

Dr. Suresh Kumar, Department of Mathematics, BITS-Pilani, Pilani Campus

**Note:** Some concepts of Probability & Statistics are briefly described here just to help the students. Therefore, the following study material is expected to be useful but not exhaustive for the Probability & Statistics course. For detailed study, the students are advised to attend the lecture/tutorial classes regularly, and consult the text book prescribed in the hand out of the course.

## 1 Why do we need statistics ?

Consider the ideal or perfect gas law that connects the pressure ( $P$ ), volume ( $V$ ) and temperature ( $T$ ) of a gas, and is given by

$$PV = RT$$

where  $R$  is gas constant and is fixed for a given gas. So given any two of the three variables  $P$ ,  $V$  and  $T$  of the gas, we can determine the third one. Such an approach is deterministic. However, gases may not follow the ideal gas law at extreme temperatures. Similarly, there are gases which do not follow the perfect gas law even at moderate temperatures. Here the point is that a given physical phenomenon need not to follow some deterministic formula or model. In such cases where the deterministic approach fails, we collect data related to the physical process and do the analysis using the statistical methods. These methods enable us to understand the nature of physical process to a certain degree of certainty.

## 2 Classification of Statistics

Statistics is mainly classified into three categories:

### 2.1 Descriptive Statistics

It belongs to the data analysis in the cases where the data set size is manageable and can be analysed analytically or graphically. It is the statistics that you studied in your school classes. Recall histograms, pie charts etc.

### 2.2 Inferential Statistics

It is applied where the entire data set (population) can not be analysed at one go or as a whole. So we draw a sample (a small or manageable portion) from the population. Then we analyse the sample for the characteristic of interest and try to infer the same about the population. For example, when you cook rice, you take out few grains and crush them to see whether the rice is properly cooked. Similarly, survey polls prior to voting in elections, TRP ratings of TV channel shows etc are samples based and therefore belong to the inferential statistics.

### 2.3 Model Building

Constructing a physical law or model or formula based on observational data pertains to model building. Such an approach is called empirical. For example, Kepler's laws about planetary motion belong to this approach.

Statistics is fundamentally based on the theory of probability. So first we discuss the theory of probability, and then we shall move to the statistical methods.

## 3 Theory of Probability

Depending on the nature of event, the probability is usually assigned/calculated in the following three ways.

### 3.1 Personal Approach

An oil spill has occurred from a ship carrying oil near a sea beach. A scientist is asked to find the probability that this oil spill can be contained before it causes widespread damage to the beach. The scientist assigns probability to this event considering the factors such as the amount of oil spilled, wind direction, distance of sea beach from the spot of oil spill etc. Naturally such a probability may not be accurate since it depends on the expertise of the scientist as well as the information available to the scientist. Similarly, the percentage of crops destroyed due to heavy rains in a particular area as estimated by an agriculture scientist belongs to the personal approach.

### 3.2 Relative Frequency Approach

An electrical engineer employed at a power house observes that 80 days out of 100 days, the peak demand of power supply occurs between 6 PM to 7 PM. One can immediately conclude that on any other day there are 80% chances of peak demand of power supply between 6 PM to 7 PM.

The probability assigned to an event (such as in the above example) after repeated experimentation and observation belongs to relative frequency approach.

### 3.3 Classical Approach

First we give some definitions related to this approach.

#### Random Experiment

An experiment whose outcome or result is random, that is, is not known before the experiment, is called random experiment. eg. Tossing a fair coin is a random experiment.

#### Sample Space

Set of all possible outcomes is called sample space of the random experiment and is usually denoted by  $S$ . eg. In the toss of a fair coin,  $S = \{H, T\}$ .

#### Events

Any subset of sample space is called an event. eg. If  $S = \{H, T\}$ , then the sets  $\phi$ ,  $\{H\}$ ,  $\{T\}$  and  $\{H, T\}$  all are events. The event  $\phi$  is called impossible event as it does not happen. The event  $\{H, T\}$  is called sure event as we certainly get either head or tail in the toss of a fair coin.

#### Elementary and Compound Events

Singleton subsets of sample space  $S$  are called elementary events. The subsets of  $S$  containing more than one element are known as compound events. eg. The singleton sets  $\{H\}$  and  $\{T\}$  are called elementary events while  $\{H, T\}$  is a compound event.

#### Equally Likely Events

The elementary events of a sample space are said to be equally likely if each one of them has same chance of occurring. eg. The elementary events  $\{H\}$  and  $\{T\}$  in the sample space of the toss of a fair coin are equally likely because both have same chance of occurring.

#### Mutually Exclusive and Exhaustive Events

Two events are said to be mutually exclusive if happening of one event precludes the happening of the other. eg. The events  $\{H\}$  and  $\{T\}$  in the sample space of the toss of a fair coin are mutually exclusive because both can not occur together. Similarly, more than two events say  $A_1, A_2, \dots, A_n$  are mutually exclusive if any two of these can not occur together, that is,  $A_i \cap A_j = \phi$  for  $i \neq j$ , where  $i, j \in \{1, 2, \dots, n\}$ . Further, the mutually exclusive events in a sample space are exhaustive if their union is equal to the sample space. eg. The events  $\{H\}$  and  $\{T\}$  in the sample space of the toss of a fair coin are mutually exclusive and exhaustive.

## Combination of Events

If  $A$  and  $B$  are any two events in a sample space  $S$ , then the event  $A \cup B$  implies either  $A$  or  $B$  or both;  $A \cap B$  implies both  $A$  and  $B$ ;  $A - B$  implies  $A$  but not  $B$ ;  $\bar{A}$  implies not  $A$ , that is,  $\bar{A} = S - A$ .

eg. Let  $S$  be sample space in a roll of a fair die. Then  $S = \{1, 2, 3, 4, 5, 6\}$ . Let  $A$  be the event of getting an even number and  $B$  be the event of getting a number greater than 3. Then  $A = \{2, 4, 6\}$  and  $B = \{4, 5, 6\}$ . So  $A \cup B = \{2, 4, 5, 6\}$ ,  $A \cap B = \{4\}$ ,  $A - B = \{2\}$  and  $\bar{A} = \{1, 3, 5\}$ .

## Classical Formula of Probability

Let  $S$  be sample space of a random experiment, where all the possible outcomes are equally likely. If  $A$  is any event in  $S$ , then probability of  $A$  denoted by  $P[A]$  is defined as

$$P[A] = \frac{\text{Number of elements in } A}{\text{Number of elements in } S} = \frac{n(A)}{n(S)}.$$

eg. If  $S$  is sample space for toss of two fair coins, then  $S = \{HH, HT, TH, TT\}$ . The coins being fair, here all the four outcomes are equally likely. Let  $A$  be the event of getting two heads. Then  $A = \{HH\}$ , and therefore  $P[A] = 1/4$ .

The classical approach is applicable in the cases (such as the above example) where it is reasonable to assume that all possible outcomes are equally likely. The probability assigned to an event through classical approach is the accurate probability.

## Axioms of Probability

The classical formula of probability as discussed above suggests the following:

- (i) For any event  $A$ ,  $0 \leq P[A] \leq 1$ .
- (ii)  $P[\phi] = 0$  and  $P[S] = 1$ .
- (iii) If  $A$  and  $B$  are mutually exclusive events, then  $P[A \cup B] = P[A] + P[B]$ .

These are known as axioms of the theory of probability.

## Deductions from Classical Formula

One may easily deduce the following from the classical formula:

- (i) If  $A$  and  $B$  are any two events, then  $P[A \cup B] = P[A] + P[B] - P[A \cap B]$ . This is called law of addition of probabilities.
- (ii)  $P[\bar{A}] = 1 - P[A]$ . It follows from the fact that  $A$  and  $\bar{A}$  are mutually exclusive and  $A \cup \bar{A} = S$  with  $P[S] = 1$ .
- (iii) If  $A$  is subset of  $B$ , then  $P[A] \leq P[B]$ .

**Ex.** From a pack of well shuffled cards, one card is drawn. Find the probability that the card is either a king or an ace. [Ans.  $4/52 + 4/52 = 2/13$ ]

**Ex.** Two dice are tossed once. Find the probability of getting an even number on the first die or a total of 8. [Ans.  $18/36 + 5/36 - 3/36 = 5/9$ ]

## Conditional Probability and Independent Events

Suppose a bag contains 10 Blue, 15 Yellow and 20 Green balls where all balls are identical except for the color. Let  $A$  be the event of drawing a Blue ball from the bag. Then  $P[A] = 10/45 = 2/9$ . Now suppose we are told after the ball has been drawn that the ball drawn is not Green. Because of this extra information/condition, we need to change the value of  $P[A]$ . Now, since the ball is not Green, the total number of balls can be considered as 25 only. Hence,  $P[A] = 10/25 = 2/5$ .

The extra information given in the above example can be considered as another event. Thus, if after the experiment has been conducted we are told that a particular event has occurred, then we need to revise the value of the probability of the previous event(s) accordingly. In other words, we find the probability of an event  $A$  under the condition that an event  $B$  has occurred. We call this changed probability of  $A$  as the **conditional probability** of  $A$  when  $B$  has occurred. We denote this conditional probability by  $P[A/B]$ . Mathematically, it is given by

$$P[A/B] = \frac{n(A \cap B)}{n(S \cap B)} = \frac{n(A \cap B)}{n(B)} = \frac{n(A \cap B)/n(S)}{n(B)/n(S)} = \frac{P[A \cap B]}{P[B]}.$$

The formula of conditional probability rewritten in the form

$$P[A \cap B] = P[A/B]P[B]$$

is known as multiplication law of probabilities.

Two events  $A$  and  $B$  are said to be independent if occurrence or non-occurrence of  $A$  does not affect the occurrence or non-occurrence of the other. Thus, if  $A$  and  $B$  are independent, then  $P[A/B] = P[A]$ . So  $P[A \cap B] = P[A]P[B]$  is the mathematical condition for the independence of the events  $A$  and  $B$ .

**Ex.** A die is thrown twice and the sum of the numbers appearing is noted to be 8. What is the probability that the number 5 has appeared at least once ? [Ans.  $P[A] = 11/36$ ,  $P[B] = 5/36$ ,  $P[A \cap B] = 2/36$ ,  $P[A/B] = 2/5$ .]

**Ex.** Two cards are drawn one after the other from a pack of well-shuffled 52 cards. Find the probability that both are spade cards if the first card is not replaced. [Ans.  $(13/52)(12/51) = 1/17$ ]

**Ex.** A problem is given to three students in a class. The probabilities of the solution from the three students are 0.5, 0.7 and 0.8 respectively. What is the probability that the problem will be solved ? [Ans.  $1 - (1-0.5)(1-0.7)(1-0.8) = 0.97$ ]

### Theorem of Total Probability

Let  $B_1, B_2, \dots, B_n$  be exhaustive and mutually exclusive events in a sample space  $S$  of a random experiment, each with non-zero probability. Let  $A$  be any event in  $S$ , then

$$P[A] = \sum_{i=1}^n P[B_i]P[A/B_i].$$

**Proof:** Since  $B_1, B_2, \dots, B_n$  are exhaustive and mutually exclusive events in the sample space  $S$ , so  $S = B_1 \cup B_2 \cup \dots \cup B_n$ . It follows that

$$A = A \cap S = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n).$$

Now  $B_1, B_2, \dots, B_n$  are mutually exclusive events. Therefore,  $A \cap B_1, A \cap B_2, \dots, A \cap B_n$  are mutually exclusive events. So we have

$$P[A] = P[A \cap B_1] + P[A \cap B_2] + \dots + P[A \cap B_n] = \sum_{i=1}^n P[A \cap B_i] = \sum_{i=1}^n P[B_i]P[A/B_i].$$

### Baye's Theorem

Let  $B_1, B_2, \dots, B_n$  be exhaustive and mutually exclusive events in a sample space  $S$  of a random experiment, each with non-zero probability. Let  $A$  be any event in  $S$  with  $P[A] \neq 0$ , then

$$P[B_i/A] = \frac{P[B_i]P[A/B_i]}{\sum_{i=1}^n P[B_i]P[A/B_i]}.$$

**Proof:** From conditional probability, we have

$$P[B_i/A] = \frac{P[A \cap B_i]}{P[A]} = \frac{P[B_i]P[A/B_i]}{P[A]}.$$

So the desired result follows from the theorem of total probability.

**Note:** In the above theorem,  $P[B_i]$  is known before the experiment and is known as **priori probability**.  $P[B_i/A]$  is known as **posteriori probability**. It gives the probability of the occurrence of the event  $B_i$  with respect to the occurrence of the event  $A$ .

**Ex.** Four units of a bulb making factory respectively produce 3%, 2%, 1% and 0.5% defective bulbs. A bulb selected at random from the entire output is found defective. Find the probability that it is produced by the fourth unit of the factory. [Ans.  $1/13$ ]

## 4 Discrete Random Variable, Density and Cumulative Distribution Functions

If a variable  $X$  takes real values  $x$  corresponding to each outcome of a random experiment, it is called a **random variable**. The random variable is said to be **discrete** if it assumes finite or countably infinite real values. The behavior of random variable is studied in terms of its probabilities. Suppose a random variable  $X$  takes real values  $x$  with probabilities  $P[X = x]$ . Then a function  $f$  defined by  $f(x) = P[X = x]$  is called **density function** of  $X$  provided  $f(x) \geq 0$  for all  $x$  and  $\sum_{X=x} f(x) = 1$ . Further, a function  $F$  defined by  $F(x) = \sum_{X \leq x} f(x)$  is called **cumulative distribution function** of  $X$ .

**Ex.** Consider the random experiment of tossing of two fair coins. Then the sample space is  $S = \{HH, HT, TH, TT\}$ . Let  $X$  denotes the number of heads. Then  $X$  is a random variable and it takes the values 0, 1 and 2 since  $X(HH) = 2$ ,  $X(HT) = 1$ ,  $X(TH) = 1$  and  $X(TT) = 0$ . Here the random variable assumes only three values and hence is discrete. We find that  $P[X = 0] = \frac{1}{4}$ ,  $P[X = 1] = \frac{1}{2}$  and  $P[X = 2] = \frac{1}{4}$ . It is easy to see that the function  $f$  given by

$$\begin{array}{lcl} X = x & : & 0 \quad 1 \quad 2 \\ f(x) = P[X = x] & : & \frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{4} \end{array}$$

is density function of  $X$ . It gives the probability distribution of  $X$ .

The cumulative distribution function  $F$  of  $X$  is given by

$$\begin{array}{lcl} X = x & : & 0 \quad 1 \quad 2 \\ F(x) = P[X \leq x] & : & \frac{1}{4} \quad \frac{3}{4} \quad 1 \end{array}$$

**Ex.** A fair coin is tossed again and again till head appears. If  $X$  denotes the number of tosses in this experiment, then  $X = 1, 2, 3, \dots$  since head can appear in the first toss, second toss, third toss and so on. So here the discrete random variable  $X$  assumes countably infinite values. The function  $f$  given by

$$\begin{array}{lcl} X = x & : & 1 \quad 2 \quad 3 \quad \dots \\ f(x) = P[X = x] & : & \frac{1}{2} \quad \left(\frac{1}{2}\right)^2 \quad \left(\frac{1}{2}\right)^3 \quad \dots \end{array}$$

or  $f(x) = \left(\frac{1}{2}\right)^x$ ,  $x = 1, 2, 3, \dots$ , is the density function of  $X$  since  $f(x) \geq 0$  for all  $x$  and

$$\sum_{X=x} f(x) = \sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^x = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1 \quad (\because \text{The sum of infinite G.P. } a + ar + ar^2 + \dots = a/(1-r)).$$

The cumulative distribution function  $F$  of  $X$  is given by

$$F(x) = \sum_{X \leq x} f(x) = \frac{\frac{1}{2} [1 - \left(\frac{1}{2}\right)^x]}{1 - \frac{1}{2}} = 1 - \left(\frac{1}{2}\right)^x, \text{ where } x = 1, 2, 3, \dots$$

## 5 Expectation, Variance, Standard Deviation and Moments

### 5.1 Expectation

Let  $X$  be a random variable with density function  $f$ . Then, the expectation of  $X$  denoted by  $E(X)$  is defined as

$$E(X) = \sum_{X=x} x f(x).$$

More generally, if  $H(X)$  is function of the random variable  $X$ , then we define

$$E(X) = \sum_{X=x} H(x) f(x).$$

**Ex.** Let  $X$  denotes the number of heads in a toss of two fair coins. Then  $X$  assumes the values 0, 1 and 2 with probabilities  $\frac{1}{4}$ ,  $\frac{1}{2}$  and  $\frac{1}{4}$  respectively. So  $E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1$ .

**Note:** (i) The variance  $E(X)$  of the random variable  $X$  is its theoretical average. In a statistical setting, the average value, mean value<sup>1</sup> and expected value are synonyms. The mean value is demoted by  $\mu$ . So  $E(X) = \mu$ .

(ii) If  $X$  is a random variable and  $c$  is a constant, then it is easy to verify that  $E(c) = c$  and  $E(cX) = cE(X)$ . Also,  $E(X + Y) = E(X) + E(Y)$ , where  $Y$  is another random variable.

(iii) The expected or the mean value of the random variable  $X$  is a measure of the location of the center of values of  $X$ .

## 5.2 Variance

Let  $X$  and  $Y$  be two random variables assuming the values  $X = 1, 9$  and  $Y = 4, 6$ . We observe that both the variables have the same mean values given by  $\mu_X = \mu_Y = 5$ . However, we see that the values of  $X$  are far away from the mean or the central value 5 in comparasion to the values of  $Y$ . Thus, the mean value of a random variable does not account for its variability. In this regard, we define a new parameter known as variance. It is defined as follows.

If  $X$  is a random variable with mean  $\mu$ , then its variance, denoted by  $\text{Var}(X)$  is defined as the expectation of  $(X - \mu)^2$ . So, we have

$$\text{Var}(X) = E[(X - \mu)^2] = E(X^2) + \mu^2 - 2\mu E(X) = E(X^2) + E(X)^2 - 2E(X)E(X) = E(X^2) - E(X)^2.$$

**Ex.** Let  $X$  denotes the number of heads in a toss of two fair coins. Then  $X$  assumes the values 0, 1 and 2 with probabilities  $\frac{1}{4}$ ,  $\frac{1}{2}$  and  $\frac{1}{4}$  respectively. So

$$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1,$$

$$E(X^2) = (0)^2 \times \frac{1}{4} + (1)^2 \times \frac{1}{2} + (2)^2 \times \frac{1}{4} = \frac{3}{2}.$$

$$\therefore \text{Var}(X) = \frac{3}{2} - 1 = \frac{1}{2}.$$

**Note:** (i) The variance  $\text{Var}(X)$  of the random variable  $X$  is also denoted by  $\sigma^2$ . So  $\text{Var}(X) = \sigma^2$ .

(ii) If  $X$  is a random variable and  $c$  is a constant, then it is easy to verify that  $\text{Var}(c) = 0$  and  $\text{Var}(cX) = c^2 \text{Var}(X)$ . Also,  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ , where  $X$  and  $Y$  are independent<sup>2</sup> random variables.

## 5.3 Standard Deviation

The variance of a random variable, by definition, is sum of the squares of the differences of the values of the random variable from the mean value. So variance carries squared units of the original data, and hence is a pure number often without any physical meaning. To overcome this problem, a second measure of variability is employed known as standard deviation and is defined as follows.

Let  $X$  be a random variable with variance  $\sigma^2$ . Then the standard deviation of  $X$  denoted by  $\sigma$  is the the non-negative square root of  $X$ , that is,

$$\sigma = \sqrt{\text{Var}(X)}.$$

**Note:** A large standard deviation implies that the random variable  $X$  is rather inconsistent and somewhat hard to predict. On the other hand, a small standard deviation is an indication of consistency and stability.

<sup>1</sup>From your high school mathematics, you know that if we have  $n$  distinct values  $x_1, x_2, \dots, x_n$  with frequencies  $f_1, f_2, \dots, f_n$  respectively and  $\sum_{i=1}^n f_i = N$ , then the mean value is

$$\mu = \sum_{i=1}^n \frac{f_i x_i}{N} = \sum_{i=1}^n \left( \frac{f_i}{N} \right) x_i = \sum_{i=1}^n f(x_i) x_i.$$

where  $f(x_i) = \frac{f_i}{N}$  is the probability of occurrence of  $x_i$  in the given data set. Obviously, the final expression for  $\mu$  is the expectation of a random variable  $X$  assuming the values  $x_i$  with probabilities  $f(x_i)$ .

<sup>2</sup>Independent random variables will be discussed later on.

## 5.4 Moments

Let  $X$  be a random variable and  $k$  be any positive integer. Then  $E(X^k)$  defines the  $k$ th ordinary moment of  $X$ .

Obviously,  $E(X) = \mu$  is the first ordinary moment,  $E(X^2)$  is the second ordinary moment and so on. Further, the ordinary moments can be obtained from the function  $E(e^{tX})$ . For, the ordinary moments  $E(X^k)$  are coefficients of  $\frac{t^k}{k!}$  in the expansion

$$E(e^{tX}) = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \dots$$

Also, we observe that

$$E(X^k) = \frac{d^k}{dt^k} [E(e^{tX})]_{t=0}.$$

Thus, the function  $E(e^{tX})$  generates all the ordinary moments. That is why, it is known as the moment generating function and is denoted by  $m_X(t)$ . Thus,  $m_X(t) = E(e^{tX})$ .

## 6 Geometric Distribution

Suppose a random experiment consists of a series of independent trials to obtain success, where each trial results into two outcomes namely success ( $s$ ) and failure ( $f$ ) which have constant probabilities  $p$  and  $1 - p$  respectively in each trial. Then the sample space of the random experiment is  $S = \{s, fs, ffs, \dots\}$ . If  $X$  denotes the number of trials in the experiment, then  $X$  is a discrete random variable with countably infinite values given by  $X = 1, 2, 3, \dots$ . Trials being independent, we have  $P[X = 1] = P[s] = p$ ,  $P[X = 2] = P[fs] = P[f]P[s] = (1 - p)p$ ,  $P[X = 3] = P[ffs] = P[f]P[f]P[s] = (1 - p)^2p, \dots$ . Consequently, the density function of  $X$  is given by

$$f(x) = (1 - p)^{x-1}p, \quad x = 1, 2, 3, \dots$$

The random variable  $X$  with this density function is called geometric<sup>3</sup> random variable. Given the value of the parameter  $p$ , the probability distribution of the geometric random variable  $X$  is uniquely described.

For the geometric random variable  $X$ , we have (please try the proofs)

- (i)  $m_X(t) = \frac{pe^t}{1 - qe^t}$ , where  $q = 1 - p$  and  $t < -\ln q$ ,
- (ii)  $E(X) = 1/p$ ,  $E(X^2) = (1 + q)/p^2$ ,
- (iii)  $\text{Var}(X) = q/p^2$ .

**Ex.** A fair coin is tossed again and again till head appears. If  $X$  denotes the number of tosses in this experiment, then  $X$  is a geometric random variable with the density function  $f(x) = (\frac{1}{2})^x$ ,  $x = 1, 2, 3, \dots$ . Here  $p = \frac{1}{2}$ .

## 7 Binomial Distribution

Suppose a random experiment consisting of a finite number  $n$  of independent trials is performed, where each trial results into two outcomes namely success ( $s$ ) and failure ( $f$ ) which have constant probabilities  $p$  and  $1 - p$  respectively in each trial. Let  $X$  denotes the number of successes in the  $n$  trials. Then  $X$  is a discrete random variable with values  $X = 0, 1, 2, \dots, n$ . Now, corresponding to  $X = 0$ , there is only one point in the sample space namely  $fff\dots f$  (where  $f$  repeats  $n$  times) with probability  $(1 - p)^n$ . Therefore,  $P[X = 0] = (1 - p)^n$ . Next, corresponding to  $X = 1$  there are  ${}^nC_1$  points  $sff\dots f$ ,  $fsf\dots f$ ,  $ffs\dots f$ , ...,  $fff\dots s$  (where  $s$  appears once and  $f$  repeats  $n - 1$  times) in the sample space each with probability  $(1 - p)^{n-1}p$ . Therefore,  $P[X = 1] = {}^nC_1(1 - p)^{n-1}p$ . Likewise,  $P[X = 2] = {}^nC_2(1 - p)^{n-2}p^2, \dots$ ,  $P[X = n] = p^n$ . Consequently, the density function of  $X$  is given by

$$f(x) = {}^nC_x(1 - p)^{n-x}p^x, \quad x = 0, 1, 2, 3, \dots, n.$$

The random variable  $X$  with this density function is called binomial<sup>4</sup> random variable. Once the values of the parameters  $n$  and  $p$  are given/determined, the density function uniquely describes the binomial distribution of  $X$ .

<sup>3</sup>The name geometric because the probabilities  $p, (1 - p)p, (1 - p)^2, \dots$  in succession constitute a geometric progression.

<sup>4</sup>The name binomial because the probabilities  $(1 - p)^n, {}^nC_1(1 - p)^{n-1}p, \dots, p^n$  in succession are the terms in the binomial expansion of  $((1 - p) + p)^n$ .

For the binomial random variable  $X$ , we have (please try the proofs)

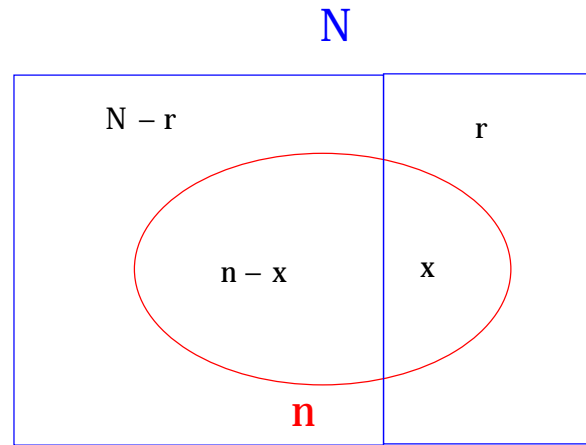
- (i)  $m_X(t) = (q + pe^t)^n$ , where  $q = 1 - p$ ,
- (ii)  $E(X) = np$ ,
- (iii)  $\text{Var}(X) = npq$ .

**Ex.** Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours ? (Here  $n = 5$ ,  $p = 1/6$ ,  $x = 2$ , and therefore  $P[X = 2] = {}^5C_2(1 - 1/6)^{5-2}(1/6)^2 = 0.161$ .)

## 8 Hypergeometric Distribution

Suppose a random experiment consists of choosing  $n$  objects without replacement from a lot of  $N$  objects given that  $r$  objects possess a trait of our interest in the lot of  $N$  objects. Let  $X$  denotes the number of objects possessing the trait in the selected sample of size  $n$ . Then  $X$  is a discrete random variable and assumes values in the range  $\max[0, n - (N - r)] \leq x \leq \min(n, r)$ . Further,  $X = x$  implies that there are  $x$  objects possessing the trait in the selected sample of size  $n$ , which should come from the  $r$  objects possessing the trait. On the other hand, the remaining  $n - x$  objects are without trait in the selected sample of size  $n$ . So these should come from the  $N - r$  objects without trait available in the entire lot of  $N$  objects. It follows that the number of ways to select  $n$  objects, where  $x$  objects with trait are to be chosen from  $r$  objects and  $n - x$  objects without trait are to be chosen from  $N - r$  objects, is  ${}^rC_x \cdot {}^{N-r}C_{n-x}$ . Also, the number of ways to select  $n$  objects from the lot of  $N$  objects is  ${}^NC_n$ .

$$\therefore P[X = x] = \frac{{}^rC_x \cdot {}^{N-r}C_{n-x}}{{}^NC_n}.$$



The random variable  $X$  with the density function  $f(x) = \frac{{}^rC_x \cdot {}^{N-r}C_{n-x}}{{}^NC_n}$ , where  $\max[0, n - (N - r)] \leq x \leq \min(n, r)$  is called hypergeometric random variable. The hypergeometric distribution is characterized by the three parameters  $N$ ,  $r$  and  $n$ .

For the hypergeometric random variable  $X$ , it can be shown that  $E(X) = n \left( \frac{r}{N} \right)$  and  $\text{Var}(X) = n \left( \frac{r}{N} \right) \left( \frac{N-r}{N} \right) \left( \frac{N-n}{N-1} \right)$ . The hypergeometric probabilities can be approximated satisfactorily by the binomial distribution provided  $n/N \leq 0.5$ .

**Ex.** Suppose we randomly select 5 cards without replacement from a deck of 52 playing cards. What is the probability of getting exactly 2 red cards ? (Here  $N = 52$ ,  $r = 26$ ,  $n = 5$ ,  $x = 2$ , and therefore  $P[X = 2] = 0.3251$ .)



## 9 Poisson Distribution

Observing discrete occurrences of an event in a continuous region or interval<sup>5</sup> is called a Poisson process or Poisson experiment. For example, observing the white blood cells in a sample of blood, observing the number of BITS-Pilani students placed with more than one crore package in five years etc. are Poisson experiments.

Let  $\lambda$  denote the number of occurrences of the event of interest per unit measurement of the region or interval. Then the expected number of occurrences in a given region or interval of size  $s$  is  $k = \lambda s$ . If  $X$  denotes the number of occurrences of the event in the region or interval of size  $s$ , then  $X$  is called a Poisson random variable. Its probability density function can be proved to be

$$f(x) = \frac{e^{-k} k^x}{x!}, \quad x = 0, 1, 2, \dots$$

We see that the Poisson distribution is characterized by the single parameter  $k$  while a Poisson process or experiment is characterized by the parameter  $\lambda$ .

It can be shown that

$$m_X(t) = e^{k(e^t - 1)}, \quad E(X) = k = \text{Var}(X).$$

**Ex.** A healthy person is expected to have 6000 white blood cells per ml of blood. A person is tested for white blood cells count by collecting a blood sample of size 0.001ml. Find the probability that the collected blood sample will carry exactly 3 white blood cells. (Here  $\lambda = 6000$ ,  $s = 0.001$ ,  $k = \lambda s = 6$  and  $x = 3$ , and therefore  $P[X = 3] = \frac{e^{-6} 6^3}{3!}$ .)

**Ex.** In the last 5 years, 10 students of BITS-Pilani are placed with a package of more than one crore. Find the probability that exactly 7 students will be placed with a package of more than one crore in the next 3 years. (Here  $\lambda = 10/5 = 2$ ,  $s = 3$ ,  $k = \lambda s = 6$  and  $x = 7$ , and therefore  $P[X = 7] = \frac{e^{-6} 6^7}{7!}$ .)

## 10 Uniform Distribution

A random variable  $X$  is said to follow uniform distribution if it assumes finite number of values all with same chance of occurrence or equal probabilities. For instance, if the random variable  $X$  assumes  $n$  values  $x_1, x_2, \dots, x_n$  with equal probabilities  $P[X = x_i] = 1/n$ , then it is uniform random variable with density function given by

$$f(x) = \frac{1}{n}, \quad x = x_1, x_2, \dots, x_n.$$

The moment generating function, mean and variance of the uniform random variable respectively read as

$$m_X(t) = \frac{1}{n} \sum_{i=1}^n e^{tx_i}, \quad \mu = \frac{1}{n} \sum_{i=1}^n x_i, \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2.$$

**Ex.** Suppose a fair die is thrown once. Let  $X$  denotes the number appearing on the die. Then  $X$  is a discrete random variable assuming the values 1, 2, 3, 4, 5, 6. Also,  $P[X = 1] = P[X = 2] = P[X = 3] = P[X = 4] = P[X = 5] = P[X = 6] = 1/6$ . Thus,  $X$  is a uniform random variable.

## 11 Continuous Random Variable

A continuous random variable is a variable  $X$  that takes all values  $x$  in an interval or intervals of real numbers, and its probability for a particular value is 0. For example, if  $X$  denotes the time of peak power demand in a power house, then it is a continuous random variable because the peak power demand happens over a continuous period of time, no matter how small or big it is. In other words, it does not happen at an instant of time or at a particular value of time variable.

<sup>5</sup>Note that the specified region could take many forms. For instance, it could be a length, an area, a volume, a period of time, etc.

A function  $f$  is called **density function** of a continuous random variable  $X$  provided  $f(x) \geq 0$  for all  $x$ ,  $\int_{-\infty}^{\infty} f(x)dx = 1$  and  $P[a \leq X \leq b] = \int_a^b f(x)dx$ . Further, a function  $F$  defined by  $F(x) = \int_{-\infty}^x f(x)dx$  is called **cumulative distribution function** (cdf) of  $X$ .

The **expectation** of a random variable  $X$  having density  $f$  is defined as

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

In general, the expectation of  $H(X)$ , a function of  $X$ , is defined as

$$E(H(X)) = \int_{-\infty}^{\infty} H(x)f(x)dx.$$

The **moment generating function** of  $X$  is defined as

$$m_X(t) = E(e^{Xt}) = \int_{-\infty}^{\infty} e^{xt}f(x)dx.$$

The **kth ordinary moment** ( $E(X^k)$ ), **mean** ( $\mu$ ) and **variance** ( $\sigma^2$ ) of  $X$  are respectively, given by

$$E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx = \left[ \frac{d^k}{dx^k} [m_X(t)] \right]_{t=0},$$

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx,$$

$$\sigma^2 = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \left( \int_{-\infty}^{\infty} xf(x)dx \right)^2.$$

**Remarks** (i) The condition  $f(x) \geq 0$  implies that the graph of  $y = f(x)$  lies on or above x-axis.

(ii) The condition  $\int_{-\infty}^{\infty} f(x)dx = 1$  graphically implies that the total area under the curve  $y = f(x)$  is 1. Therefore,  $P[a \leq X \leq b] = \int_a^b f(x)dx$  implies the area under the curve  $y = f(x)$  from  $x = a$  to  $x = b$ . Also,

$$P[a \leq X \leq b] = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx = F(b) - F(a).$$

(iii)  $P[a \leq X \leq b] = P[a < X \leq b] = P[a < X < b]$  since  $P[X = a] = 0$ ,  $P[X = b] = 0$ .

(iv)  $F'(x) = f(x)$  provided the differentiation is permissible.

**Ex.** Verify whether the function

$$f(x) = \begin{cases} 12.5x - 1.25, & 0.1 \leq x \leq 0.5 \\ 0 & \text{elsewhere} \end{cases}$$

is density function of  $X$ . If so, find  $F(x)$ ,  $P[0.2 \leq X \leq 0.3]$ ,  $\mu$  and  $\sigma^2$ .

**Sol.** Please try the detailed calculations yourself. You will find  $\int_{-\infty}^{\infty} f(x)dx = 1$ . So  $f$  is a density function. Further,

$$F(x) = \begin{cases} 0, & x < 0.1 \\ 6.25x^2 - 1.25x + 0.625, & 0.1 \leq x \leq 0.5 \\ 1, & x > 0.5 \end{cases}$$

$$P[0.2 \leq X \leq 0.3] = 0.1875.$$

$$\mu = 0.3667.$$

$$\sigma^2 = 0.00883.$$

## 12 Uniform or Rectangular Distribution

A continuous random variable  $X$  is said to have uniform distribution if its density function reads as

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{elsewhere} \end{cases}$$

In this case, the area under the curve is in the form of a rectangle. That is why the name rectangular is there. You may easily derive the following for the uniform distribution.

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & x \geq b \end{cases}$$

$$m_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}.$$

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

$$\sigma^2 = \frac{(b-a)^2}{12}.$$

## 13 Gamma Distribution

A continuous random variable  $X$  is said to have gamma distribution with parameters  $\alpha$  and  $\beta$  if its density function reads as

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x > 0, \alpha > 0, \beta > 0,$$

where  $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx$  is the gamma function.<sup>6</sup>

The moment generating function, mean and variance of the gamma random variable can be derived as

$$m_X(t) = (1 - \beta t)^{-\alpha} \quad \left( t < \frac{1}{\beta} \right).$$

$$\mu = \alpha\beta,$$

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

**Note:** The special case of gamma distribution with  $\alpha = 1$  is called **exponential distribution**. Therefore, density function of exponential distribution reads as

$$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}, \quad x > 0, \beta > 0,$$

On the other hand, the special case of gamma distribution with  $\beta = 2$  and  $\alpha = \gamma/2$ ,  $\gamma$  being some positive integer, is named as **Chi-Squared ( $\chi^2$ ) distribution**. Its density function is

$$f(x) = \frac{1}{\Gamma\left(\frac{\gamma}{2}\right) 2^{\frac{\gamma}{2}}} x^{\frac{\gamma}{2}-1} e^{-\frac{x}{2}}, \quad x > 0.$$

The  $\chi^2$  distribution arises so often in practice, extensive tables for its cumulative distribution function have been derived.

**Ex.** If  $W$  denotes the time of occurrence of the first event in a Poission process with parameter  $\lambda$ , then  $W$  has exponential distribution.

**Sol.** We have  $F(w) = P[W \leq w] = 1 - P[W > w]$ .

Let  $X$  denote the number of occurrences of the event in the time interval  $[0, w]$ . Then  $X$  is Poission random variable with parameter  $\lambda w$ . So

$$P[W > w] = P[X = 0] = e^{-\lambda w} (\lambda w)^0 / 0! = e^{-\lambda w}.$$

So  $F(w) = 1 - e^{-\lambda w}$  gives the density  $F'(w) = f(w) = \lambda e^{-\lambda w}$  of  $W$ , which shows that  $W$  follows exponential distribution.

<sup>6</sup>One should remember that  $\Gamma(1) = 1$ ,  $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ ,  $\Gamma(1/2) = \sqrt{\pi}$  and  $\Gamma(\alpha) = (\alpha - 1)!$  when  $\alpha$  is an integer.

## 14 Normal Distribution

A continuous random variable  $X$  is said to follow Normal distribution<sup>7</sup> with parameters  $\mu$  and  $\sigma$  if its density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0.$$

For the normal random variable  $X$ , one can immediately verify the following:

$$\int_{-\infty}^{\infty} f(x)dx = 1, \quad m_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, \quad \text{Mean} = \mu, \quad \text{Variance} = \sigma^2.$$

This shows that the two parameters  $\mu$  and  $\sigma$  in the density function of normal random variable  $X$  are its mean and standard deviation, respectively.

**Note:** If  $X$  is a normal random variable with mean  $\mu$  and variance  $\sigma^2$ , then we write  $X \sim N(\mu, \sigma^2)$ .

### Standard Normal Distribution

Let  $Z = \frac{X-\mu}{\sigma}$ . Then  $E(Z) = 0$  and  $\text{Var}(Z) = 1$ . We call  $Z$  as the standard normal variate and we write  $Z \sim N(0, 1)$ . Its density function reads as

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty.$$

The corresponding cumulative distribution function is given by

$$\Phi(z) = \int_{-\infty}^z \phi(z)dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz.$$

The normal probability curve is symmetric about the line  $X = \mu$  or  $Z = 0$ . Therefore, we have

$$P[X < 0] = P[X > 0] = 0.5, \quad P[-a < Z < 0] = P[0 < Z < a].$$

The probabilities of the standard normal variable  $Z$  in the probability table of normal distribution are given in terms of cumulative distribution function  $\Phi(z) = F(z) = P[Z \leq z]$  (See Table 5 on page 697 in the text book). So we have

$$P[a < Z < b] = P[Z < b] - P[Z < a] = F(b) - F(a).$$

From the normal table, it can be found that

$$P[|X - \mu| < \sigma] = P[\mu - \sigma < X < \mu + \sigma] = P[-1 < Z < 1] = F(1) - F(-1) = 0.8413 - 0.1587 = 0.6826.$$

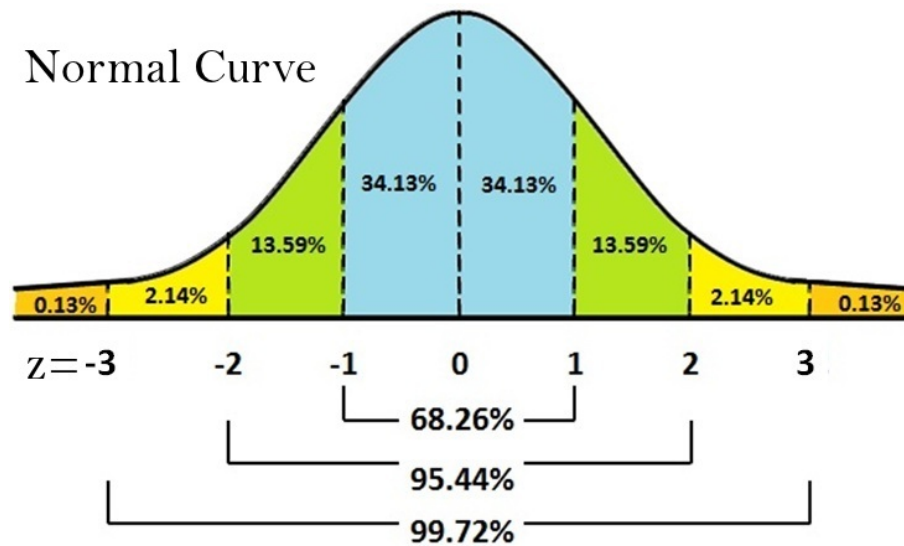
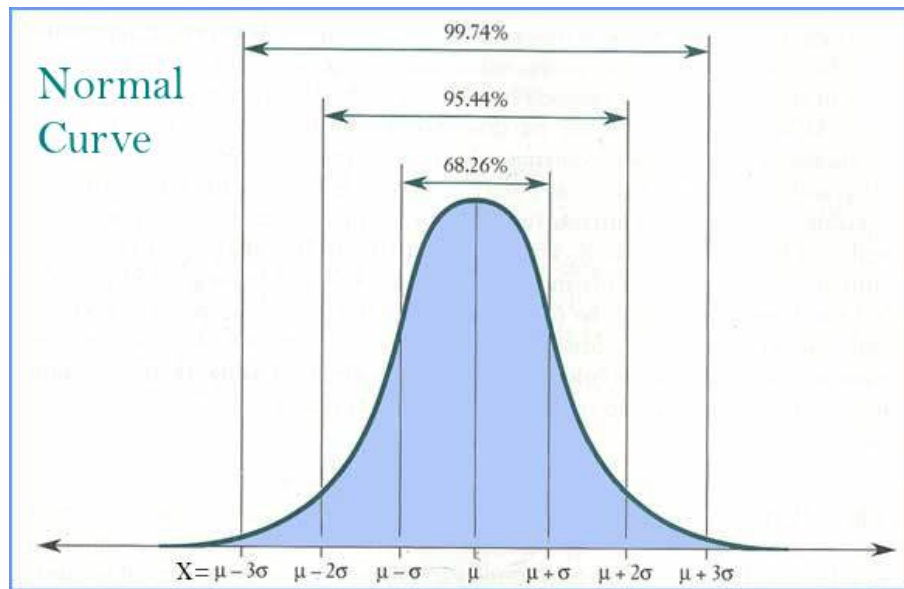
This shows that there is approximately 68% probability that the normal variable  $X$  lies in the interval  $(\mu - \sigma, \mu + \sigma)$ . We call this interval as the  $1\sigma$  confidence interval of  $X$ . Similarly, the probabilities of  $X$  in  $2\sigma$  and  $3\sigma$  confidence intervals are respectively, are given by

$$P[|X - \mu| < 2\sigma] = P[\mu - 2\sigma < X < \mu + 2\sigma] = P[-2 < Z < 2] = 0.9544,$$

$$P[|X - \mu| < 3\sigma] = P[\mu - 3\sigma < X < \mu + 3\sigma] = P[-3 < Z < 3] = 0.9973.$$

---

<sup>7</sup>Normal distribution was first described by De Moivre in 1733 as the limiting case of Binomial distribution when number of trials is infinite. This discovery did not get much attention. Around fifty years later, Laplace and Gauss rediscovered normal distribution while dealing with astronomical data. They found that the errors in astronomical measurements are well described by normal distribution. The normal distribution is also known as Gaussian distribution.



**Ex.** A random variable  $X$  is normally distributed with mean 9 and standard deviation 3. Find  $P[X \geq 15]$ ,  $P[X \leq 15]$  and  $P[0 \leq X \leq 9]$ .

**Sol.** We have  $Z = \frac{X-9}{3}$ .  $\therefore P[X \geq 15] = P[Z \geq 2] = 1 - F(2) = 1 - 0.9772 = 0.0228$ .

$$P[X \leq 15] = 1 - 0.0228 = 0.9772$$

$$P[0 \leq X \leq 9] = P[-3 \leq Z \leq 0] = F(0) - F(-3) = 0.5 - 0.0013 = 0.4987.$$

**Ex.** In a normal distribution, 12% of the items are under 30 and 85% are under 60. Find the mean and standard deviation of the distribution

**Sol.** Let  $\mu$  be mean and  $\sigma$  be standard deviation of the distribution. Given that  $P[X < 30] = 0.12$  and  $P[X < 60] = 0.85$ . Let  $z_1$  and  $z_2$  be values of the standard normal variable  $Z$  corresponding to  $X = 30$  and  $X = 60$  respectively so that  $P[Z < z_1] = 0.12$  and  $P[Z < z_2] = 0.85$ . From the normal table, we find  $z_1 \approx -1.17$  and  $z_2 \approx 1.04$  since  $F(-1.17) = 0.121$  and  $F(1.04) = 0.8508$ .

Finally, solving the equations,  $\frac{30-\mu}{\sigma} = -1.17$  and  $\frac{60-\mu}{\sigma} = 1.04$ , we find  $\mu = 45.93$  and  $\sigma = 13.56$ .

**Note:** If  $X$  is normal random variable with mean  $\mu$  and variance  $\sigma^2$ , then  $P[|X - \mu| < k\sigma] = P[|Z| < k] = F(k) - F(-k)$ . However, if  $X$  is not a random variable, then the rule of thumb for the required probability is given by the Chebyshev's inequality.

## Chebyshev's Inequality

If  $X$  is a random variable with mean  $\mu$  and variance  $\sigma^2$ , then

$$P[|X - \mu| < k\sigma] \geq 1 - \frac{1}{k^2}$$

Note that the Chebyshev's inequality does not yield the exact probability of  $X$  to lie in the interval  $(\mu - k\sigma, \mu + k\sigma)$  rather it gives the minimum probability for the same. However, in case of normal random variable, the probability obtained is exact. For example, consider the  $2\sigma$  interval  $(\mu - 2\sigma, \mu + 2\sigma)$  for  $X$ . Then, Chebyshev's inequality gives  $P[|X - \mu| < 2\sigma] \geq 1 - \frac{1}{4} = 0.75$ . In case,  $X$  is normal variable, we get the exact probability  $P[|X - \mu| < 2\sigma] = 0.9544$ . However, the advantage of Chebyshev's inequality is that it applies to any random variable of known mean and variance.

## Approximation of Binomial distribution by Normal distribution

If  $X$  is a Binomial random variable with parameters  $n$  and  $p$ , then  $X$  approximately follows a normal distribution with mean  $np$  and variance  $np(1-p)$  provided  $n$  is large. Here the word "large" is quite vague. In strict mathematical sense, large  $n$  means  $n \rightarrow \infty$ . However, for most of the practical purposes, the approximation is acceptable if the values of  $n$  and  $p$  are such that either  $p \leq 0.5$  and  $np > 5$  or  $p > 0.5$  and  $n(1-p) > 5$ . <sup>x<sup>25</sup></sup>

MORE CONCEPTS SHALL BE ADDED SOON.

Cheers!