Applied Mathematics-II (BIT-203-CR)

Lecture Notes

BIT-203-CR *S.P. College*

UNIT-I (Probability)

Probability is a measure of uncertainty and deals with the phenomenon of chance or randomness. The theory of probability has its origins in gambling and games of chances. It is known that a French gambler approached a French mathematician Blais Pascal for a solution of a problem concerning gambling. Pascal gave a solution and subsequently he corresponded with another French Pierre de Fermat and established the foundations of the theory of probability.

In this section we first define certain important elementary terms related to probability.
Random Experiments:
A random experiment is an experiment in which

(1) The experiment can be repeated any number of times under identical conditions.

(2) All possible outcomes of the experiment are known in advance.

(3) The actual outcome in a particular experiment is not known in advance.

Examples of Random Experiment:

(i) Rolling of an unbiased die.

(ii) Tossing of an unbiased Coin.

Sample Space: A set of all possible outcomes in a random experiment is called sample space . For example; In rolling an unbiased die the sample space is $\{1, 2, 3, 4, 5, 6\}$ and in tossing an unbiased Coin the sample space is $\{\text{Head}, \text{Tail}\}$ or $\{H, T\}$. Similarly, in tossing two unbiased coins simultaneously the sample space is $\{HH, HT, TH, TT\}$.

Event: Any subset of sample space is called an event. For example, since $\{2, 4, 6\} \subset \{1, 2, 3, 4, 5, 6\}$, therefore $\{2, 4, 6\}$ is an event in the random experiment of rolling of an unbiased die.

Exhaustive Events: Two or more events are said to be exhaustive if their union is whole sample space. For example in tossing 2 coins the events $\{HT, TH, HH\}$ and $\{TT, TH, HT\}$ are mutually exhaustive.

Mutually Exclusive events: Two events A and B are said to be mutually exclusive if the happening of A excludes the happening of B, i.e., they are disjoint, $A \cap B = \varnothing$. For example in rolling a die the events $A = \{2, 4\}$ and $B = \{3, 5\}$ are mutually exclusive because $A \cap B = \varnothing$.

Classical definition of probability: Suppose in a random experiment there are n possible exhaustive and equally likely outcomes out of which m favors the happening of an event A, then probability of A is defined as

$$P(A) := \frac{m}{n} = \frac{\text{Number of outcomes favorable to event } A}{\text{Total number of possible outcomes}} = \frac{n(A)}{n(S)}$$

where S is a sample space. One can easily note that $0 \leq P(A) \leq 1$.

Example Suppose that a bag contains 6 red, 5 black and 4 blue balls. Find the probability that three balls drawn simultaneously are one blue, one black and one red.

Solution. Total number of balls is $6+5+4=15$. Out of these 15 balls, 3 balls can be drawn simultaneously in $\binom{15}{3}$ ways. Therefore, 3 balls can be drawn simultaneously in

$$\binom{15}{3} = \frac{15!}{(15-3)!3!} = \frac{15.14.13}{1.2.3} = 455 \text{ ways.}$$

One red ball out of 6 can be drawn in $\binom{6}{1} = 6$ ways. Similarly one black ball and one blue ball can be drawn respectively in 5 and 4 ways. By Counting principle one red, one black and one blue ball can be drawn simultaneously in $6.5.4 = 120$ ways. Thus

$$\text{Required probability} = \frac{120}{455} = \frac{24}{91}.$$

Exercises

(1) Suppose an integer is picked from among 1 to 20. What is probability of picking a prime.

(2) Suppose that 2 dice are thrown. Find probability that none of the dice shows 3.

(3) Out of 20 consecutive positive integers two are drawn at random. Find probability that their sum is odd.

Theorem 1.1 *Let S be a sample space in a random experiment with $n(S)$ finite, then*

(i) $P(\varnothing) = 0$ *and* $P(S) = 1$.

(ii) If events A and B are such that $A \subset B$, then $P(A) \leq P(B)$.

(iii) For any event A, $0 \leq P(A) \leq 1$.

(iv) For any two event A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(v) $P(A) = 1 - P(A')$.

Proof: (i)

$$P(\varnothing) = \frac{n(\varnothing)}{n(S)} = 0 \quad \text{and} \quad P(S) = \frac{n(S)}{n(S)} = 1.$$

(ii) If $A \subset B$, then $n(A) \leq n(B)$, therefore

$$P(A) = \frac{n(A)}{n(S)} \leq \frac{n(B)}{n(S)} = P(B).$$

(iii) For any event A, we have $\varnothing \subset A \subset S$. By (ii) we get $P(\varnothing) \leq P(A) \leq P(S)$ and by using (i) we obtain $0 \leq P(A) \leq 1$.
(iv) By set inclusion-exclusion principle

$$n(A \cup B) = n(A) + n(B) - n(A \cap B).$$

This implies by dividing both sides by $n(S)$

$$\frac{n(A\cup B)}{n(S)}=\frac{n(A)}{n(S)}+\frac{n(B)}{n(S)}-\frac{n(A\cap B)}{n(S)},\quad \text{or}\quad P(A\cup B)=P(A)+P(B)-P(A\cap B).$$

(v) Since $A\cup A'=S$ then by (iv), we have

$$P(S)=P(A\cup A')=P(A)+P(A')-P(A\cap A')\quad \text{or}\quad 1=P(A)+P(A')-0$$
$$\Rightarrow P(A)=1-P(A')$$

■

Exercises

1. Let A, B and C are three events. Show that

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

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

(iii) If A, B and C are mutually exclusive events then $P(A\cup B\cup C)=P(A)+P(B)+P(C)$

(iv) $P(A'\cap B')=1-P(A\cup B)=1-P(A)-P(B)+P(A\cap B)$.

(v) $P(A'\cup B')=1-P(A\cap B)$.

2. Find the probability of drawing an ace or a spade from a well-shuffled pack of 52 cards.
3. Let A, B and C are three newspapers from a city. 25% of the population read A, 20% read B, 15% read C, 16% read both A and B, 10% read both B and C, 8% read both A and C and 4% read all the three. A person is selected what is probability that he reads atleast one of the news papers.

Conditional Probability

Given two events A and B with $P(B)>0$, the *conditional probability* of A given B is defined as the quotient of the probability of the joint of events A and B, and the probability of B, i.e.,

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

Similarly, the conditional probability of B given A is defined as

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

Example 1. In a card game, suppose a player needs to draw two cards of the same suit in order to win. Of the 52 cards, there are 13 cards in each suit. Suppose first the player draws a heart. Now the player wishes to draw a second heart. Since one heart has already been chosen, there are now 12 hearts remaining in a deck of 51 cards. So the conditional probability $P(\text{Draw second heart}|\text{First card a heart})=12/51$.

2. In a school of 1200 students, 250 are seniors, 150 students take math, and 40 students are seniors and are also taking math. What is the probability that a randomly chosen student who is a senior, is taking math?
Solution. These questions can be confusing. It sounds, at first read, that they are asking for the probability

of choosing a student who is a senior and who is taking math. Not quite right! It helps to re-word the question into:

Find the probability that the student is taking math, given that the student is a senior.
B = the student is taking math
$n(A)$ = the student is a senior = 250
$n(A \cap B)$ = the student is a senior and is taking math = 40.

$$P(A \cap B) = \frac{40}{1200} \quad P(A) = \frac{250}{1200}.$$

Thus,

$$P(B|A) = \frac{40/1200}{250/1200} = \frac{4}{25}$$

Exercises

1. A math teacher gave her class two tests. 25% of the class passed both tests and 42% of the class passed the first test. What percent of those who passed the first test also passed the second test?
2. A bag contains 12 red balls, 12 blue balls, and 12 green balls. What is the probability of drawing two balls of the same color in a row?

Multiplication theorem on Conditional probability

Theorem. Let A and B are two events of a random experiment such that $P(A) > 0$. Then $P(A \cap B) = P(A)P(A|B)$.
Proof: This is an immediate consequence of the definition of the conditional probability $P(A|B)$. ■

The above theorem can be extended to any finite number of events by using principle of mathematical induction.

Exercise

Let A, B and C are events such that $P(A \cap B \cap C) > 0$ then $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$.

Examples

1. A bag contains 20 identical balls of which 8 are black and 12 are blue. Three balls are taken out at random from the bag one after the other without replacement. Find the probability that all the three balls drawn are blue.
Solution. Let A : the event that first ball is blue, B : the event that second ball is blue, C : the event that third ball is blue. The probability that the first ball drawn is blue is 12/20, i.e., $P(A) = 12/20$. Since there are 12 blue balls among 20 balls in the bag. If the first ball is blue, then the probability that the second ball drawn is blue is 11/19, i.e., $P(B|A) = 11/19$, since 11 of the remaining 19 ae blue. Similarly, if the first two balls drawn are blue, then the probability that third ball drawn is blue is 10/18, i.e., $P(C|A \cap B) = 10/18$. By multiplication theorem on conditional probability, the probability that all the three balls drawn are blue is given as:

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B) = \frac{12}{20}.\frac{11}{19}.\frac{10}{18} = \frac{11}{57}.$$

1.0.1 Independent events

Two events, A and B, are independent if the fact that A occurs does not affect the probability of B occurring. Some examples of independent events are:

- Landing on heads after tossing a coin AND rolling a 5 on a single 6-sided die.
- Choosing a marble from a jar AND landing on heads after tossing a coin.
- Choosing a 3 from a deck of cards, replacing it, AND then choosing an ace as the second card.
- Rolling a 4 on a single 6-sided die, AND then rolling a 1 on a second roll of the die.

If A and B are independent then occurrence of A does not affect the occurrence of B, this implies that $P(B|A) = P(B)$. Thus if A and B are independent then by multiplication rule on conditional probability,

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

This is also considered as the definition of independent events. Hence, two events A and B are independent iff $P(A \cap B) = P(A)P(B)$, i.e., two events are independent if the probability of their simultaneous occurrence is equal to the product of their probabilities.

Examples

1. A coin is tossed and a single 6-sided die is rolled. Find the probability of landing on the head side of the coin and rolling a 3 on the die.
Solution. Let A : head on coin and B : 3 on the die are two events. Clearly A and B are independent as the outcome on coin does not affect the outcome on the die.

$$\text{Now,} \quad P(A) = \frac{1}{2} \quad \text{and} \quad P(B) = \frac{1}{6}.$$

Thus,

$$P(A \cap B) = P(A)P(B) = \frac{1}{2}.\frac{1}{6} = \frac{1}{12}.$$

2. A card is chosen at random from a deck of 52 cards. It is then replaced and a second card is chosen. What is the probability of choosing a jack and then an eight?
Solution. Let A : first card is jack and B : second card is an eight are two events. Clearly A and B are independent as the card is replaced before drawing the second card. Here $P(A) = 4/52$ and $P(B) = 4/52$. Thus

$$P(A \cap B) = P(A)P(B) = \frac{4}{52}\frac{4}{52} = \frac{1}{169}.$$

Exercises

1. A jar contains 6 red balls, 3 green balls, 5 white balls and 7 yellow balls. Two balls are chosen from the jar, with replacement. What is the probability that both balls chosen are green?

2. If A and B are two independent events then show that

   (i) A' and B' are also independent.

(ii) A' and B are also independent.

Partition of a sample space A collection of sets $B_1, B_2, \ldots, B_n$ is said to partition the sample space if the sets (i) are mutually disjoint (ii) mutually exhaustive and (iii) each have non-zero probability. A simple example of a partition is given by a set B, together with its complement B.

Law of total probability

Let $B_1, B_2, \ldots, B_n$ constitute a partition of a sample space and let A be any event, then

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \ldots + P(B_n)P(A|B_n) = \sum_{j=0}^{n} P(B_j)P(A|B_j).$$

Proof: Since $B_1, B_2, \ldots, B_n$ partition the sample space S (say) therefore they are mutually exclusive and exhaustive events. Thus $\bigcup_{j=1}^{n} B_j = S$ and by distributive law, $A = A \cap S = A \cap \left(\bigcup_{j=1}^{n} B_j\right) = \bigcup_{j=1}^{n} (A \cap B_j)$. Again, since $B_1, B_2, \ldots, B_n$ are mutually exclusive therefore $A \cap B_1, A \cap B_2, \ldots, A \cap B_n$ are also mutually exclusive as $A \cap B_j$ is a part/subset of B_j, $j = 1, 2, \ldots, n$. Therefore by multiplication rule of total probability,

$$P(A) = P\left(\bigcup_{j=1}^{n} (A \cap B_j)\right) = \sum_{j=1}^{n} P(A \cap B_j) = \sum_{j=0}^{n} P(B_j)P(A|B_j).$$

■

Bayes' theorem

Let $B_1, B_2, \ldots, B_n$ partition the sample space S and A be any event with non-zero probability, then for each $k = 1, 2, \ldots, n$,

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \ldots + P(B_n)P(A|B_n)} = \frac{P(B_k)P(A|B_k)}{\sum_{j=0}^{n} P(B_j)P(A|B_j)}.$$

Proof: Since $P(A) > 0$ then by multiplication rule on conditional probability and theorem of total probability, we have for each $k = 1, 2, \ldots, n$,

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(B_k)P(A|B_k)}{\sum_{j=0}^{n} P(B_j)P(A|B_j)}.$$

This completes the proof the Bayes' theorem. ■

When to apply Bayes' theorem?

Part of the challenge in applying Bayes' theorem involves recognizing the types of problems that warrant its use. You should consider Bayes' theorem when the following conditions exist.

* The sample space is partitioned into a set of mutually exclusive events $\{B_1, B_2, \ldots, B_n\}$.
* Within the sample space, there exists an event A, for which $P(A) > 0$.
* The analytical goal is to compute a conditional probability of the form: $P(B_k|A)$.
* You know at least one of the two sets of probabilities $P(B_k \cap A)$ for each B_k **OR** $P(B_k)$ and $P(A|B_k)$ for each B_k.

Example

Three bags b_1, b_2 and b_3 contain balls as given in the following table

	Red	White	Black
b_1	2	2	1
b_2	4	3	2
b_3	2	4	3

A die is thrown. b_1 is chosen if either 1 or 2 turns up, b_2 is chosen if either 3 or 4 turns up and b_3 is chosen if either 5 or 6 turns up. Having chosen a bag in this way, a ball is chosen at random from this bag. If the ball chosen is of red colour, find the probability that it comes from bag b_2?

Solution. Let B_j be the event of choosing the bag b_j for $j = 1, 2, 3$. Let A be the event of choosing a red ball. Then $P(B_j) = 1/3$ for $j = 1, 2, 3$. Having chosen the bag b_j, the probability of choosing a red ball is $P(A|B_j)$ and is given by

$$P(A|B_1) = \frac{2}{5}, \quad P(A|B_2) = \frac{4}{9}, \quad P(A|B_3) = \frac{2}{9}.$$

We want to find the probability $P(B_2|A)$. By Bayes' theorem, we get

$$P(B_2|A) = \frac{P(B_2)P(A|B_2)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)} = \frac{\frac{1}{4} \times \frac{4}{9}}{\left(\frac{1}{3} \times \frac{2}{5}\right) + \left(\frac{1}{3} \times \frac{4}{9}\right) + \left(\frac{1}{3} \times \frac{2}{9}\right)} = \frac{5}{12}.$$

2. **Monty Hall problem** Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say number 3, which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors, i.e., does probability of winning a car increases or remains same by switching your choice of doors?

Solution. Let B_1, B_2, B_3 denote the events "the car is behind door number 1", "the car is behind the door number 2", "the car is behine the door number 3." Let also A denote the event of Monty opening door number 3. Now

$P(A|B_1) = \frac{1}{2}$ because Monty has to choose between two carless doors, 2 and 3

$P(A|B_2) = 1$ because Monty never opens the door with a car behind, therefore he will open door 3

$P(A|B_3) = 0$ for the very same reason that$P(A|B_2) = 1$.

The events B_1, B_2 and B_3 are mutually exclusive and exhaustive, i.e., B_1, B_2 and B_3 partition the sample space, also

$$P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) = \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 1 + \frac{1}{3} \times 0 = \frac{1}{2}.$$

Now you are given a chance to switch to another door, 1 or 2 (depending on which one remains closed.) If you stick with your original selection (1), by Bayes' theorem,

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)} = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{2}} = \frac{1}{3}.$$

However, if you switch,

$$P(B_2|A) = \frac{P(B_2)P(A|B_2)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)} = \frac{\frac{1}{3} \times 1}{\frac{1}{2}} = \frac{2}{3}.$$

Summing up, if you do not switch your chance of winning is 1/3 whereas if you do switch your chance of winning is 2/3. You'd be remiss not to switch!

Exercises

1. Assume that in a family, each children is equally likely to be a boy or a girl. A family with three children is chosen at random. Find the probability that the eldest child is a girl given that the family has at least one girl.

2. A shopkeeper sells three types of flower seeds A_1, A_2 and A_3. They are sold as a mixture where the proportions are $4:4:2$ respectively. The germination rates of the three types of seeds are 45%, 60% and 35% respectively. Calculate the probability

(i) of a randomly chosen seed to germinate

(ii) that it will not germinate given that the seed is of type A_3

(iii) that it is of the type A_3 given that a randomly chosen seed does not germinate.

1.1 UNIT-II (Statistics)

Measures of Central Tendency

A measure of central tendency is a single value that describes the way in which a group of data cluster around a central value, i.e., A score that indicates where the center of the distribution tends to be located. To put in other words, it is a way to describe the center of a data set. There are three measures of central tendency: the mean, the median, and the mode.

Arithmetic mean

The arithmetic mean of a set of data is found by taking the sum of the data, and then dividing the sum by the total number of values in the set, i.e., sum of observations divided by number of observations. Let $x_1, x_2, \ldots, x_n$ be n observations of a data then their arithmetic mean is given by:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

For example, let us consider the monthly salary of 10 employees of a firm: 2500, 2700, 2400, 2300, 2550, 2650, 2750, 2450, 2600, 2400. The arithmetic mean is

$$\bar{x} = \frac{2500 + 2700 + 2400 + 2300 + 2550 + 2650 + 2750 + 2450 + 2600 + 2400}{10} = \frac{25300}{10} = 2530.$$

Mean of a Frequency distribution

The **frequency** is the number of times a particular data point occurs in the set of data. A frequency distribution is a table that list each data point and its frequency. Suppose we have the following (ungrouped) frequency distribution of the data:

observation	x_1	x_2	$\cdots$	x_n
frequency	f_1	f_2	$\cdots$	f_n

The observation x_i is repeated f_i, $i = 1, 2, \ldots, n$ times in the data. Thus, sum of the observations is $f_1x_1 + f_2x_2 + \ldots + f_nx_n$ and number of observations is $f_1 + f_2 + \ldots + f_n$. Thus, mean is

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$$

Example

Find the arithmetic mean of the following frequency distribution:

x :	1	2	3	4	5	6	7
f :	5	9	12	17	14	10	6

Solution.

x_i	f_i	x_if_i
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
	$\sum f_i = 73$	$\sum f_ix_i = 299$

$$\therefore \quad \bar{x} = \frac{\sum f_ix_i}{\sum f_i} = \frac{299}{73} = 4.09.$$

Mean of Grouped frequency distribution: In a grouped frequency distribution, data is sorted and separated into groups called *class-intervals*. To find arithmetic mean in case of a grouped (continuous) frequency distribution, x is taken as the mid-value of the corresponding class. This is described in the following example.

Example

Calculate the arithmetic mean of the marks from the following table:

Marks :	$0-10$	$10-20$	$20-30$	$30-40$	$40-50$	$50-60$
No. of students :	12	18	27	20	17	6

Solution.

Marks	f_i	Mid-point x_i	x_if_i
$0-10$	12	5	60
$10-20$	18	18	270
$20-30$	27	36	675
$30-40$	20	68	700
$40-50$	17	70	765
$50-60$	6	60	330
	$\sum f_i = 100$		$\sum f_ix_i = 2800$

$$\therefore \quad \bar{x} = \frac{\sum f_ix_i}{\sum f_i} = \frac{2800}{100} = 28.$$

Properties of Arithmetic mean

Property-I Alegbraic sum of the deviations of a set of values from their arithmetic mean is zero.

Proof: Let $x_i|f_i$, $i = 1, 2, \ldots, n$ be the frequency distribution, then sum of deviations of the set of values from mean is given as:

$$\sum f_i(x_i - \bar{x}) = \sum f_i x_i - \sum \bar{x} f_i = \sum f_i x_i - \bar{x} \sum f_i.$$

Since $\bar{x} = \dfrac{\sum f_i x_i}{\sum f_i} \Rightarrow \bar{x} \sum f_i = \sum f_i x_i$. Hence,

$$\sum f_i(x_i - \bar{x}) = \sum f_i x_i - \sum f_i x_i = 0.$$

■

Property-II The sum of the squares of the deviations of a set of values is minimum when taken about mean.

Proof: Let $x_i|f_i$, $i = 1, 2, \ldots, n$ be the frequency distribution, let

$$z = \sum f_i(x_i - A)^2$$

be the sum if the squares of the deviations of given values from any arbitrary point A. We have to prove z is minimum when $A = \bar{x}$. We prove it by the second derivative test. That is, z will be minimum for variations in A if

$$\frac{\partial z}{\partial A} = 0, \qquad \frac{\partial^2 z}{\partial A^2} > 0.$$

Now,

$$\frac{\partial z}{\partial A} = -2 \sum f_i(x_i - A) = 0 \quad \Rightarrow \quad \sum (f_i - A) = 0 \quad \Rightarrow \quad \sum f_i(x_i - A) = 0 \ \text{ or } \ A = \frac{\sum f_i x_i}{\sum f_i} = \bar{x}.$$

Again,

$$\frac{\partial^2 z}{\partial A^2} = -2 \sum f_i(-1) = 2 \sum f_i > 0.$$

Hence z is minimum at the point $A = \bar{x}$. ■

Median

Median of a distribution is the value of the variable which divides it into equal parts. It is the value which exceeds and is exceeded by the same number of observations, i.e., it is the value such that the number of observations above it equal to the number of observations below it. The median is thus a positional average.

The median of a finite list of numbers can be found by arranging all the numbers from smallest to greatest. If there is an odd number of numbers, the middle one is picked. For example, consider the set of numbers:

$$1, 3, 3, 6, 7, 8, 9$$

This set contains seven numbers. The median is the fourth of them, which is 6.

If there are an even number of observations, then there is no single middle value; the median is then usually defined to be the mean of the two middle values. For example, in the data set:

$$1, 2, 3, 4, 5, 6, 8, 9$$

The median is the mean of the middle two numbers: this is $(4 + 5) \div 2$, which is 4.5.

Median of (ungrouped) frequency distribution

The cumulative frequency is the total of a frequency and all frequencies so far in a frequency distribution. Cumulative frequency is defined as a running total of frequencies. The frequency of an element in a set refers to how many of that element there are in the set. Cumulative frequency can also defined as the sum of all previous frequencies up to the current point. In case of an (ungrouped) frequency distribution median is obtained by considering the cumulative frequencies.

The steps for calculating median are given below:

(i) Find $N/2$ where $N = \sum_i f_i$.

(ii) See the (less than) cumulative frequency ($c.f.$) just greater than $N/2$.

(iii) The corresponding value of x is median.

Example

Find the median of the following frequency distribution:

x :	1	2	3	4	5	6	7
f :	5	9	12	17	14	10	6

Solution.

x	f	$c.f.$
1	5	5
2	9	14
3	12	26
4	17	43←
5	14	57
6	10	67
7	6	73
	$N = \sum f_i = 73$	

$\therefore$ Hence $N = 73 \Rightarrow N/2 = 36.5$.

Cumulative frequency ($c.f.$) just greater than $N/2$ is 43 and the value of x corresponding to 43 is 4. Therefore, median is 4.

Median of (grouped) frequency distribution

In the case of grouped (continuous) frequency distribution, the class corresponding to the $c.f.$ just greater than $N/2$ is called the *median class* and the value of median is obtained by the following formula:

$$\text{Median} = \ell + \frac{h}{f}\left(\frac{N}{2} - c\right), \quad \text{where}$$

- ℓ is the lower limit of the median class.
- f is the frequency of the median class.
- h is the length of the median class.
- c is the cumulative frequency of the class preceding the median class and $N = \sum f_i$.

Example

Calculate the median of the marks from the following table:

Marks :	$0-10$	$10-20$	$20-30$	$30-40$	$40-50$	$50-60$
No. of students :	12	18	27	20	17	6

Solution.

Marks	f_i	$c.f.$
$0-10$	12	12
$10-20$	18	30
$20-30$	27	57←
$30-40$	20	77
$40-50$	17	94
$50-60$	6	100
		$N = \sum f_i = 100$

Here $N/2 = 50$, Median class is $20-30$, $\ell = 20$ $f = 27$, $h = 10$ and $c = 30$.

$$\text{Median} = \ell + \frac{h}{f}\left(\frac{N}{2} - c\right) = 20 + \frac{10}{27}(50 - 30) \approx 27.4$$

Mode

Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster densely. Thus in the case of discrete frequency distribution mode is the value of x corresponding to maximum frequency. For example, in the following frequency distribution:

x	:	1	2	3	4	5	6	7	8
f	:	4	9	16	25	22	15	7	3

the value of x corresponding to the maximum frequency, viz., 25 is 4. Hence mode is 4. We can have more than one mode. That is their can be more than one observations having maximum frequency. If a distribution have two modes then it is called 'bimodal' distribution and if the distribution have more than two modes then it is called 'multimodal' distribution.

Mode of grouped frequency distribution

In case of grouped (continuous) frequency distribution, the class-interval with maximum frequency is called *modal class* and mode is given by the formula:

$$\text{Mode} = \ell + \frac{h(f_1 - f_0)}{2f_1 - f_0 - f_2} \quad \text{where}$$

- ℓ is the lower limit of modal class
- h is the length of modal class
- f_1 is the frequency of the modal class
- f_0 and f_2 are the frequencies of the classes preceding and succeeding the modal class respectively.

Example

Calculate the mode of the marks from the following table:

Marks :	$0-10$	$10-20$	$20-30$	$30-40$	$40-50$	$50-60$
No. of students :	12	18	27	20	17	6

Solution. Here maximum frequency is 27. Thus the class $20-30$ is the modal class. The mode is given as:

$$\text{Mode} = \ell + \frac{h(f_1 - f_0)}{2f_1 - f_0 - f_2} = 20 + \frac{10(27-18)}{2(27) - 18 - 20} = 20 + \frac{90}{16} = 25.62(\text{approx.})$$

Exercise

The median and mode of the following wage distribution are known to be ₹ 33.5 and ₹ 34 respectively. Find the values of f_3, f_4 and f_5.

Wages (in ₹)	$0-10$	$10-20$	$20-30$	$30-40$	$40-50$	$50-60$	$60-70$	Total
Frequency	4	16	f_3	f_4	f_5	6	4	230

Measure of dispersion

The measure of central tendency give us an idea of the concentration of the observations about the central part of the distribution. If we know the average alone we cannot form a complete idea about the distribution. In other words, they fail to reveal the degree of the spread out or the extent of the variability in individual items of the distribution. This can be explained by certain other measures, known as 'Measures of Dispersion' or Variation. Simplest meaning that can be attached to the word 'dispersion' is a lack of uniformity in the sizes or quantities of the items of a group or series. The word dispersion may also be used to indicate the spread of the data. In all these definitions, we can find the basic property of dispersion as a value that indicates the extent to which all other values are dispersed about the central value in a particular distribution

Range

The simplest of our methods for measuring dispersion is range. Range is the difference between the largest value and the smallest value in the data set. While being simple to compute, the range is often unreliable as a measure of dispersion since it is based on only two values in the set. For example the range of $13, 21, 26, 32, 41, 49, 53$ is $53 - 13 = 40$. A range of 40 tells us very little about how the values are dispersed. Are the values all clustered to one end with the low value (13) or the high value (53) being an outlier? Or are the values more evenly dispersed among the range?

Range is the simplest but crude measure of dispersion. Since it is based on two extreme observations. It is not at all a reliable measure of dispersion.

Mean deviation

If $x_i|f_i$, $i = 1, 2, \ldots, n$ is the frequency distribution then mean deviation from average A, (usually mean, median or mode) is given by

$$\text{Mean deviation} = \frac{\sum_i f_i|x_i - A|}{N}, \quad \text{where} \quad N = \sum_i f_i$$

where $|x_i - A|$ represents the modulus of the deviation $(x_i - A)$.

Example

Find the mean deviation about mean of the following frequency distribution:

x :	1	2	3	4	5	6	7
f :	5	9	12	17	14	10	6

Solution.

x_i	f_i	$x_i f_i$	$\lvert x_i - \bar{x}\rvert$	$f_i\lvert x_1 - \bar{x}\rvert$
1	5	5	3.09	15.45
2	9	18	2.09	18.81
3	12	36	1.09	13.08
4	17	68	0.09	01.53
5	14	70	0.91	12.74
6	10	60	1.91	19.1
7	6	42	2.91	17.46
Total	$N = 73$	299		98.17

$$\therefore \quad \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{299}{73} = 4.09.$$

Thus mean deviation about mean is given as:

$$\text{Mean deviation} = \frac{\sum_i f_i|x_i - A|}{N} = \frac{98.17}{73} \approx 1.344$$

Standard deviation & variance

The standard deviation, which is shown by greek letter σ (read as sigma) is extremely useful in judging the representativeness of the mean. The concept of standard deviation, which was introduced by Karl Pearson has a practical significance because it is free from all defects, which exists in a range or mean deviation.

Standard deviation is calculated as the square root of average of squared deviations taken from actual mean. It is also called root mean square deviation. For the frequency distribution $x_i|f_i$, $i = 1, 2, \ldots, n$ standard deviation given as:

$$\sigma = \sqrt{\frac{\sum_i f_i(x_i - \bar{x})^2}{N}} \quad \text{where} \quad \bar{x} \ \text{ is the arithmetic mean and } \ N = \sum_i f_i.$$

The square of standard deviation i.e., σ^2 is called 'variance'. Thus variance is given as:

$$\sigma^2 = \frac{\sum_i f_i(x_i - \bar{x})^2}{N}.$$

Exercises

1. Calculate the mean, standard deviation and variance for the following table giving age distribution of 542 members.

Age in years	$20-30$	$30-40$	$40-50$	$50-60$	$60-70$	$70-80$	$80-90$
No. of members	3	61	132	153	140	51	2

2. Prove that for any discrete distribution standard deviation is not less than mean deviation from mean.

3. For a group of 200 candidates, the mean and standard deviation of scores were found to be 40 and 15 respectively. Later on it was discovered that the scores 43 and 35 were misread as 34 and 53 respectively. Find the corrected mean and standard deviation corresponding to the corrected figures.

UNIT III & IV

1.1 The Königsberg Bridge Problem

Graph theory is usually said to have been invented in 1736 by the great Leonhard Euler, who used it to solve the Königsberg Bridge Problem. I used to find this hard to believe—the graph-theoretic graph is such a natural and useful abstraction that it's difficult to imagine that no one hit on it earlier—but Euler's paper about graphs[1] is generally acknowledged[2] as the first one and it certainly provides a satisfying solution to the bridge problem. The sketch in the left panel of Figure 1.1 comes from Euler's original paper and shows the main features of the problem. As one can see by comparing Figures 1.1 and 1.2, even this sketch is already a bit of an abstraction.

The question is, can one make a walking tour of the city that (a) starts and finishes in the same place and (b) crosses every bridge exactly once. The short answer to this question is "No" and the key idea behind proving this is illustrated in the right panel of Figure 1.1. It doesn't matter what route one takes while walking around on, say, the smaller island: all that really matters are the ways in which the bridges connect the four land masses. Thus we can shrink the small island to a point—and do the same with the other island, as well as with the north and south

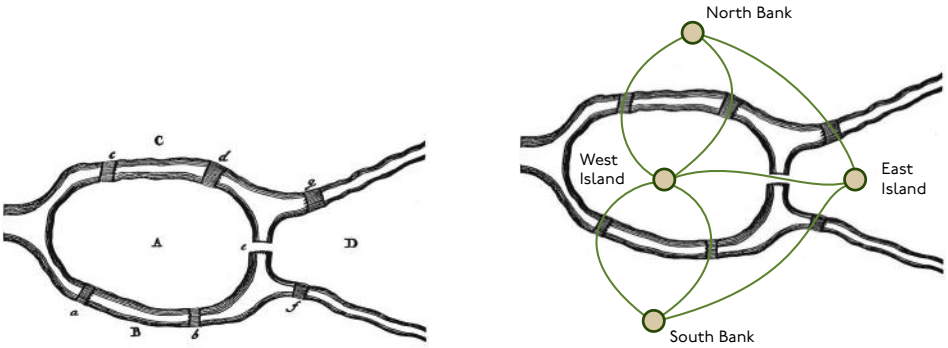


Figure 1.1: *The panel at left shows the seven bridges and four land masses that provide the setting for the Königsberg bridge problem, which asks whether it is possible to make a circular walking tour of the city that crosses every bridge exactly once. The panel at right includes a graph-theoretic abstraction that helps one prove that no such tour exists.*

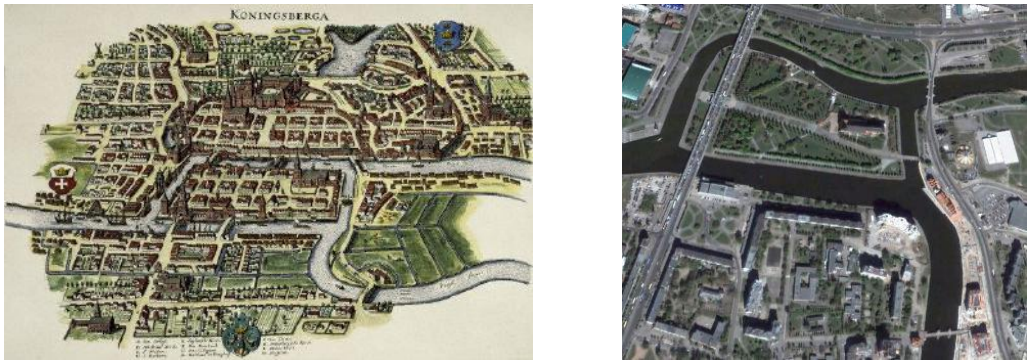


Figure 1.2: *Königsberg is a real place—a port on the Baltic—and during Euler's lifetime it was part of the Kingdom of Prussia. The panel at left is a bird's-eye view of the city that shows the celebrated seven bridges. It was made by Matthäus Merian and published in 1652. The city is now called Kaliningrad and is part of the Russian Federation. It was bombed heavily during the Second World War: the panel at right shows a recent satellite photograph and one can still recognize the two islands, but very little else appears to remain.*

banks of the river—and then connect them with arcs that represent the bridges. The problem then reduces to the question whether it is possible to draw a path that starts and finishes at the same dot, but traces each of over the seven arcs exactly once.

One can prove that such a tour is impossible by contradiction. Suppose that one exists: it must then visit the easternmost island (see Figure 1.3) and we are free to imagine that the tour actually starts there. To continue we must leave the island, crossing one of its three bridges. Then, later, because we are required to cross each bridge exactly once, we will have to return to the eastern island via a

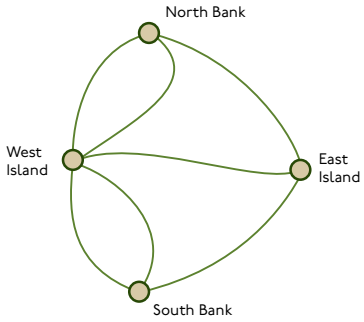


Figure 1.3: *The Königsberg Bridge graph on its own: it is not possible to trace a path that starts and ends on the eastern island without crossing some bridge at least twice.*

different bridge from the one we used when setting out. Finally, having returned to the eastern island once, we will need to leave again in order to cross the island's third bridge. But then we will be unable to return again without recrossing one of the three bridges. And this provides a contradiction: the walk is supposed to start and finish in the same place and cross each bridge exactly once.

1.2 Definitions: graphs, vertices and edges

The abstraction behind Figure 1.3 turns out to be very powerful: one can draw similar diagrams to represent "connections" between "things" in a very general way. Examples include: representations of social networks in which the points are people and the arcs represent acquaintance; genetic regulatory networks in which the points are genes and the arcs represent activation or repression of one gene by another and scheduling problems in which the points are tasks that contribute to some large project and the arcs represent interdependence among the tasks. To help us make more rigorous statements, we'll use the following definition:

Definition. *A* ***graph*** *is a finite, nonempty set V, the* ***vertex set****, along with a set E, the* ***edge set****, whose elements $e \in E$ are pairs $e = (a, b)$ with $a, b \in V$.*

We will often write $G(V, E)$ to mean the graph G with vertex set V and edge set E. An element $v \in V$ is called a *vertex* (plural *vertices*) while an element $e \in E$ is called an *edge*.

The definition above is deliberately vague about whether the pairs that make up the edge set E are ordered pairs—in which case (a, b) and (b, a) with $a \neq b$ are distinct edges—or unordered pairs. In the unordered case (a, b) and (b, a) are just two equivalent ways of representing the same pair.

Definition. *An* ***undirected graph*** *is a graph in which the edge set consists of unordered pairs.*

Definition. *A* ***directed graph*** *is a graph in which the edge set consists of ordered pairs. The term "directed graph" is often abbreviated as* ***digraph****.*

Figure 1.4: *Diagrams representing graphs with vertex set* $V = \{A, B\}$ *and edge set* $E = \{(A, B)\}$*. The diagram at left is for an undirected graph, while the one at right shows a directed graph. Thus the arrow on the right represents the ordered pair* (A, B)*.*

Although graphs are defined abstractly as above, it's very common to draw *diagrams* to represent them. These are drawings in which the vertices are shown as points or disks and the edges as line segments or arcs. Figure 1.4 illustrates the graphical convention used to mark the distinction between directed and undirected edges: the former are drawn as line segments or arcs, while the latter are shown as arrows. A directed edge $e = (A, B)$ appears as an arrow that points from A to B.

Sometimes one sees graphs with more than one edge[3] connecting the same two vertices; the Königsberg Bridge graph is an example. Such edges are called *multiple* or *parallel* edges. Additionally, one sometimes sees graphs with edges of the form $e = (v, v)$. These edges, which connect a vertex to itself, are called *loops* or *self loops*. All these terms are illustrated in Figure 1.5.

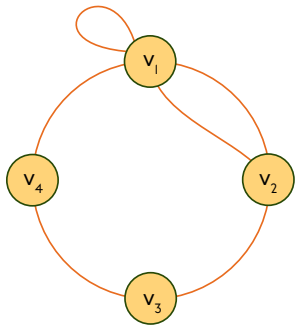


Figure 1.5: *A graph whose edge set includes the self loop* (v_1, v_1) *and two parallel copies of the edge* (v_1, v_2)*.*

It is important to bear in mind that diagrams such as those in Figures 1.3–1.5 are only illustrations of the edges and vertices. In particular, the arcs representing edges may cross, but this does not necessarily imply anything: see Figure 1.6.

Remark. In this course when we say "graph" we will normally mean an undirected graph that contains no loops or parallel edges: if you look in other books you may see such objects referred to as *simple graphs*. By contrast, we will refer to a graph that contains parallel edges as a *multigraph*.

[3]In this case it is a slight abuse of terminology to talk about the edge "set" of the graph, as sets contain only a single copy of each of their elements. Very scrupulous books (and students) might prefer to use the term *edge list* in this context, but I will not insist on this nicety.

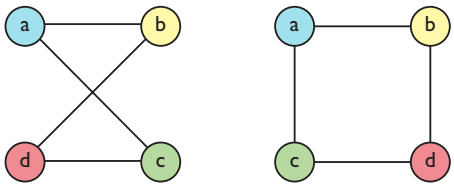


Figure 1.6: *Two diagrams for the same graph: the crossed edges in the leftmost version do not signify anything.*

Definition. *Two vertices $a \neq b$ in an undirected graph $G(V, E)$ are said to be* ***adjacent*** *or to be* ***neighbours*** *if $(a, b) \in E$. In this case we also say that the edge $e = (a, b)$ is* ***incident on*** *the vertices a and b.*

Definition. *If the directed edge $e = (u, v)$ is present in a directed graph $H(V', E')$ we will say that u is a* ***predecessor*** *of v and that v is a* ***successor*** *of u. We will also say that u is the* ***tail*** *or* ***tail vertex*** *of the edge (u, v), while v is the* ***tip*** *or* ***tip vertex****.*

1.3 Standard examples

In this section I'll introduce a few families of graphs that we will refer to throughout the rest of the term.

The complete graphs K_n

The *complete graph* K_n is the undirected graph on n vertices whose edge set includes every possible edge. If one numbers the vertices consecutively the edge and vertex set are

$$\begin{aligned} V &= \{v_1, v_2, \ldots, v_n\} \\ E &= \{(v_j, v_k) \,|\, 1 \leq j \leq (n-1),\ (j+1) \leq k \leq n\}\,. \end{aligned}$$

There are thus

$$|E| = \binom{n}{2} = \frac{n(n-1)}{2}$$

edges in total: see Figure 1.7 for the first few examples.

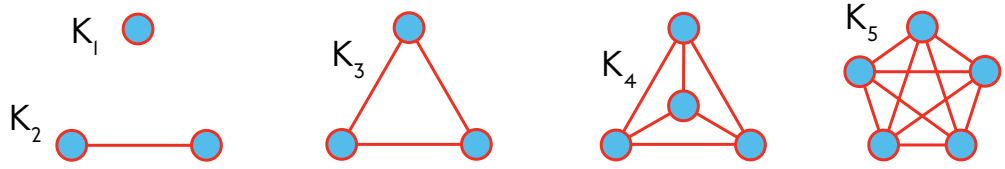


Figure 1.7: *The first five members of the family K_n of complete graphs.*

The path graphs P_n

These graphs are formed by stringing n vertices together in a path. The word "path" actually has a technical meaning in graph theory, but you needn't worry about that today. P_n has vertex and edge sets as listed below,

$$\begin{aligned} V &= \{v_1, v_2, \ldots, v_n\} \\ E &= \{(v_j, v_{j+1}) \,|\, 1 \le j < n\}\,, \end{aligned}$$

and Figure 1.8 shows two examples.

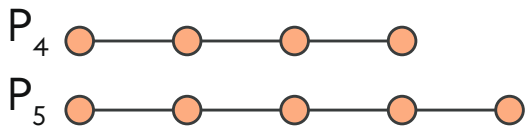


Figure 1.8: *Diagrams for the path graphs P_4 and P_5.*

The cycle graphs C_n

The *cycle graph* C_n, sometimes also called the *circuit graph*, is a graph in which $n \ge 3$ vertices are arranged in a ring. If one numbers the vertices consecutively the edge and vertex set are

$$\begin{aligned} V &= \{v_1, v_2, \ldots, v_n\} \\ E &= \{(v_1, v_2), (v_2, v_3), \ldots, (v_j, v_{j+1}), \ldots, (v_{n-1}, v_n), (v_n, v_1)\}\,. \end{aligned}$$

C_n has n edges that are often written (v_j, v_{j+1}), where the subscripts are taken to be defined periodically so that, for example, $v_{n+1} \equiv v_1$. See Figure 1.9 for examples.

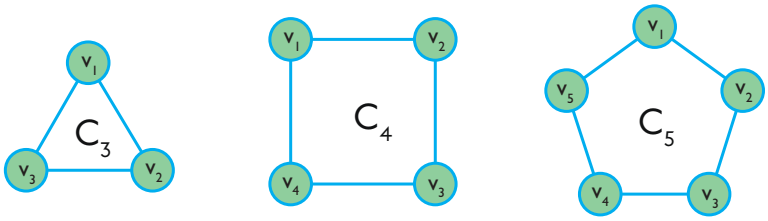


Figure 1.9: *The first three members of the family C_n of cycle graphs.*

The complete bipartite graphs $K_{m,n}$

The *complete bipartite graph* $K_{m,n}$ is a graph whose vertex set is the union of a set V_1 of m vertices with second set V_2 of n different vertices and whose edge set includes every possible edge running between these two subsets:

$$\begin{aligned} V &= V_1 \cup V_2 \\ &= \{u_1, \ldots, u_m\} \cup \{v_1, \ldots, v_n\} \\ E &= \{(u,v) \,|\, u \in V_1,\, v \in V_2\}\,. \end{aligned}$$

$K_{m,n}$ thus has $|E| = mn$ edges: see Figure 1.10 for examples.

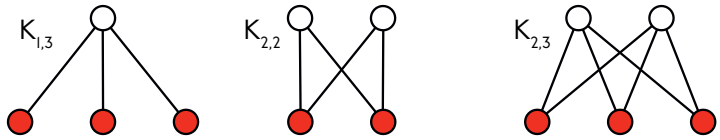


Figure 1.10: *A few members of the family $K_{m,n}$ of complete bipartite graphs. Here the two subsets of the vertex set are illustrated with colour: the white vertices constitute V_1, while the red ones form V_2.*

There are other sorts of bipartite graphs too:

Definition 1.1. *A graph $G(V,E)$ is said to be a* ***bipartite graph*** *if*

- *it has a nonempty edge set: $E \neq \emptyset$ and*
- *its vertex set V can be decomposed into two nonempty, disjoint subsets*

  $$V = V_1 \cup V_2 \;\text{ with }\; V_1 \cap V_2 = \emptyset \;\text{ and }\; V_1 \neq \emptyset \;\text{ and }\; V_2 \neq \emptyset$$

  in such a way that all the edges in E connect a member of V_1 with a member of V_2. That is, we need

  $$(u,v) \in E \Rightarrow \begin{cases} u \in V_1 \text{ and } v \in V_2 \\ \text{or } u \in V_2 \text{ and } v \in V_1. \end{cases}$$

The cube graphs I_d

These graphs are specified in a way that's closer to the purely combinatorial, set-theoretic definition of a graph given above. I_d, the *d-dimensional cube graph*, has vertices that are strings of d zeroes or ones, and all possible labels occur. Edges connect those vertices whose labels differ in exactly one position. Thus, for example, I_2 has vertex and edge sets

$$V = \{00,\, 01,\, 10,\, 11\} \qquad \text{and} \qquad E = \{(00,01),\, (00,10),\, (01,11),\, (10,11)\}\,.$$

Figure 1.11 shows diagrams for the first few cube graphs and these go a long way toward explaining the name. More generally, I_d has vertex and edge sets given by

$$\begin{aligned} V &= \left\{w \,|\, w \in \{0,1\}^d\right\} \\ E &= \{(w,w') \,|\, w \text{ and } w' \text{ differ in a single position}\}\,. \end{aligned}$$

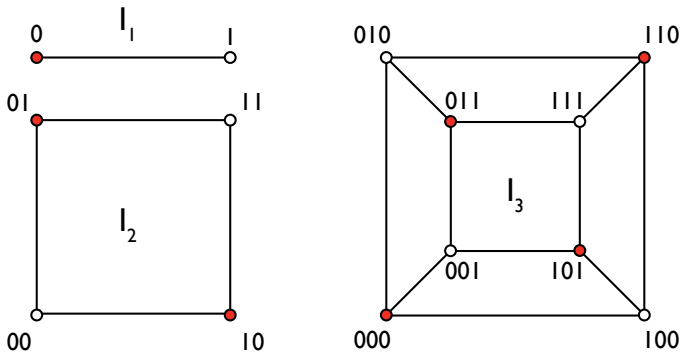


Figure 1.11: *The first three members of the family I_d of cube graphs. Notice that all the cube graphs are bipartite (the red and white vertices are the two disjoint subsets from Definition 1.1), but that, for example, I_3 is* not *a complete bipartite graph.*

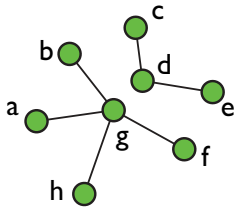
This means that I_d has $|V| = 2^d$ vertices, but it's a bit harder to count the edges. In the last part of today's lecture we'll prove a theorem that enables one to show that I_d has $|E| = d\,2^{d-1}$ edges.

1.4 A first theorem about graphs

I find it wearisome to give, or learn, one damn definition after another and so I'd like to conclude the lecture with a small, but useful theorem. To do this we need one more definition:

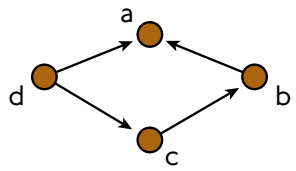
Definition. *In an undirected graph $G(V, E)$ the* ***degree*** *of a vertex $v \in V$ is the number of edges that include the vertex. One writes* $\deg(v)$ *for "the degree of v".*

So, for example, every vertex in the complete graph K_n has degree $n - 1$, while every vertex in a cycle graph C_n has degree 2; Figure 1.12 provides more examples. The generalization of degree to directed graphs is slightly more involved. A vertex v in a digraph has two degrees: an *in-degree* that counts the number of edges having v at their tip and an *out-degree* that counts number of edges having v at their tail. See Figure 1.13 for an example.



v	a	b	c	d	e	f	g	h
$\deg(v)$	1	1	1	2	1	1	4	1

Figure 1.12: *The degrees of the vertices in a small graph. Note that the graph consists of two "pieces".*



v	$\deg_{in}(v)$	$\deg_{out}(v)$
a	2	0
b	1	1
c	1	1
d	0	2

Figure 1.13: *The degrees of the vertices in a small digraph.*

Once we have the notion of degree, we can formulate our first theorem:

Theorem 1.2 (Handshaking Lemma, Euler 1736)**.** *If $G(V, E)$ is an undirected graph then*

$$\sum_{v \in V} \deg(v) = 2|E|. \tag{1.1}$$

Proof. Each edge contributes twice to the sum of degrees, once for each of the two vertices on which it is incident. □

The following two results are immediate consequences:

Corollary 1.3. *In an undirected graph there must be an even number of vertices that have odd degree.*

Corollary 1.4. *The cube graph I_d has $|E| = d\,2^{d-1}$.*

The first is fairly obvious: the right hand side of (1.1) is clearly an even number, so the sum of degrees appearing on the left must be even as well. To get the formula for the number of edges in I_d, note that it has 2^d vertices, each of degree d, so the Handshaking Lemma tells us that

$$2|E| \;=\; \sum_{v \in V} \deg(v) \;=\; 2^d \times d$$

and thus $|E| = (d \times 2^d)/2 = d\,2^{d-1}$.

Eulerian graphs

In this section, the graphs can have loops and multiple edges. Such graphs are called multi-graphs. Let $X = (V, E)$ be a graph. Then recall that a trail in X is a walk in which each edge is distinct. A graph is said to be an Eulerian graph or a closed Eulerian trail (in short Eulerian) if there is a closed trail that traverses each edge of X exactly once. Note that this is equivalent to saying that a graph X is Eulerian, if one can find a walk that traverses every edge of X exactly once and finishes at the starting vertex. A non-Eulerian graph is called an Eulerian trail if there is a walk that traverses every edge of X exactly once. The graphs that have a closed trail traversing each edge exactly once have been name "Eulerian graphs" due to the solution of Königsberg bridge problem by Euler in 1736. The problem is as follows: The city Königsberg (the present day Kaliningrad) is divided into 4 land masses by the river Pregel. These land masses are joined by 7 bridges (see Figure 1). The question required one to answer "is there a way to start from a land mass that passes through all the seven bridges in Figure 4.11 and returns back to the starting land mass"? Euler, rephrased the problem along the following lines: Let the four land masses be denoted by the vertices A, B, C and D of a graph and let the 7 bridges correspond to 7 edges of the graph. Then he asked "does this graph has a closed trail that traverses each edge exactly once"? He gave a necessary and sufficient condition for a graph to have such a closed trail and thus giving a negative answer to Königsberg bridge problem.

Observe that the definition implies that either the graph X is a connected multi-graph or in more generality, X may have isolated vertices but it has exactly one component that contains all the edges of X. So, let us assume that the multi-graphs in this section are connected. One can also relate this with the problem of drawing a given figure with pencil such that neither the pencil is lifted from the paper nor a line is repeated.

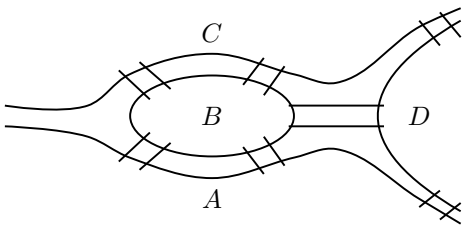


Figure 1 : Königsberg bridge problem

To state and prove the result of Euler that states that "a connected graph $X = (V, E)$ is Eulerian if and only if $\deg(v)$ is even, for each $v \in V$, we need the following result.

Lemma 4.5.1. *Let $X = (V, E)$ be a connected multi-graph such that $\deg(v) \geq 2$, for each $v \in V$. Then X contains a circuit.*

Proof. The result is clearly true if X has either a loop or a multiple edge. So, let us assume that X is a simple graph. The proof is constructive in nature. Let us start with a vertex $v_0 \in V$.

As X is connected, there exists a vertex $v_1 \in V$ that is adjacent to v_0. Since X is a simple graph and $\deg(v) \geq 2$, for each $v \in V$, there exists a vertex $v_2 \in V$ adjacent to v_1 with $v_2 \neq v_0$. Similarly, there exists a vertex $v_3 \in V$ adjacent to v_2 with $v_3 \neq v_1$. Note that either $v_3 = v_0$, in which case, one has a circuit $[v_0v_1v_2v_0]$ or else one can proceed as above to get a vertex $v_4 \in V$ and so on.

As the number of vertices is finite, the process of getting a new vertex will finally end with a vertex v_i being adjacent to a vertex v_k, for some i, $0 \leq i \leq k-2$. Hence, $[v_iv_{i+1}v_{i+2}\dots v_kv_i]$ forms a circuit. Thus, the proof of the lemma is complete. ■

Let us now prove the following theorem.

Theorem 4.5.2 (Euler 1736)**.** *Let $X = (V, E)$ be a connected graph. Then X is an Eulerian graph if and only if each vertex of X has even degree.*

Proof. Let $X = (V, E)$ be an Eulerian graph with a closed Eulerian trail $T \equiv [v_0v_1\dots v_{k-1}v_k = v_0]$. By the very nature of the trail, for each $v \in V$, the trail T enters v through an edge and departs v from another edge of X. Thus, at each stage, the process of coming in and going out, contributes 2 to degree of v. Also, the trail T passes through each edge of X exactly once and hence each vertex must be of even degree.

Conversely, let us assume that each vertex of X has even degree. We need to show that X is Eulerian. We prove the result by induction on the number of edges of X. As each vertex has even degree and X is connected, by Lemma 4.5.1 X contains a circuit, say C. If C contains every edge of X then C gives rise to a closed Eulerian trail and we are done. So, let us assume that C is a proper subset of E. Now, consider the graph X' that is obtained from X by removing all the edges in C. Then, X' may be a disconnected graph but each vertex of X' still has even degree. Hence, we can use induction to each component to X' to get a closed Eulerian trail for each component of X'.

As each component of X' has at least one vertex in common with C, we use the following method to construct the required closed Eulerian trail: start with a vertex, say v_0 of C. If there is a component of X' having v_0 as a vertex, then traverse this component and come back to v_0. This is possible as each component is Eulerian. Now, proceed along the edges of C until we get another component of X', say at v_1. Traverse the new component of X' starting with v_1 and again come back to v_1. This process will come to an end as soon as we return back to the vertex v_0 of C. Thus, we have obtained the required closed Eulerian trail. ■

We state two consequences of Theorem 4.5.2 to end this section. The proofs are omitted as they can be easily obtained using the arguments used in the proof of Theorem 4.5.2.

Corollary 4.5.3. *Let $X = (V, E)$ be a connected graph. Then X has an Eulerian trail if and only if X has exactly two vertices of odd degree.*

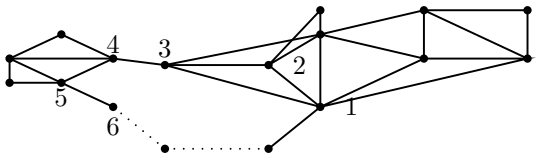


Figure 4.12: Constructing a closed Eulerian trail

Corollary 4.5.4. *Let $X = (V, E)$ be a connected graph. Then X is an Eulerian graph if and only if the edge set of X can be partitioned into cycles.*