5.6 Alternate form of Hyper-geometric Distribution

Sometimes the PMF of a hyper-geometric random variable is written in the following form.

$$P(X=i) = \frac{\binom{Np}{i}\binom{Nq}{n-i}}{\binom{N}{n}} \qquad i = 0, 1,\ldots\ldots n$$

$$= 0 \qquad\qquad \text{Otherwise}$$

Where, p = proportion of the items belonging to the class possessing the characteristics of interest and $q = 1 - p$

Note that $p = \dfrac{m}{N}$ and $q = 1 - \dfrac{m}{N} = \dfrac{N-m}{N}$

The mean and variance is given by,

$$\mu = E(X) = np$$

$$Var(X) = npq\left(\frac{N-n}{N-1}\right)$$

14.6 Multinomial Distribution

Multinomial random variable is very useful in practice, which counts the number of outcomes of particular categories, when 'n' random trials are performed which has r different possible outcomes. For example, the product from a machine when inspected falls in three categories viz. satisfactory, reject, and requires rework. Based on the past data, we know the proportion of these categories in the production lot. Now, we randomly pick, say 100 items and put them in packet and want to know the probability that out of 100 a specific number of items are satisfactory, reject and requiring rework. This probability is given by multinomial distribution. Multinomial random variable is a generalised version of binomial random variable when the outcomes of the experiment are more than two. In real life problems, there are many such situations.

Suppose n independent trials are performed. Each trial results in any of the r possible outcomes namely $a_1, a_2, a_3, \ldots a_r$. Further, suppose that probabilities of getting these outcomes in any of the trials are $p_1, p_2, p_3, \ldots p_r$ respectively. That is probability of getting a_1 as an outcome in any of the trial is p_1, probability of getting a_2 as an outcome in any of the trial is p_2, and so on. If X represents multi-dimensional variable ($X: n_1, n_2, n_3, \ldots n_r$) where $n_1, n_2, n_3, \ldots n_r$ are total number occurrences of outcomes $a_1, a_2, a_3, \ldots a_r$ in n trials, then X is said to be a Multinomial random variable and the probability distribution of $p_1, p_2, p_3, \ldots p_r$ is known as multinomial distribution with parameters ($n; n_1, n_2, n_3, \ldots n_r; p_1, p_2, p_3, \ldots p_r$). Note that $n_1 + n_2 + \ldots + n_r = n$ and $p_1 + p_2 + \ldots + p_r = 1$

Thus, Binomial distribution is just a Multinomial distribution with $r = 2$ or with parameters ($n; n_1, n_2; p_1, p_2$) i.e. ($n; n_1, 1- n_1; p_1, 1- p_1$) or just ($n, p$).

Probability Mass Function (PMF)

PMF of Multinomial random variable with parameters $(n; n_1, n_2, n_3, \ldots n_r;\ p_1, p_2, p_3, \ldots p_r)$ is given by,

$$P(n_1, n_2, \ldots n_r) = \binom{n}{n_1, n_2 \ldots n_r}(p_1)^{n_1}(p_2)^{n_2}\ldots\ldots(p_r)^{n_r}$$

$$= \frac{n!}{n_1! n_2! \ldots n_r!}(p_1)^{n_1}(p_2)^{n_2}\ldots\ldots(p_r)^{n_r}$$

Where, $n_1 + n_2 + \ldots + n_r = n$

$$\binom{n}{n_1, n_2 \ldots n_r} = \frac{n!}{n_1! n_2! \ldots n_r!} \text{ is called a multinomial coefficient.}$$

Proof:

Let $A_1, A_2, A_3, \ldots A_r$ be r mutually exclusive and collectively exhaustive events associated with a random experiment such that probability of $P(A_i$ occurs$) = p_i$ where i = 1, 2, ...,r with $(p_1 + p_2 + \ldots + p_r) = 1$. This is obvious because sum of the probabilities of mutually exclusive and collectively exhaustive events is always 1 as per the probability Axiom II. If the experiment is repeated n times, then we have to show that probability of A_1 occurring n_1 times, A_2 occurring n_2 times, $\ldots A_r$ occurring nr times, is given by

$$P(n_1, n_2 \cdots, n_r) = \frac{n!}{n_1! n_2! \ldots n_r!}(p_1)^{n_1}(p_2)^{n_2}\ldots\ldots(p_r)^{n_r}$$

Since $n, n_1, n_2, n_3, \ldots n_r$ are number of occurrences, these are non-negative integers. Also since $A_1, A_2, A_3, \ldots A_r$ be r mutually exclusive and collectively exhaustive events, and their occurrences in total trials n are $n_1, n_2, \ldots, n_r$; we know that $n_1 + n_2 + \ldots + n_r = n$

Now out of n random trials, any n_1 trials have event A_1 as outcomes. However, probability of any event to be A_1 is p_1. Hence, by product rule, the probability of event A_1 occurring n_1 times is, $p_1 \times p_1 \times \ldots\ldots \times p_1$ etc. n_1 times $= p_1^{n_1}$. Further, trials in which event A_1 occurs could be any of the n_1 trials out of n. These could be in $\binom{n}{n_1}$ ways.

Similarly, out of n random trials, any n_2 trials have event A_2 as outcome. However, probability of any event to be A_2 is p_2. Hence, by product rule, the probability of events A_2 occurring n_2 times is, $p_2^{n_2}$. Further, trials in which event A_2 occurs could be any of the n_2 trials out of remaining $(n - n_1)$. These could be in $\binom{n - n_1}{n_2}$ ways.

Progressing in a similar way, till last event A_r occurs n_r times, each with probability p_r. Now using the product rule, the probability of event A_1 occurring n_1 times, A_2 occurring n_2 times, $\ldots A_r$ occurring n_r times out of total n trials is,

$$P(n_1, n_2 \cdots, n_r) = p_1^{n_1}\binom{n}{n_1} \times p_2^{n_2}\binom{n - n_1}{n_2} \times p_3^{n_3}\binom{n - n_1 - n_2}{n_3} \times \cdots\cdots \times p_r^{n_r}\binom{n - n_1 - n_2 - \cdots\cdots - n_{r-1}}{n_r}$$

$$= \frac{n!}{n_1!(n - n_1)!} \times \frac{(n - n_1)!}{n_2!(n - n_1 - n_2)!} \times \frac{(n - n_1 - n_2)!}{(n - n_1 - n_2 - n_3)! n_3!} \times \cdots\cdots$$

$$\times \frac{(n-n_1-n_2-n_3-\cdots-n_{r-1})!}{(n-n_1-n_2-n_3-\cdots-n_{r-1}-n_r)!n_r!} \times p_1^{\,n_1} \times p_2^{\,n_2} \times p_3^{\,n_3} \times \ldots\ldots \times p_r^{\,n_r}$$

$$= \frac{n!}{n_1!n_2!\ldots n_r!} \times p_1^{\,n_1} \times p_2^{\,n_2} \times p_3^{\,n_3} \times \ldots\ldots \times p_r^{\,n_r}$$

Hence, the result is proved.

Here while cancelling the terms from denominator and numerator, we have used relation,

$$n_1 + n_2 + \ldots + n_r = n$$

or, $\qquad (n - n_1 - n_2 - n_3 - \cdots - n_{r-1} - n_r) = 0$

Hence, $\qquad (n - n_1 - n_2 - n_3 - \cdots - n_{r-1} - n_r)! = 0! = 1$

Notes $\qquad$ Because $\binom{n}{i} \geq 0$, $p_i \geq 0$ for all i we can say that, $P(n_1, n_2 \cdots, n_r) \geq 0$

Also,

$$\sum_{all} P(n_1, n_2 \cdots, n_r) = \sum_{all} \frac{n!}{n_1!n_2!\ldots n_r!} \times p_1^{\,n_1} \times p_2^{\,n_2} \times p_3^{\,n_3} \times \ldots\ldots \times p_r^{\,n_r}$$

$$= (p_1 + p_2 + \ldots + p_r)^n \qquad \qquad \text{Using Multinomial Theorem}$$

$$= 1 \qquad \qquad \text{Because } (p_1 + p_2 + \ldots + p_r) = 1$$

$$\therefore \sum_{all} p(n_1, n_2, n_3, \ldots, n_r) = 1 \qquad \qquad \text{Axiom II is satisfied.}$$

Hence, $\qquad\qquad 0 \leq p(n_1, n_2, n_3, \ldots, n_r) \leq 1 \qquad \qquad \text{Axiom I is satisfied.}$

Case Study

Probability of vehicle breakdown of a transport company is $\frac{1}{1200}$. Calculate the probability that in a fleet of 300 vehicles, there will be 2 or more vehicles break down on a given day.

Solution:

Now, average rate of failure for the fleet is, $\lambda = np = 300 \times \frac{1}{1200} = 0.25$. Thus, this is a case of a Poisson distribution with parameter ($\lambda = 0.25$). Now,

$$P(X \geq 2) = 1 - P(0) - P(1)$$

$$= 1 - \frac{e^{-0.25}(0.25)^0}{0!} - \frac{e^{-0.25}(0.25)^1}{1!} = 0.265$$

14.7 Summary

- Binomial random variable is very useful in practice, which counts the number of successes when 'n' Bernoulli trials are performed.

- When number of trials or sample space is infinitely large, and where each trial is Bernoulli trial, Poisson random variable is used.

- Suppose that independent trials are performed until a success occurs, and each trial has a probability of success as p ($0 < p < 1$). Now if a random variable X is number of trials required till success occurs, then X is called as Geometric random variable.

- Suppose that independent trials, each having probability of being successful as '' , are performed until a total of successes is accumulated, then number of trials required to achieve successes follow negative binomial distribution.

- Binomial distribution is applied whenever we draw a random sample with replacement. This is because in such case probability of success p remains constant in every draw.

- Multinomial random variable is very useful in practice, which counts the number of outcomes of particular categories, when 'n' random trials are performed which has r different possible outcomes.

14.8 Keywords

Expected Value: In probability theory, the expected value (or expectation, or mathematical expectation, or mean, or the first moment) of a random variable is the weighted average of all possible values that this random variable can take on.

Moment Generating Function: In probability theory and statistics, the moment-generating function of any random variable is an alternative definition of its probability distribution.

Variance: In probability theory and statistics, the variance is used as a measure of how far a set of numbers are spread out from each other. It is one of several descriptors of a probability distribution, describing how far the numbers lie from the mean (expected value).

14.9 Self Assessment

1. The lifetime in hours of a certain kind of IC is a random variable having a p.d.f. of failure given by,

$$f(x) = \begin{cases} 0 & if \quad x \leq 100 \\ \dfrac{100}{x^2} & if \quad x > 100 \end{cases}$$

   Assume that the lifetimes of all the five IC_s are independent of each other.

   (a) What is the probability of exactly 2 out of 5 IC_s in equipment had to be replaced within the first 150 hours of operation?

   (b) The equipment fails to operate, if more than half IC_s fail. What is the probability that the equipment will survive 150 hours?

2. Find the value of c for which following define PMF with $i = 1, 2, 3, ...$

    (a) $P(X = i) = c(2)^{-i}$

    (b) $P(X = i) = \dfrac{c(2)^{-i}}{i}$

    (c) $P(X = i) = \dfrac{c(2)^{i}}{i!}$

    (d) $P(X = i) = c(i)^{-2}$

3. Calculate MGF for a discrete random variable with PMF

    $$P(X = m) = \binom{n+m-1}{m} p^{n}(1-p)^{m} \qquad \text{for } m \ge 0$$

4. A purchaser of transistors buys them in lots of 20. It is his policy to randomly inspect 4 components from a lot and to accept the lot only if all 4 are non-defective. If each component in a lot is independently defective with probability 0.1, what proportion of lot is rejected?

5. The PMF of a random variable X is,

    $$P(X = x_i) = \begin{cases} \dfrac{1}{15} & \text{for } x_i = 1,2,3,.....,15 \\ 0 & \text{otherwise} \end{cases}$$

    Find

    (i) $E(X)$,

    (ii) $Var(X)$

6. If X is a discrete random variable with probability mass function PMF as,

    $$P(x) = Cp^{x} \quad \text{for} \quad x = 1,2,3,.....$$

    Find

    (i) C

    (ii) $E(X)$

7. Let $A_1, A_2, A_3, \ldots A_k$ be k mutually exclusive and collectively exhaustive events associated with a random experiment, such that probability of $P(A_i \text{ occurs}) = p_i$ where $i = 1, 2, ..., k$ with $p_1 + p_2 + \cdots + p_k = 1$. If the experiment is repeated n times, then we have to show that probability of A_1 occurring r_1 times, A_2 occurring r_2 times, $\ldots A_k$ occurring r_k times, is given by

    $$P(r_1, r_2 \cdots, r_k) = \frac{n!}{r_1! r_2! \ldots r_k!} (p_1)^{r_1} (p_2)^{r_2} \ldots (p_k)^{r_k}$$

    Where $(r_1 + r_2 + \ldots + r_k) = n$

8. In a binomial distribution consisting of 5 independent trials, probabilities of 1 and 2 successes are 0.4096 and 0.2048 respectively. Find parameter p of the distribution.

9. The probability of failure of a certain component within one year is 0.40. Total six of these components are fitted on a machine. The machine will continue to be working if at least 4 out of the 6 components fitted on a machine survive. What is the probability that the machine will not fail during the one-year warranty period?

10. Let the probability of defective bolts be 0.1. Find the mean and standard deviation of defective bolts in a total lot of 400 bolts, if the defective bolts follow Binomial distribution.

11. Among 200 employees of a company, 160 are workers and 40 are managers. If 4 employees are to be chosen randomly to serve on staff welfare committee, find the probability that two of them will be workers and the other two managers using,

    (a) Hyper-geometric distribution.

    (b) Binomial distribution as an approximation.

    (c) What is the error of approximation?

12. In a quality control department of casting production division, 10 castings are randomly selected from each lot for destructive testing. If not more than 1 casting is defective, the lot is approved and sent to machining department. Otherwise, the lot is rejected and sent for recycling. What is the probability of rejecting the lot if the true population of defectives in the lot is 0.3?

13. In a factory past records show that on an average 3 workers are absent without leave per shift. Find the probability that in a shift,

    (a) Exactly two workers are absent

    (b) More than 4 workers will be absent.

    (c) At least 3 workers will be absent.

14. Multinomial random variable is a generalised version of .... random variable when the outcomes of the experiment are more than two.

15. Poisson introduced it for application of probability theory to .... suits.

14.10 Review Questions

1. It is known that screws produced by a certain machine will be defective with probability 0.01 independent of each other. The screws are sold in packages of 10 offered a money-back guarantee that at the most 1 of the 10 screws is defective. What proportion of packages sold would have to be replaced?

   (*Hint:* This is a binomial distribution with $n = 10$ and $p = 0.01$)

2. A purchaser of electronic components buys them in lots of size 10. He checks 3 randomly selected components from each lot and accepts the lot only if all these three are non-defective. If 30% of the lots received by him have 4 components defective and 70% have only 1 defective, what proportion of the lots will be rejected?

   (*Hint:* This is the case of Hyper Geometric Random Variable with $N = 10$, $n = 3$, $m_1 = 4$ for 30% lots Type A and $m_2 = 1$ for 70% lots type B.)

   P(reject)=1–P(accept)=1–P(accept|Type A)P(Type B) – P(accept|Type B)P(Type B)

3. Independent trials that result in success with probability p are performed until a total of r successes are obtained. Show that the probability that exactly n trials required is,

$$\binom{n-1}{r-1}p^r(1-p)^{n-r}$$

4. A random variable X has PMF $P(X) = pq^{x-1}$ *for* $x = 1, 2, 3, ...$

   Find

   (a) Mean

   (b) Variance of random variable X

5. When does multinomial probability distribution occur? State its probability distribution function.

6. Let X be a random variable with PMF $P(X) = q^{x-1}p$ *for* $x = 1, 2, 3, ...$ $0 < p < 1$ and $q = 1 - p$.

   Prove that $P(X > s + t | X > s) = P(X > t)$ for any $s, t > 0$.

7. If random variable X has PMF $P(x) = \binom{n}{x} q^{n-x} p^x$ Find mean and variance of X.

8. An experiment succeeds twice as many times as it fails. Find the chance that in 6 trials, there will be at least 5 successes.

9. Find the most probable number of heads, if a biased coin is tossed 100 times. Given the probability of head in every toss is 0.6

10. Two warring countries A and B have 4 bases each. When one wave of attack is carried out by both on each other, B is expected to destroy only one of the bases of A. However, due to superiority of weapons, A is expected to destroy two bases of B. Show that nevertheless, there may be possibility that A suffers more loss than B and find the probability of that happening. (Assume Binomial distribution for number of bases destroyed for each side).

11. The overall % of returns of certain items under warranty is 30. What is the probability that there would be maximum of 2 returns in 6 items sold?

12. Show that product of n Bernoulli (p) random variable is a Bernoulli random variable with parameter (p^n).

13. If X is a discrete uniform random variable distributed over 11, 12, ……, 20. Find,

    (a) $P(X > 15)$

    (b) $P(13 \leq X \leq 18)$

    (c) $P(X \leq 14)$

    (d) Mean
    (e) Standard Deviation

14. Let $X \to B\left(n = 8, p = \frac{1}{4}\right)$ Find

    (a) $P(X = 3)$

    (b) $P(X < 3)$

    (c) $P(X \leq 6)$

15. A radar system has a probability of 0.1 of detecting target at 60 km. range during a single scan. Find the probability that the target will be detected,

    (a) At least twice in four scans.

    (b) At most once in four scans.

16. Random variable *X* follows binomial distribution with mean = 10 and variance = 5. Find,

    (a) $P(X < 5)$

    (b) $P(2 < X < 10)$

    (c) $P(X \leq 10)$

17. A parcel of 12 books contains 4 books with loose binding. What is the probability that a random selection of 6 books (without replacement) will contain 3 books with loose binding?

18. If probability that individual suffers bad reaction from a particular injection is 0.001, determine the probability that out of 2000 patients, following patients will suffer bad reaction.

    (a) Exactly 3

    (b) More than 2

19. In a certain factory manufacturing bolts, there is 0.2% probability that any bolt can be defective. Bolts are supplied in packets of 10. Using Poisson distribution, calculate the approximate number of packets containing no defective, 1 defective, 2 defectives and 3 defectives bolts respectively in a consignment of 2000 packets.

20. On a highway 20 accidents take place in a span of 100 days. Assume that the number of accidents per day follow the Poisson distribution. Find the probability that there will be 3 or more accidents in a day.

21. Obtain approximation of Binomial distribution as Poisson distribution.

Answer: Self Assessment

1. First we find the probability of each IC to survive 150 hours, in other words, cumulative probability of upto 150 hours. Now, CDF of the random variable can be found out as,

$$F(a) = \int_{-\infty}^{a} f(x)dx = \int_{0}^{a} f(x)dx$$

Therefore, $F(150) = \int_{0}^{100} 0\,dx + \int_{100}^{150} \frac{100}{x^2}dx = \left[\frac{-100}{x}\right]_{100}^{150} = \left[\frac{100}{150} - \frac{100}{100}\right] = \frac{1}{3}$

Thus, probability of failure of IC within 150 hours is $\frac{1}{3}$. Now there are 5 IC$_s$ in equipment with the probability of failure of each is $\frac{1}{3}$.

Thus, it is a Binomial distribution with parameters, ($n = 5$, $p = \frac{1}{3}$).

(a) Probability that exactly 2 IC$_s$ will fail is,

$$P(X = 2) = \binom{5}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 = 0.329$$

(b) Probability of equipment survival for 150 hours

$$= \ (1 - \text{probability of equipment failure within 150 hours})$$

$$= \ 1 - [P(3) + P(4) + P(5)] = P(0) + P(1) + P(2)$$

Now, $P(0) \ = \ \binom{5}{0}\left(\dfrac{0}{3}\right)^0 \left(\dfrac{2}{3}\right)^5 = 0.132$

$$P(1) \ = \ \binom{5}{1}\left(\dfrac{1}{3}\right)^1 \left(\dfrac{2}{3}\right)^4 = 0.329$$

Probability of equipment survival for 150 hours = 0.132 + 0.329 + 0.329 = 0.79

2. (a) $P(X = i) = c(2)^{-i} = c\left(\dfrac{1}{2}\right)^i$

Comparing with geometric random variable with parameter $p = \dfrac{1}{2}$ we get $c = 1$

Or to be a PMF, $P(X = i)$ must satisfy Axiom II, $\displaystyle\sum_{i=1}^{n} P(X = i) = 1$

Hence $\displaystyle\sum_{i=1}^{n} P(X = i) \ = \ \sum_{i=1}^{\infty} c(2)^{-i} = \sum_{i=1}^{\infty} c\left(\dfrac{1}{2}\right)^i = c\left[\left(\dfrac{1}{2}\right)^1 + \left(\dfrac{1}{2}\right)^2 + \left(\dfrac{1}{2}\right)^3 + ...\right]$

$$= c\,\dfrac{\dfrac{1}{2}}{1 - \left(\dfrac{1}{2}\right)} = 1$$

Therefore, $c \ = \ 1$

Note that $0 \le P(X = i)$ for all values of i and Axiom II is satisfied. Therefore,

$0 \le P(X = i) \le 1$ Axiom I is satisfied.

Hence, the given expression is PMF

Note: Here we have used sum of geometric series,

$$r^0 + r^1 + r^2 + ... = \dfrac{1}{1 - r} \qquad \text{if } r < 1.$$

(b) $P(X = i) = \dfrac{c(2)^{-i}}{i}$

To be a PMF, $P(X = i)$ must satisfy Axiom II, $\displaystyle\sum_{i=1}^{n} P(X = i) = 1$

Hence, $\displaystyle\sum_{i=1}^{n} P(X = i) = \sum_{i=1}^{\infty} \dfrac{c(2)^{-i}}{i} = c\left[\dfrac{\left(\dfrac{1}{2}\right)^1}{1} + \dfrac{\left(\dfrac{1}{2}\right)^2}{2} + \dfrac{\left(\dfrac{1}{2}\right)^3}{3} + ...\right]$

$$= c\left[-\ln(1-\frac{1}{2})\right] = c \times \ln 2$$

Or, $c \times \ln 2 = 1$

Or, $c = \dfrac{1}{\ln 2}$

Note that $0 \le P(X = i)$ for all values of i and Axiom II is satisfied. Therefore,

$0 \le P(X = i) \le 1$ Axiom I is satisfied

Hence, the given expression is PMF

Note: Here we have used sum of logarithmic series,

$$\ln(1 + x) = \frac{x^1}{1} - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

(c) $P(X = i) = \dfrac{c(2)^i}{i!}$

To be a PMF, $P(X = i)$ must satisfy Axiom II, $\displaystyle\sum_{i=1}^{n} P(X = i) = 1$

Hence, $\displaystyle\sum_{i=1}^{n} P(X = i) = \sum_{i=1}^{\infty} \frac{c(2)^i}{i!} = c\left[\frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \dots\right] = 1$

Or, $c\left[\dfrac{2^0}{0!} + \dfrac{2^1}{1!} + \dfrac{2^2}{2!} + \dfrac{2^3}{3!} + \dots - 1\right] = c\left[e^2 - 1\right] = 1$

Or, $c = \dfrac{1}{e^2 - 1}$

Now that, $0 \le P(X = i)$ for all values of i and Axiom II is satisfied. Therefore,

$0 \le P(X = i) \le 1$

Hence, the given expression is PMF

Note: Here we have used sum of exponential series,

$$e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

(d) $P(X = i) = c(i)^{-2}$

To be a PMF, $P(X = i)$ must satisfy Axiom II, $\displaystyle\sum_{i=1}^{n} P(X = i) = 1$

Hence, $\displaystyle\sum_{i=1}^{n} P(X = i) = \sum_{i=1}^{\infty} c(i)^{-2} = c\left(\frac{1}{i}\right)^2 = c\left[\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \dots\right] = 1$

Or, $\qquad 2 \times c = 1$

Or, $\qquad c = \dfrac{1}{2}$

Now that $0 \le P(X = i)$ for all values of i and Axiom II is satisfied. Therefore,

$0 \le P(X = i) \le 1$. $\hfill$ Axiom I is satisfied

Hence, the given expression is PMF

Note: Here we have used sum of series,

$$\left[\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \dots \right] = 2$$

3. Generating function is given as,

$$M(t) \quad = \quad E(e^{tx}) = \sum_{m=0}^{n+m-1} e^{tm} \binom{n+m-1}{m} p^n (1-p)^m = \sum_{m=0}^{n+m-1} (qe^t) \binom{n+m-1}{m} (1-q)^n$$

$$= \quad (1-q) \sum_{m=0}^{n+m-1} (qe^t) \binom{n+m-1}{m} (1-q)^{n-1} = (1-q)[qe^t + (1-q)]^{m+n-1}$$

$$= \quad p[p + (1-p)e^t]^{m+n-1}$$

4. This is a sampling problem. Hence, probability of drawing defective item changes with every draw according to the previous draw (sampling without replacement). Therefore, number of defective items in sample is a Hyper-geometric random variable. Given lot size $N = 20$, probability of each item to be defective $p = 0.1$, sample size $n = 4$. Now probability of each item to be defective is 0.1 and total lot size is 20, hence, expected number of defective items in the lot of 20 i.e. $m = 20 \times 0.1 = 2$. Thus, number of defective items in sample is a Hyper-geometric random variable X with parameters ($m = 2$, $N = 20$, $n = 4$). The lot is accepted only if no item in the sample is defective i.e. $X = 0$. Its probability is,

$$P(X = 0) \quad = \quad \frac{\binom{m}{0}\binom{N-m}{n-0}}{\binom{N}{n}} = \frac{\binom{2}{0}\binom{20-2}{4-0}}{\binom{20}{4}} = \frac{\binom{18}{4}}{\binom{20}{4}} = 0.632$$

5. (a) $\qquad E(X) \quad = \quad \displaystyle\sum_{i=1}^{15} x_i P(x_i) = \frac{1}{15} + \frac{2}{15} + \frac{3}{15} + \dots + \frac{15}{15} = \frac{1}{15}(1 + 2 + 3 + \dots + 15)$

$$= \quad \frac{1}{15}\left(\frac{15 \times 16}{2}\right) = 8$$

(b) $\qquad E(X^2) \quad = \quad \displaystyle\sum_{i=1}^{15} x_i^2 P(x_i) = \frac{1^2}{15} + \frac{2^2}{15} + \frac{3^2}{15} + \dots + \frac{15^2}{15} = \frac{1}{15}(1^2 + 2^2 + 3^2 + \dots + 15^2)$

$$= \quad \frac{1}{15}\left(\frac{15 \times 16 \times 31}{6}\right) = \frac{8 \times 31}{3}$$

Now variance is,

$$Var(X) \quad = \quad E(X^2) - [E(X)]^2 = \frac{8 \times 31}{3} - 8^2 = \frac{56}{3} = 18.66$$

Alternatively

We observe that the given PMF is of **Uniform Distribution** with $n = 15$. Therefore, using standard results,

(a) $E(X) = \frac{n+1}{2} = \frac{15+1}{2} = 8$

(b) $Var(X) = \frac{n^2 - 1}{12} = \frac{15^2 - 1}{12} = \frac{224}{12} = 18.66$

6. (a) To be a PMF, $P(x)$ must satisfy the axioms of probability. For Axiom II, $\sum_{x=1}^{\infty} P(x) = 1$

Thus,

$$\sum_{x=1}^{\infty} P(x) \quad = \quad \sum_{x=1}^{\infty} Cp^x = C\{p^1 + p^2 + p^3 + ...\} = 1$$

Now using sum of infinite G.P. series with first term as p and common ratio as p, also taking $0 \le p \le 1$, we get,

$$C\left(\frac{p}{1-p}\right) \quad = \quad 1 \qquad\qquad \text{Or,}$$

$$C \quad = \quad \frac{1-p}{p}$$

Substituting the value of C, the PMF becomes, $P(x) = (1-p)p^{x-1} \quad for \quad x = 1,2,3,.....$

Note that for $0 \le p \le 1$, for any value of x, $0 \le P(x) \le 1$. Thus, Axiom I also satisfied.

(b) The given PMF is of Geometric random variable with probability of failure p and probability of success $(1 - p)$. The solution is,

$$E(X) = \frac{1}{1-p}$$

7. This is a multinomial distribution. *For solution, refer to section 14.6.*

8. For binomial distribution probability of i successes in n independent trials is given by,

$$P(X = i) \quad = \quad \binom{n}{i} p^i (1-p)^{n-i}$$

Now, with $n = 5$ and given values of i, we get,

$$P(X = 1) \quad = \quad \binom{5}{1} p^1 (1-p)^4 = 5p(1-p)^4 = 0.4096 \qquad\qquad ...(1)$$

$$P(X = 2) = \binom{5}{2} p^2 (1-p)^3 = 10 p^2 (1-p)^3 = 0.2048 \qquad \text{...(2)}$$

Dividing (2) by (1) equations,

$$\frac{10 p^2 (1-p)^3}{5 p (1-p)^4} = \frac{2p}{(1-p)} = \frac{0.2048}{0.4096} = \frac{1}{2}$$

$$p = \frac{1}{5}$$

9. This is a case of Binomial distribution. Let q be the probability of failure and p the probability of success. Number of trials is n and i is number of components survive. As per the given data, $q = 0.4$ Hence, $p = 1 - q = 0.6$ Also, $n = 6$

    Now probability that at least 4 components will survive is,

    $$P(4) + P(5) + P(6) = \binom{6}{4}(0.6)^4(0.4)^2 + \binom{6}{5}(0.6)^5(0.4)^1 + \binom{6}{6}(0.6)^6(0.4)^0$$

    $$= 0.5443$$

10. Given is $p = 0.1$ and $n = 400$. Also $q = (1-p) = 0.9$

    $$Mean = 400 \times 0.1 = 40$$

    $$\text{Standard Deviation} = \sqrt{npq} = \sqrt{400 \times 0.1 \times 0.9} = 6$$

11. Let X be the random variable denoting number of workers selected in the committee.

    (a) The given data is Total employees $N = 200$, Number of workers $m = 160$, Size of the committee $n = 4$. Thus, if X follows Hyper-geometric distribution,

    $$X \to H(n = 4, N = 200, m = 160) \quad \text{Hence,}$$

    $$P(X = 2) = \frac{\binom{m}{2}\binom{N-m}{2}}{\binom{N}{n}} = \frac{\binom{160}{2}\binom{40}{2}}{\binom{200}{4}} = 0.1534$$

    (b) For binomial distribution parameter $p = \dfrac{m}{N} = 0.8$ Thus, if X follows Binomial distribution,

    $$X \to B(n = 4, p = 0.8) \quad \text{Hence,}$$

    $$P(X = 2) = \binom{4}{2}(0.8)^2(0.2)^2 = 0.1536$$

    (c) $Error = P(X = 2)_{Hyper-geometric} - P(X = 2)_{Binomial} = -0.002$

12. Let X denote the number of defective castings in the randomly selected sample. Therefore,

    $$X \to B(n = 10, p = 0.3)$$

The lot is accepted if only 0 or 1 casting is found to be defective.   **Notes**

$$P(Accepting) = P(X = 0) + P(X = 1) = \binom{n}{0}(p)^0(1-p)^n + \binom{n}{1}(p)^1(1-p)^{n-1}$$

$$= \binom{10}{0}(0.3)^0(0.7)^{10} + \binom{10}{1}(0.3)^1(0.7)^9 = 0.1493$$

Therefore,

$$\text{P } (Rejecting) = P(rejecting) = 1 - P(Accepting) = 1 - 0.1493 = 0.8507$$

13. This is a case of a Poisson distribution with parameter (l = 3). Now,

    (a) $P(X = 2) = \dfrac{e^{-\lambda}\lambda^2}{2!} = 0.2241$

    (b) $P(X > 4) = 1 - P(0) - P(1) - P(2) - P(3) - P(4) = 0.1845$

    (c) $P(X \geq 3) = 1 - P(0) - P(1) - P(2) - P(3) = 0.5767$

14. Binomial

15. Law

14.11 Further Readings

Books   Béla Bollobás, *Modern graph theory*, Springer

Martin Charles Golumbic, Irith Ben-Arroyo Hartman, *Graph theory, Combinatorics, and Algorithms*, Birkhäuser

Online links   http://en.wikipedia.org/wiki/Poisson_distribution

http://en.wikipedia.org/wiki/Binomial_distribution

http://en.wikipedia.org/wiki/Geometric_distribution