# Module 2: Probability Distribution

By
Dr. Bidisha Bhabani

# Introduction to Probability

- The word Probability has its root in Latin words "Probare"(to test, approve) and "ilis" (able to be).
- We can use probability:
  - For studying future events
  - For understanding the underlying phenomena when our knowledge about it is incomplete
  - When it is too expensive to gather complete information
- We can apply probability
  - Insurance
  - Decision under risk
  - Effect of medicine
  - Sampling theory
  - Estimation of growth prospects of a company etc.

# Introduction to Probability

## Terminologies : Random Experiment

- Random experiment is an experiment in which the outcome is not known with certainty.
- Predictive analysis mainly deals with random experiment like:
  - Predicting quarterly revenue of an organization
  - Demand for a product at future time period etc.

# Introduction to Probability

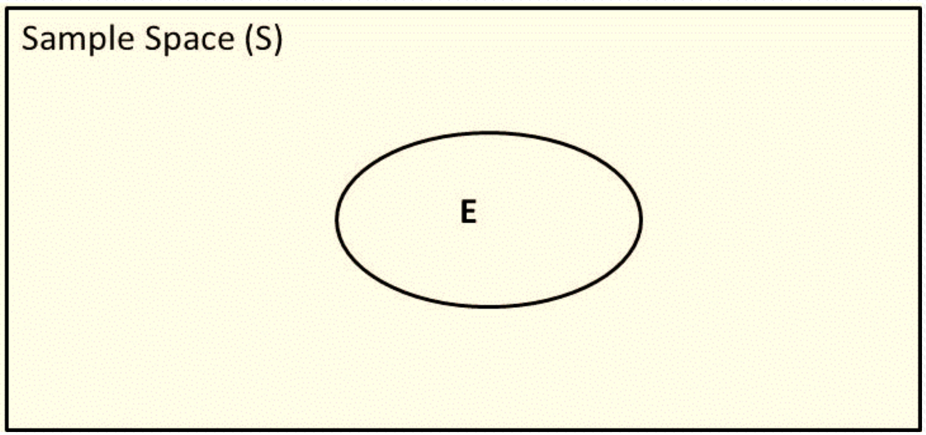
## Terminologies : Sample Space

- It is the universal set that consist of all possible outcomes of an experiment.
- It is represented using letter "S"
- Individual outcomes are called elementary events
- Sample Space can be finite or infinite.

# Introduction to Probability

## Terminologies : Event

- An Event(E) is one or several outcomes of an experiments and is a subset of a sample space and probability is usually calculated with respect to an event.




- The Venn diagram indicates that the event E is a subset of the sample space S, that is, $E \subset S$ (E is a subset of S)

# Probability Estimation using Relative Frequency

- The classical approach to probability estimation of an event is based on the relative frequency of the occurrence of that event

- According to frequency estimation, the probability of an event X, P(X), is given by

$$P(X) = \frac{\text{Number of observations in favour of event } X}{\text{Total number of observations}} = \frac{n(X)}{N}$$

# Example

- A website displays 10 advertisements and the revenue generated by the website depends on the number of visitors to the site clicking on any of the advertisements displayed in the website. The data collected by the company has revealed that out of 2500 visitors, 30 people clicked on 1 advertisement, 15 clicked on 2 advertisements, and 5 clicked on 3 advertisements. Remaining did not click on any of the advertisements. Calculate

(a) The probability that a visitor to the website will click on at least one advertisement.

(b) The probability that the visitor will click on at least two advertisements.

(c) The probability that a visitor will not click on any advertisements.

# Solution

- Number of customers clicking an advertisement is 50 and the total number of visitors is 2500. Thus, the probability that a visitor to the website will click on an advertisement is $\frac{50}{2500} = 0.02$
- Number of customers clicking on at least 2 advertisements is 20. Thus, the probability that a visitor will click on at least 2 advertisements is $\frac{20}{2500} = 0.008$
- Probability that a visitor will not click on any advertisement is $\frac{2450}{2500} = 0.98$

# Algebra of Events

- Assume that $X$, $Y$ and $Z$ are three events of a sample space. Then the following algebraic relationships are valid and are useful while deriving probabilities of events:

- **Commutative rule:** $X \cup Y = Y \cup X$ and $X \cap Y = Y \cap X$

- **Associative rule:** $(X \cup Y) \cup Z = X \cup (Y \cup Z)$ and $(X \cap Y) \cap Z = X \cap (Y \cap Z)$

- **Distributive rule:** $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$
  $$X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$$

- The following rules known as ***DeMorgan's*** *Laws* on complementary sets are useful while deriving probabilities:
  $$(X \cup Y)^C = X^C \cap Y^C$$
  $$(X \cap Y)^C = X^C \cup Y^C$$

where $X^C$ and $Y^C$ are the complementary events of $X$ and $Y$, respectively

# Axioms of Probability

According to axiomatic theory of probability, the probability of an event $E$ satisfies the following axioms

1. The probability of event $E$ always lies between 0 and 1. That is, $0 \leq P(E) \leq 1$.

2. The probability of the universal set $S$ is 1. That is, $P(S) = 1$

3. $P(X \cup Y) = P(X) + P(Y)$, where $X$ and $Y$ are two mutually exclusive events.

4. $P(X \cap Y) = \mathrm{P}(\mathrm{X}) \,.\, \mathrm{P}(\mathrm{Y})$, where X and Y are two independent events.

# Elementary Rules of Probability

The elementary rules of probability are directly deduced from the original three axioms of probability, using the set theory relationships

1. For any event $A$, the probability of the complementary event, written $A^C$, is given by $P(A) = 1 - P(A^C)$
   If $P(A)$ is a probability of observing a fraudulent transaction at an e-commerce portal, then $P(A^C)$ is the probability of observing a genuine transaction.
2. The probability of an empty or impossible event, $\phi$, is zero:
   $$P(\phi) = 0$$
3. If occurrence of an event $A$ implies that an event $B$ also occurs, so that the event class $A$ is a subset of event class $B$, then the probability of $A$ is less than or equal to the probability of $B$:

$$P(A) \leq P(B)$$

# Elementary Rules of Probability

4. The probability that either events A or B occur or both occur when A and B are not mutually exclusive is given by

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

5. If A and B are mutually exclusive events, so that $P(A \cap B) = 0$, then

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

6. 6. If A1, A2, …, An are n events that form a partition of sample space S, then their probabilities must add up to 1:

$$P(A_1) + P(A_2) + \cdots + P(A_n) = \sum_{i=1}^{n} P(A_i) = 1$$

# Joint and Marginal Probability

- When A and B are two events, then $P(A \cap B)$ is called joint probability and P(A) and P(B) are called marginal probabilities.
- **Example1**:

Find the joint and marginal probabilities for A and B, $A^c$ and $B^c$ . $P(A \cap B)$ , $P(A^c \cap B)$, $P(A \cap B^c)$ , $P(Ac \cap B^c)$

| Events | B | $B^c$ |
|---|---|---|
| A | 25 | 35 |
| $A^c$ | 45 | 15 |

**Solution:** We draw the contingency table along with row and column total for the given data.

Contingency table

| Event | $B$ | $B'$ | Total |
|---|---|---|---|
| $A$ | 25 | 35 | 60 |
| $A'$ | 45 | 15 | 60 |
| Total | 70 | 50 | 120 |

The total frequency is 120. We have to calculate the relative frequencies for $A \cap B$, $A \cap B'$ … and $A$, $A'$, $B$, $B'$

$P(A \cap B) = 25/120 = 0.206$, $P(A \cap B') = 35/120 = 0.292$

$P(A' \cap B) = 45/120 = 0.375$, $P(A' \cap B') = 15/120 = 0.125$

$P(A) = 60/120 = 0.5$, $P(A') = 60/120 = 0.5$

$P(B) = 70/120 = .583$, $P(B') = 15/120 = 0.125$

- **Example 2:**
- At an e-commerce customer service centre a total of 112 complaints were received. 78 customers complained about late delivery of the items and 40 complained about poor product quality.

(a) Calculate the probability that a customer will complain about both late delivery and product quality.

(b) What is the probability that a complaint is only about poor quality of the product?

- **Solution**
- Let $A$ = Late delivery and $B$ = Poor quality of the product. Let $n(A)$ and $n(B)$ be the number of events in favour of $A$ and $B$. So $n(A) = 78$ and $n(B) = 40$. Since the total number of complaints is 112, hence

$$n(A \cap B) = 118 - 112 = 6$$

- Probability of a complaint about both delivery and poor product quality is

$$P(A \cap B) = \frac{n(A \cap B)}{\text{Total number of complaints}} = \frac{6}{112} = 0.0535$$

- Probability that the complaint is only about poor quality = $P(A)$ = $1 - \frac{78}{112} = 0.3035$

# Conditional Probability

- If A and B are events in a sample space, then the conditional probability of the event B given that the event A has already occurred, denoted by P(B|A), is defined as

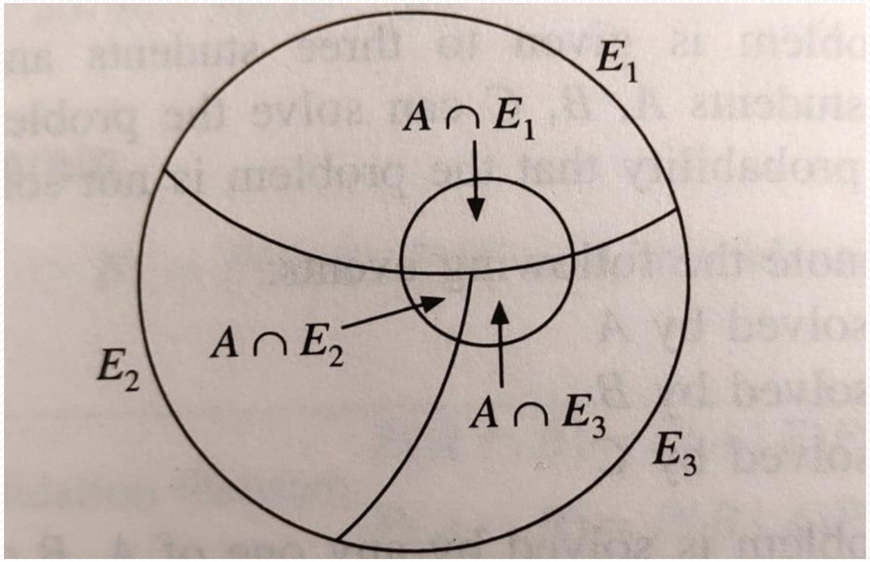
$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0$$

# Bayes Theorem

- Bayes theorem is one of the most important concepts in analytics since several problems are solved using Bayesian statistics

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

- Using the two equations, we can show that $P(B|A) = \dfrac{P(A|B)P(B)}{P(A)}$



# Terminologies used to describe various components in Bayes Theorem

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

1. $P(B)$ is called the *prior probability* (estimate of the probability without any additional information).

1. $P(B|A)$ is called the *posterior probability* (that is, given that the event $A$ has occurred, what is the probability of occurrence of event $B$). That is, *post* the additional information (or additional evidence) that $A$ has occurred, what is estimated probability of occurrence of $B$.

2. P(A|B) is called the likelihood of observing evidence A if B is true.

3. P(A) is the prior probability of A

# Example 1

- Black boxes used in aircrafts manufactured by three companies $A$, $B$ and $C$. 75% are manufactured by $A$, 15% by $B$, and 10% by $C$. The defect rates of black boxes manufactured by $A$, $B$, and $C$ are 4%, 6%, and 8%, respectively. If a black box tested randomly is found to be defective, what is the probability that it is manufactured by company $A$?

# Solution

- Let $P(A)$, $P(B)$, $P(C)$ be events corresponding to the black box being manufactured by companies $A$, $B$, and $C$, respectively, and $P(D)$ be the probability of defective black box. We are interested in calculating the probability $P(A|D)$.

$$P(A \mid D) = \frac{P(D \mid A) \times P(A)}{P(D)}$$

- Now $P(D|A) = 0.04$ and $P(A) = 0.75$. Using Eq.

$$P(D) = 0.75 \times 0.04 + 0.15 \times 0.06 + 0.10 \times 0.08 = 0.047$$

So, $$P(A \mid D) = \frac{0.04 \times 0.75}{0.047} = 0.6382$$

# Example 2

- The probability that a person has a certain disease is 0.02. If the disease is present, the medical test confirms it with probability of 0.9. If the disease is not present, the medical test confirms the absence of disease with a probability of 0.95. If the medical test indicates that a disease is present, find the probability that:
- The disease is actually present
- The disease is actually absent

**Solution:** Let $E_1$, $E_2$ and $A$ denote the following events

$E_1$: The disease is present

$E_2$: The disease is not present

$A$: The medical test gives a positive result (the presence of disease)

The given probabilities are:

$$P(E_1) = 0.02 \quad P(E_2) = 1 - P(E_1) = 0.98$$
$$P(A|E_1) = 0.9 \quad P(A'|E_2) = 0.95$$

$$\therefore \quad P(A|E_2) = \frac{P(A \cap E_2)}{P(E_2)} = \frac{P(E_2) - P(A' \cap E_2)}{P(E_2)} \quad [\text{since } P(E_2) = P(E_2 \cap A) + P(E_2 \cap A')]$$

$$= 1 - \frac{P(A' \cap E_2)}{P(E_2)} = 1 - P(A'|E_2)$$

$$= 1 - 0.95 = 0.05$$

| Event $E_i$ | Prior probability $P(E_i)$ | Conditional probability $P(A\|E_i)$ | Joint probability $P(A \cap E_i)$ | Posterior probability $P(E_i\|A)$ |
|---|---|---|---|---|
| $E_1$ | 0.02 | 0.9 | (0.02) (0.9) = 0.018 | (0.018)/0.067 = 0.269 |
| $E_2$ | 0.98 | 0.05 | (0.98) (0.05) = 0.049 | (0.049)/0.067 = 0.731 |
| | | | 0.067 | |

(a) $P(E_1|A) = 0.269$

(b) $P(E_2|A) = 0.731$

# Example 3

- An oil company has purchased an ‘oil-reserve’ on auction which can yield high-quality oil with probability 0.6, medium-quality oil with probability 0.2 and no oil with probability 0.2. The site is drilled and a particular type of soil is found. The probability of getting this type of soil in presence of high-quality oil, medium-quality oil and no oil are 0.3, 0.5, 0.2, respectively. Find the probability of finding high-quality oil, medium-quality oil and no oil.

**Solution:** Let $E_1$, $E_2$, $E_3$ and $A$ denote the following events

$E_1$: High-quality oil is found

$E_2$: Medium-quality oil is found

$E_3$: No oil is found

$A$: A particular type of soil is found

The given probabilities are

$P(E_1) = 0.6$ $\quad P(E_2) = 0.2$ $\quad P(E_3) = 0.2$

$P(A|E_1) = 0.3$ $\quad P(A|E_2) = 0.5$ $\quad P(A|E_3) = 0.2$

| Event $E_i$ | Prior probability $P(E_i)$ | Conditional probability $P(A\|E_i)$ | Joint probability $P(A \cap E_i)$ | Posterior probability $P(E_i\|A)$ |
|---|---|---|---|---|
| $E_1$ | 0.6 | 0.3 | (0.6) (0.3) = 0.18 | 0.18/0.32 = 0.5625 |
| $E_2$ | 0.2 | 0.5 | (0.2) (0.5) = 0.10 | 0.10/0.32 = 0.3125 |
| $E_3$ | 0.2 | 0.2 | (0.2) (0.2) = 0.04 | 0.04/0.32 = 0.1250 |
| | | | 0.32 | |

Probability of finding high-quality oil = 0.5625

Probability of finding medium-quality oil = 0.3125

Probability of finding no oil = 0.1250

# Application of Simple Probability Rules in Analytics

- Association rule mining is one of the popular algorithms used to solve problems such as *market basket analysis* and *recommender systems*

- Market basket analysis (MBA) is used frequently by retailers to predict products a customer is likely to buy together, which further can be used for designing planogram and product promotions

# Association Rule Mining

- Association rule learning (also known as association rule mining) is a method of finding association between different entities in a database

- Association rule is a relationship of the form

  $X \rightarrow Y$ (that is, $X$ implies $Y$).

# Association rule learning Example - Binary representation of point of sale data

| Transaction ID | Apple | Orange | Grapes | Strawberry | Plums | Green Apple | Banana |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 2 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 4 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| 6 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| 7 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |

- In Table , transaction ID is the transaction reference number and apple, orange, etc. are the different SKUs sold by the store. Binary code is used to represent whether the SKU was purchased (equal to 1) or not (equal to 0) during a transaction. The strength of association between two mutually exclusive subsets can be measured using 'support', 'confidence', and 'lift'

- **Support** between two sets (of products purchased) is calculated using the joint probability of those events:

$$\text{Support} = P(X \cap Y) = \frac{n(X \cap Y)}{N}$$

- Where $n(X \cap Y)$ is the number of times both $X$ and $Y$ is purchased together and $N$ is the total number of transactions

- **Confidence** is the conditional probability of purchasing product *Y* given the product *X* is purchased. It measures probability of event *Y* (customer buying a product *Y*) given the event *X* has occurred (the customer has already purchased product *X*). That is,

$$\text{Confidence} = \quad P(Y \mid X) = \frac{P(X \cap Y)}{P(X)}$$

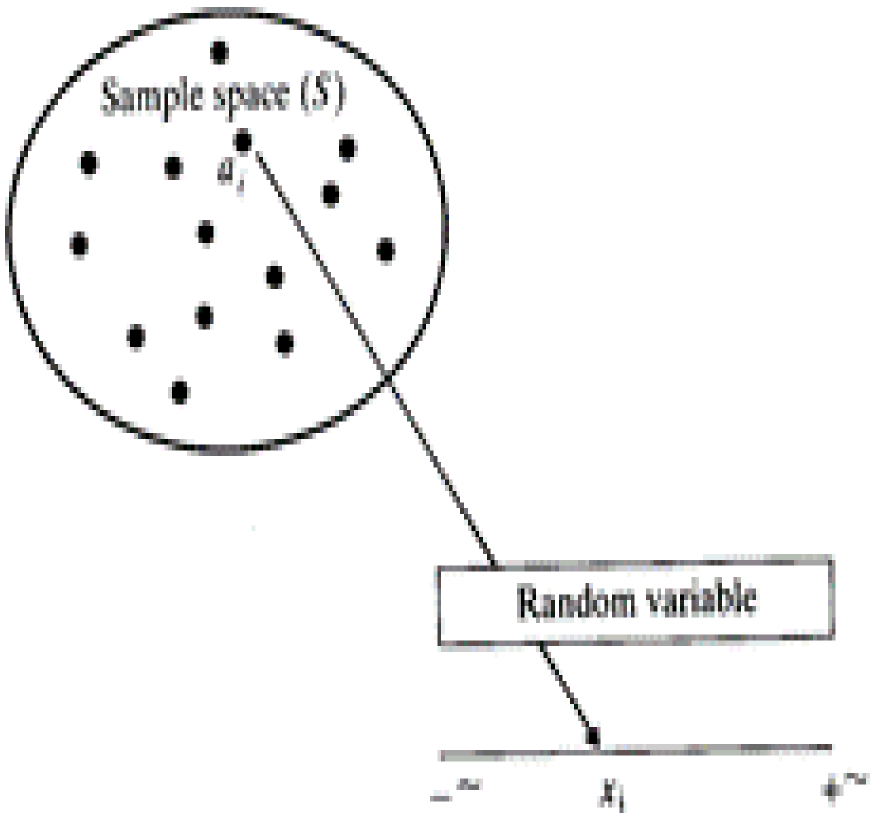
- **Lift**: The third measure in association rule mining is lift, which is given by

$$\text{Lift} = \quad \frac{P(X \cap Y)}{P(X) \times P(Y)}$$

Association rules can be generated based on threshold values of support, confidence and lift. For example, assume that the cut-off for support is 0.25 and confidence is 0.5 (Lift should be more than 1)

# Random Variables

- Random variable is a function that maps every outcome in the sample space to a real number.
- A function that assigns a real number to each sample point in the sample space S.
- Random variable is a robust and convenient way of representing the outcome of a random experiment



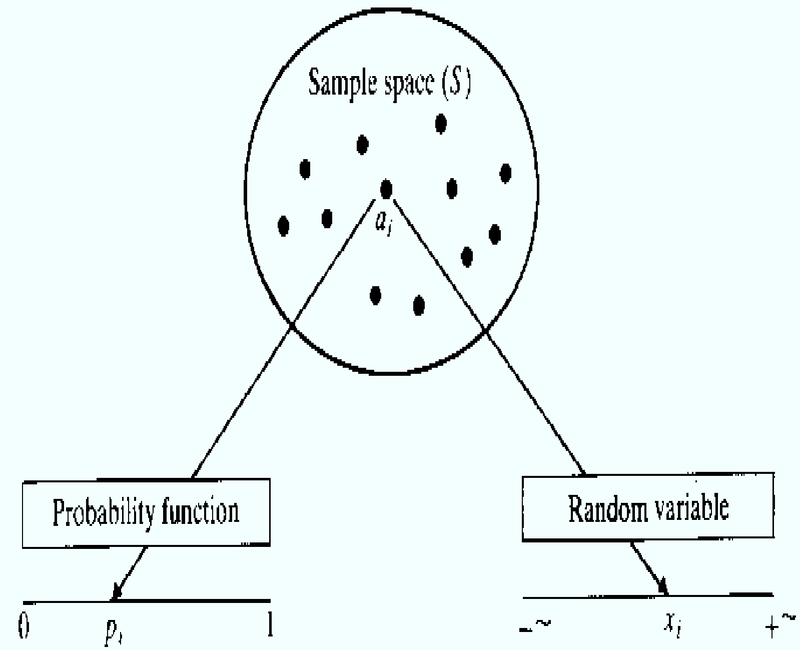
# Discrete Random Variables

- If the random variable *X* can assume only a finite or countably infinite set of values, then it is called a discrete random variable.
- Examples of discrete random variables are:
  - Credit rating (usually classified into different categories such as low, medium and high or using labels such as AAA, AA, A, BBB, etc.).
  - Number of orders received at an e-commerce retailer which can be countably infinite.
  - Customer churn (the random variables take binary values, 1. Churn and 2. Do not churn).
  - Fraud (the random variables take binary values, 1. Fraudulent transaction and 2. Genuine transaction).
  - Any experiment that involves counting (for example, number of returns in a day from customers of e-commerce portals such as Amazon, Flipkart; number of customers not accepting job offers from an organization).

# Probability mass function

- For a discrete random variable, the probability that a random variable $X$ taking a specific value $x_i$, $P(X = x_i)$, is called the probability mass function $P(x_i)$.

- That is, a probability mass function is a function that maps each outcome of a random experiment to a probability




| | | | | |
|---|---|---|---|---|
| $X$ | $x_1$ | $x_2$ | ............. | $x_n$ |
| $P$ | $p_1$ | $p_2$ | ............. | $p_n$ |

Probability distribution of a random variable

# Expected Value

- Expected value (or mean) of a discrete random variable is given by

$$\mu \quad = \qquad E(X) = \sum_{i=1}^{n} x_i P(x_i)$$

- Where $x_i$ is the specific value taken by a discrete random variable $X$ and $P(x_i)$ is the corresponding probability, that is, $P(X = x_i)$.

# Variance and Standard Deviation

- Variance of a discrete random variable is given by

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \left[x_i - E(X)\right]^2 \times P(x_i)$$

- Standard deviation of a discrete random variable is given by

$$\sigma = \sqrt{VAR(X)}$$

# Probability Distribution

Certain distributions share characteristics, so we separate them into types. The well-defined types of distributions we often deal with have elegant statistics. We distinguish between two big types of distributions based on the type of the possible values for the variable – discrete and continuous.

**Discrete**

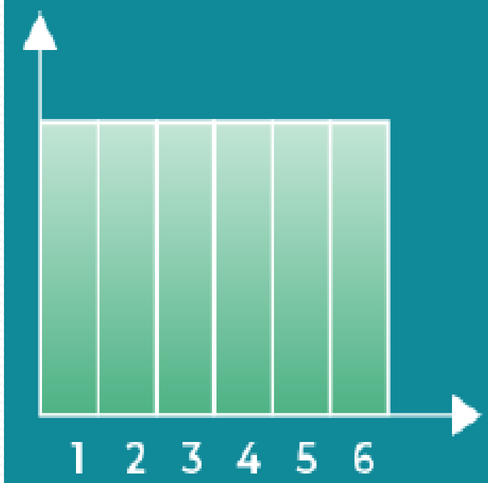
- Have a finite number of outcomes.
- Can add up individual values to determine probability of an interval.
- Can be expressed with a table, graph or a piece-wise function.
- Expected Values might be unattainable.
- Graph consists of bars lined up one after the other.

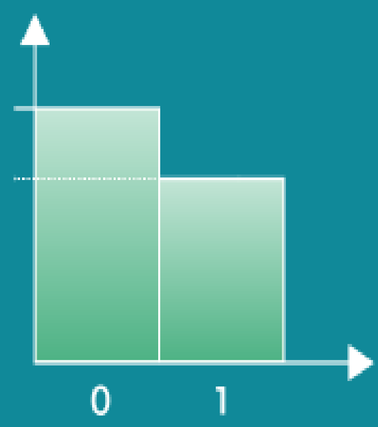
**Continuous**

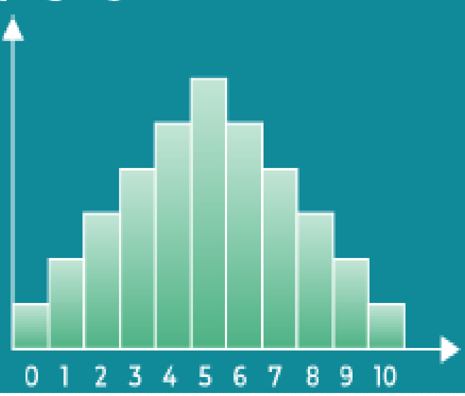
- Have infinitely many consecutive possible values.
- Use new formulas for attaining the probability of specific values and intervals.
- Cannot add up the individual values that make up an interval because there are infinitely many of them.
- Can be expressed with a graph or a continuous function.
- Graph consists of a smooth curve.

# Discrete Distribution

Discrete Distributions have finitely many different possible outcomes. They possess several key characteristics which separate them from continuous ones.




**Key characteristics of discrete distribution**

- Have a finite number of outcomes.
- Use formulas we already talked about.
- Can add up individual values to determine probability of an interval.
- Can be expressed with a table, graph or a piece-wise function.
- Expected Values might be unattainable.
- Graph consists of bars lined up one after the other.
- $P(Y \leq y) = P(Y < y + 1)$

**Examples of Discrete Distributions:**

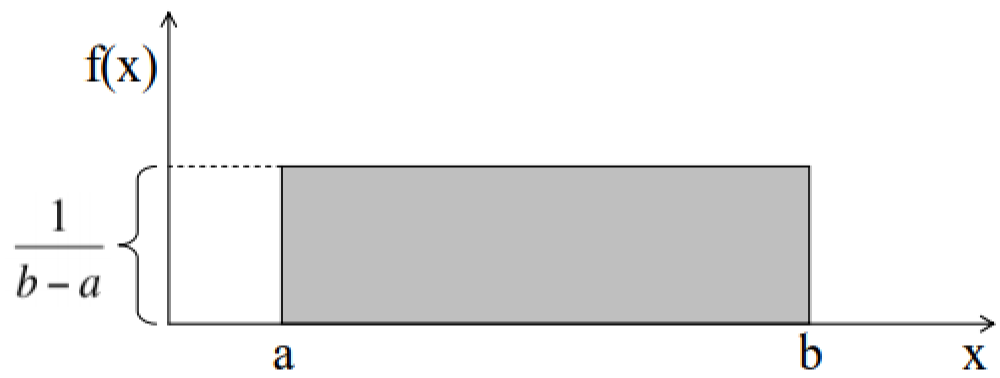
- Discrete Uniform Distribution
- Bernoulli Distribution
- Binomial Distribution
- Poisson Distribution

# Uniform Distribution

- When you roll a fair dice, the outcomes are 1 to 6. The probabilities of getting these outcomes are equally likely, which is the basis of a uniform distribution. All the n number of possible outcomes of a uniform distribution are equally likely.
- A variable X is said to be uniformly distributed if the density function is:

$$f(x) = \frac{1}{b-a} \qquad for\ -\infty < a \leq x \leq b < \infty$$

- The graph of a uniform distribution curve looks like




- You can see that the shape of the Uniform distribution curve is rectangular, the reason why Uniform distribution is called rectangular distribution.
- For a Uniform Distribution, a and b are the parameters.

# Uniform Distribution

- For the conditional probability = P( c < x < d ) = (d – c ) * f(x) = (d – c)/(b – a)
- The mean and variance of X following a uniform distribution are:
- **Mean (Expected Value) = E(X) = (a+b)/2**
- **Variance = V(X) = $(b\text{-}a)^2/12$**
- Uniform Distribution Example

  The number of bouquets sold daily at a flower shop is uniformly distributed, with a maximum of 40 and a minimum of 10.

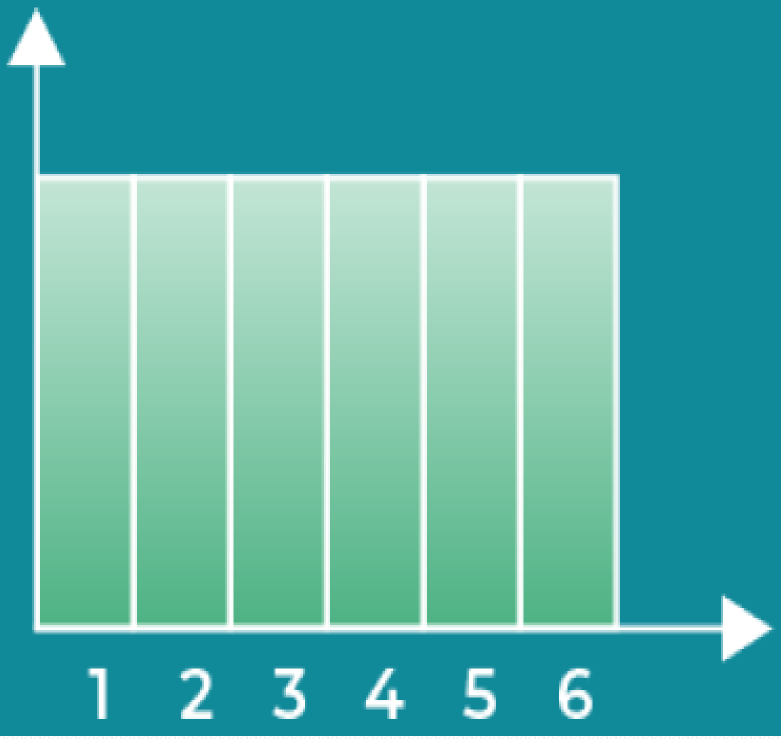
  Let's try to calculate the probability that the daily sales will fall between 15 and 30.

  The probability that daily sales will fall between 15 and 30 is (30-15)/(40-10) = 0.5

  Similarly, the probability that daily sales are greater than 20 is = 0.667
- The standard uniform density has parameters a = 0 and b = 1, so the PDF for standard uniform density is given by:

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$

A distribution where all the outcomes are equally likely is called a **Uniform Distribution**.




**Notation:**

- $Y \sim U(a, b)$
- * alternatively, if the values are categorical, we simply indicate the number of categories, like so: $Y \sim U(a)$

**Key characteristics**

- All outcomes are equally likely.
- All the bars on the graph are equally tall.
- The expected value and variance have no predictive power.

**Example and uses:**

- Outcomes of rolling a single die.
- Often used in shuffling algorithms due to its fairness.

# Bernoulli Distribution

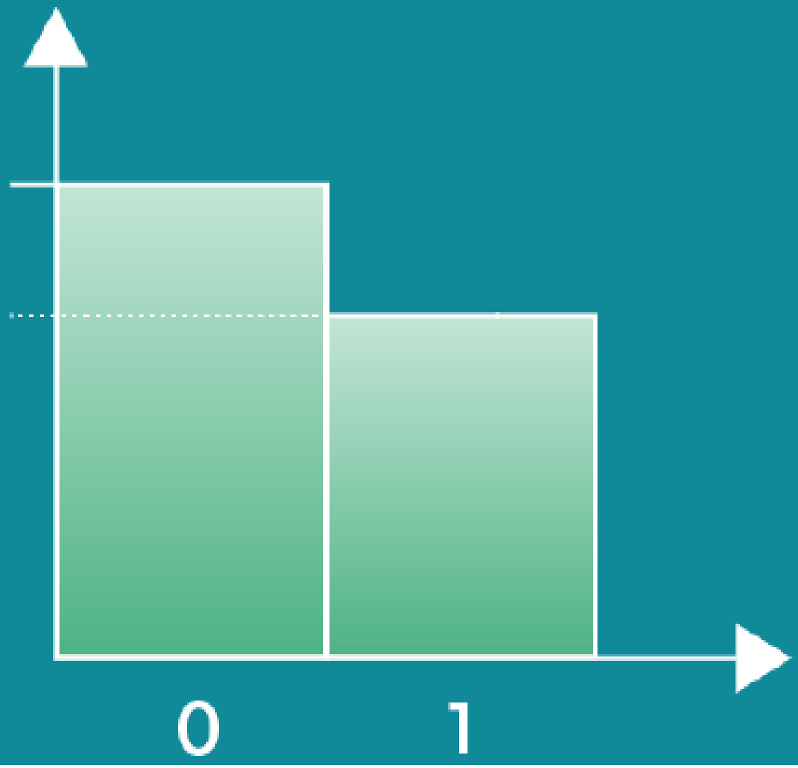
- Let's start with the easiest distribution, which is Bernoulli Distribution. It is actually easier to understand than it sounds!
- At the beginning of any cricket match, how do you decide who will bat or ball? A toss! It all depends on whether you win or lose the toss, right? Let's say if the toss results in a head, you win. Else, you lose. There's no midway.
- A **Bernoulli distribution** has only two Bernoulli trials or possible outcomes, namely 1 (success) and 0 (failure), and a single trial. So the random variable X with a Bernoulli distribution can take the value 1 with the probability of success, say p, and the value 0 with the probability of failure, say q or 1-p.
- Here, the occurrence of a head denotes success, and the occurrence of a tail denotes failure.
  Probability of getting a head = 0.5 = Probability of getting a tail since there are only two possible outcomes.
- The probability mass function is given by: $f(x) = p^x(1-p)^{1-x}$ where $x \in (0, 1)$
- It can also be written as:

$$P(x) = \begin{cases} 1-p, & x = 0 \\ p, & x = 1 \end{cases}$$

# Bernoulli Distribution

- Bernoulli Distribution Example
- The probabilities of success and failure need not be equally likely, like the result of a fight between a Grandmaster and a novice player. The Grandmaster is pretty much certain to win. So, in this case probability of the novice player's success is 0.15, while his failure is 0.85
- Here, the probability of success(p) is not the same as the probability of failure.
- Here, the probability of success = 0.15, and the probability of failure = 0.85. The expected value is exactly what it sounds like. If I punch you, I may expect you to punch me back. Basically expected value of any distribution is the mean of the distribution. The expected value of a random variable X from a Bernoulli distribution is found as follows:
- **$E(X) = 1*p + 0*(1-p) = p$**
- The variance of a random variable from a Bernoulli distribution is:
- **$V(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p)$**
- There are many examples of Bernoulli distribution, such as whether it will rain tomorrow or not, where rain denotes success and no rain denotes failure and Winning (success) or losing (failure) the game.

A distribution consisting of a single trial and only two possible outcomes – success or failure is called a Bernoulli Distribution.




Notation:

- $Y \sim Bern(p)$

Key characteristics

- One trial.
- Two possible outcomes.
- $E(Y) = p$
- $Var(Y) = p \times (1 - p)$

Example and uses:

- Guessing a single True/False question.
- Often used in when trying to determine what we expect to get out a single trial of an experiment.

# Binomial Distribution

- Suppose you won the toss today, indicating a successful event. You toss again, but you lose this time. If you win a toss today, this does not necessitate that you will win the toss tomorrow. Let's assign a random variable, say X, to the number of times you won the toss. What can be the possible value of X? It can be any number depending on the number of times you tossed a coin.
- There are only two possible outcomes. Head denoting success and tail denoting failure. Therefore, the probability of getting a head = 0.5 and the probability of failure can be easily computed as: $q = 1 - p = 0.5$.
- A distribution where only two outcomes are possible, such as success or failure, gain or loss, win or lose and where the probability of success and failure is the same for all the trials is called a Binomial Distribution.

# Binomial Distribution

- Binomial Distribution Example

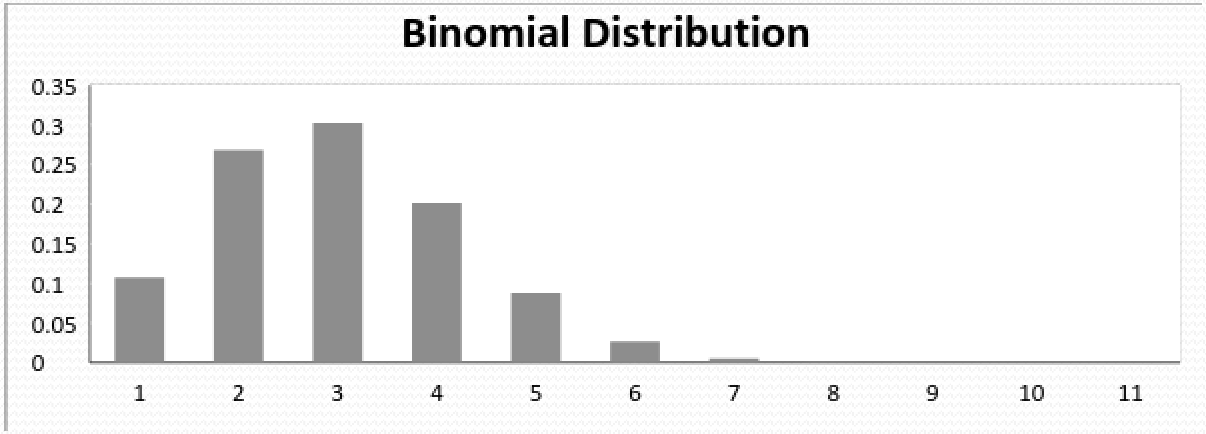
  The outcomes need not be equally likely. So, if the probability of success in an experiment is 0.2, then the probability of failure can be easily computed as q = 1 – 0.2 = 0.8.

  Each trial is independent since the outcome of the previous toss doesn't determine or affect the outcome of the current toss. An experiment with only two possible outcomes repeated n number of times is called binomial. The parameters of a binomial distribution are n and p, where n is the total number of trials and p is the probability of success in each trial.
- Based on the above explanation, the properties of a Binomial Distribution are:
- Each trial is independent.
- There are only two possible outcomes in a trial – success or failure.
- A total number of n identical trials are conducted.
- The probability of success and failure is the same for all trials. (Trials are identical.)
- The mathematical representation of binomial distribution is given by:

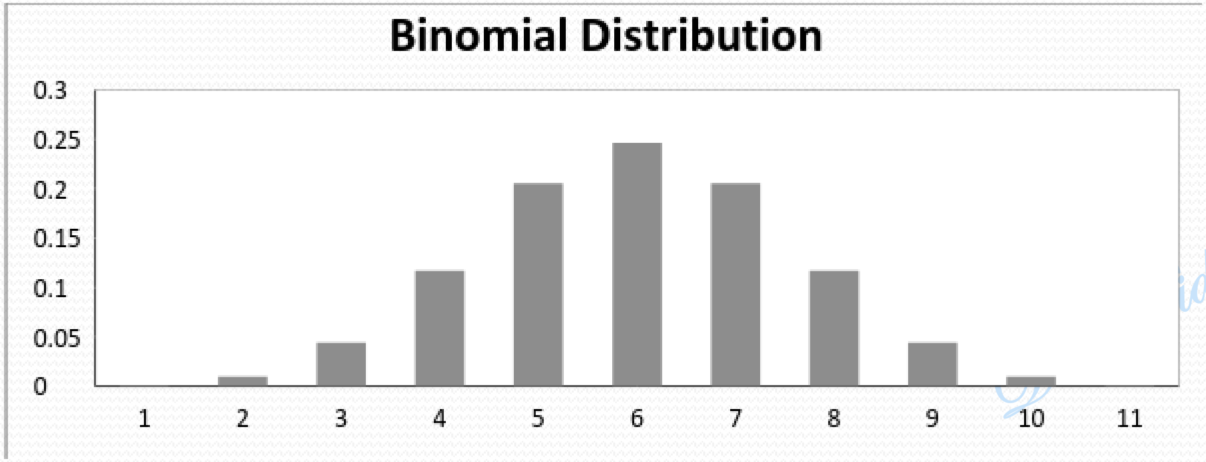
$$P(x) = \frac{n!}{(n-x)!\,x!} p^x q^{n-x} = {}^nC_x\, p^x q^{n-x}$$

# Binomial Distribution

- A binomial distribution graph where the probability of success does not equal the probability of failure looks like this.




- Now, when the probability of success = probability of failure, in such a situation, the graph of binomial distribution looks like



# Binomial Distribution

- The mean and variance of a binomial distribution are given by:
- **Mean = μ = E(X) = n*p**
- **Variance = Var(X) = n*p*q**

# Example1

- If a coin is flipped 10 times, what is the probability that it will fall heads 3 times?
- **Solution**
- Let p denote the probability of obtaining a head, and q the probability of getting a tail.
- Clearly, $n=10$, $x=3$, $p=1/2$, and $q=1/2$.
- Therefore, $b(10,3;1/2)={}^{10}C_3(1/2)^3(1/2)^7=0.1172$

# Example2

- If a medicine cures 80% of the people who take it, what is the probability that of the eight people who take the medicine, 5 will be cured?
- **Solution**
- Here p=.80, q=.20, n=8, and x=5.
- $b(8,5;.80)={}^{8}C_{5}(.80)^{5}(.20)^{3}=.1468$

# Example 3

- A telemarketing executive has determined that 15% of people contacted will purchase the product. 12 people are contacted about this product.
  a) Find the probability that among 12 people contacted, 2 will buy the product.
  b) Find the probability that among 12 people contacted, at most 2 will buy the product?

**Solution**

a) If Success denoted the probability that a person will buy the product, and Failure the probability that the person will not buy the product, then p=.15, q=.85, n=12, and x=2.

$b(12,2;.15)={}^{12}C_2(.15)^2(.85)^{10}=.2924.$

The probability that 2 people buy the product is 0.2924.

b) Again p=.15, q=.85, n=12.

But to find the probability that **at most 2** buy the product, we need to find the probabilities for x=0, x=1, x=2 and add them together.

$b(12,0;.15)= {}^{12}C_0(.15)^0(.85)^{12}=.1422$, $b(12,1;.15)={}^{12}C_1(.15)^1(.85)^{11}=.3012$

Adding all three probabilities gives: .1422 +0.3012 +.2924 = .7358

The probability that at most 2 people buy the product is 0.7358.

# Exercise 1

- A branch of a nationalized bank is giving educational loans to students. The persons in-charge of disbursement of loans claim that 40% of the students do not repay the loan. The manager is not convinced and takes random sample of 10 students. If the person in-charge of sanction of loans is correct, find the probability that:

a) Three of the 10 students do not repay
b) None of the 10 students repay

# Exercise 2

- An engineering college has the following data regarding placement of their studies in the past:
  - 30% of the students got placed in their first sitting
  - 25% of the students got placed in their second sitting
  - 20% of the students got placed in their third sitting
  - No student is allowed
- A sample of 10 students is drawn from the present set of students eligible for placement. Find the probability that:

a) At least 5 students will be placed in their first sitting

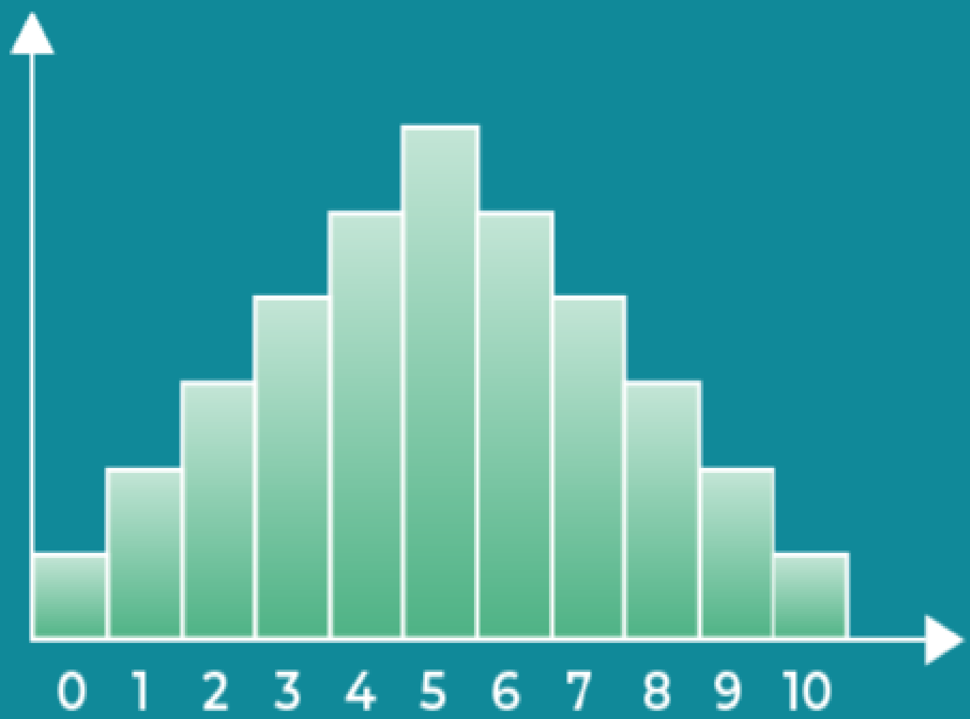
b) At least 3 students will be placed in their second sitting

c) Not more than 3 students will be placed in their third sitting

d) At least 7 students will not be placed

Also find the mean and standard deviation of the number of students who will be placed.

A sequence of identical Bernoulli events is called Binomial and follows a **Binomial Distribution**.




**Notation:**

- $Y \sim B(n, p)$

**Key characteristics**

- Measures the frequency of occurrence of one of the possible outcomes over the n trials.
- $P(Y = y) = C(y, n) \times p^y \times (1 - p)^{n-y}$
- $E(Y) = n \times p$
- $Var(Y) = n \times p \times (1 - p)$

**Example and uses:**

- Determining how many times we expect to get a heads if we flip a coin 10 times.
- Often used when trying to predict how likely an event is to occur over a series of trials.

# Relation Between Bernoulli and Binomial Distribution

- Bernoulli Distribution is a special case of Binomial Distribution with a single trial.
- There are only two possible outcomes of a Bernoulli and Binomial distribution, namely success and failure.
- Both Bernoulli and Binomial Distributions have independent trials.

# Relation Between Bernoulli and Binomial Distribution

- Bernoulli deals with the outcome of the single trial of the event, whereas Binomial deals with the outcome of the multiple trials of the single event.
- Bernoulli is used when the outcome of an event is required for only one time, whereas the Binomial is used when the outcome of an event is required multiple times.

# Poisson Distribution

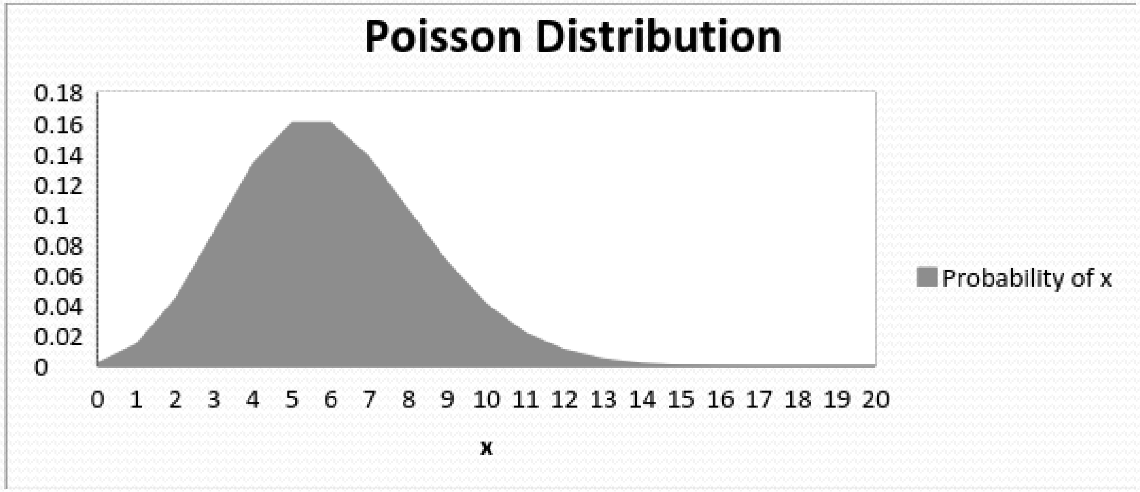
- Suppose you work at a call center; approximately how many calls do you get in a day? It can be any number. Now, the entire number of calls at a call center in a day is modeled by Poisson distribution. Some more examples are:
  - The number of emergency calls recorded at a hospital in a day.
  - The number of thefts reported in an area in a day.
  - The number of customers arriving at a salon in an hour.
  - The number of suicides reported in a particular city.
  - The number of printing errors on each page of the book.
- You can now think of many examples following the same course. Poisson Distribution is applicable in situations where events occur at random points of time and space wherein our interest lies only in the number of occurrences of the event.

# Poisson Distribution Example

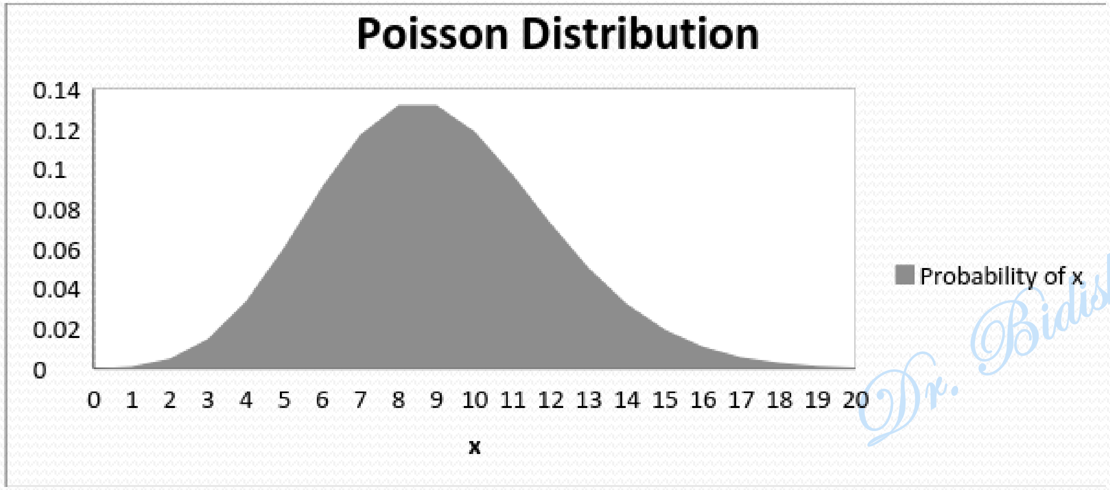
- A distribution is called a **Poisson distribution** when the following assumptions are valid:
  - Any successful event should not influence the outcome of another successful event.
  - The probability of success over a short interval must equal its probability over a longer interval.
  - The probability of success in an interval approaches zero as the interval becomes smaller.
- Now, if any distribution validates the above assumptions, then it is a Poisson distribution. Some notations used in Poisson distribution are:
  - $\lambda$ is the rate at which an event occurs,
  - t is the length of a time interval,
  - And X is the number of events in that time interval.
- Here, X is called a Poisson Random Variable, and the probability distribution of X is called Poisson distribution.
- Let $\mu$ denote the mean number of events in an interval of length t. Then, $\mu = \lambda * t$.
- The PMF of X following a Poisson distribution is given by:

$$P(X = x) = e^{-\mu} \frac{\mu^{x}}{x!} \qquad for\ x = 0, 1, 2, \ldots\ldots$$

- The mean μ is the parameter of this distribution. μ is also defined as the λ times the length of that interval. The graph of a Poisson distribution is shown below:



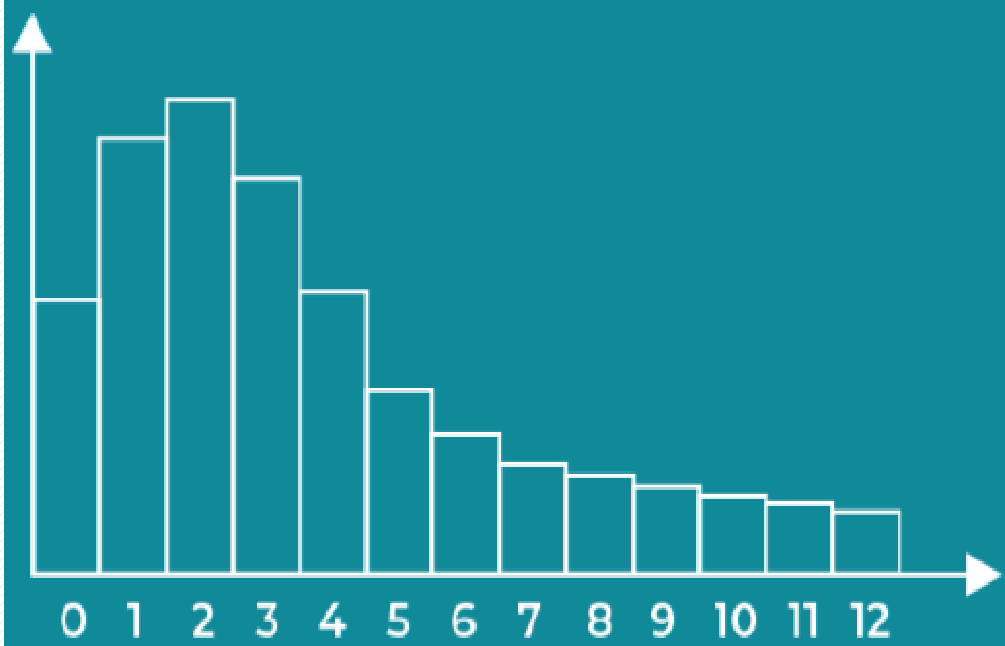

- The graph shown below illustrates the shift in the curve due to the increase in the mean.



# Poisson Distribution

- It is perceptible that as the mean increases, the curve shifts to the right.
- The mean and variance of X following a Poisson distribution:
- **Mean = E(X) = μ**
- **Variance = Var(X) = μ**

When we want to know the likelihood of a certain event occurring over a given interval of time or distance we use a Poisson Distribution.




**Notation:**

- $Y \sim Po(\lambda)$

**Key characteristics**

- Measures the frequency over an interval of time or distance. (Only non-negative values.)
- $P(Y = y) = \frac{\lambda^y}{y! e^{-\lambda}}$
- $E(Y) = \lambda$
- $Var(Y) = \lambda$

**Example and uses:**

- Used to determine how likely a specific outcome is, knowing how often the event **usually** occurs.
- Often incorporated in marketing analysis to determine whether above average visits are out of the ordinary or not.

# Example 1

The number of industrial injuries per working week in a particular factory is known to follow a Poisson distribution with mean 0.5.

Find the probability that

a) In a particular week there will be:
   i. Less than 2 accidents
   ii. More than 2 accidents

b) In a three week period there will be no accidents

**Solution**:

Let A be the number of accidents in one week. $\mu = 0.5$

a) i. $P(A<2) = P(A=0) + P(A=1) = e^{-0.5} + \frac{e^{-0.5}\, x\, 0.5}{1!} = \frac{3}{2} e^{-0.5} = 0.9098$

   ii. $P(A>2) = 1 - P(A \leq 2) = 1 - [P(A=0) + P(A=1) + P(A=2)]$

$$= 1 - [e^{-0.5} + e^{-0.5}\, x\, 0.5 + \frac{e^{-0.5}\ \ 0.5^2}{2!}]$$

$$= 1 - e^{-0.5}\, [1 + 0.5 + 0.125]$$

$$= 1 - 1.625 \text{ x } e^{-0.5}$$

$$= 0.0144$$

b) $P(0 \text{ in } 3 \text{ weeks}) = e^{-0.5x3} = 0.223$

# Example 2

- A Bank Manager found that 3 customers arrive on an the average in every 5 minutes in a savings bank counter. Assume, that the customers arrive at random:
  a) Find the probability that 5 customers arrive in a 5 minute interval.
  b) The manager wants to add one more counter for SB customers if the probability that more than 5 customers arrive in a 5-minute interval exceeds 0.2. Will the Manager add one more counter?
  c) Do subdivisions (a) and (b) for a 10 minute interval.

# Solution

- As the arrival of customers is random, we can assume that the arrival in a 5-minute interval is a Poisson distribution. μ = The average number of arrivals = 3

a) Probability (5 customers arrive in a 5-minute interval)

$$= P(5) = \frac{e^{-3} x\ 3^5}{5!} = \frac{0.0498\ x\ 243}{120} = 0.1008$$

b) P(X>5) = 1 – P(X ≤ 5)

= 1 – P(0) – P(1) – p(2) – P(3) – P(4) – P(5)

$$= 1 - e^{-3}\left(1 + \frac{3}{1!} + \frac{3^2}{2!} + \frac{3^3}{3!} + \frac{3^4}{4!} + \frac{3^5}{5!}\right)$$

= 1 – 0.0498 x 18.4

= 1 – 0.9160 = 0.0840

As P(X>5) is less than 0.2, the manager will not add one more counter.

c) The average number of arrivals in a 10-minute interval = 2 x 3 = 6

$$P(5) = \frac{e^{-6}\ x\ 6^5}{5!} = 0.1620 \qquad P(X>5) = 0.5505$$

# Example 3

- A large private sector bank has given a toll free number which is available to the customers to register their complaints regarding the service provided by its branches. It is found that 48 calls are received by the toll-free phone in one hour on the average. Assuming the number of calls follow a Poisson distribution, find the probability that:

a) No calls will be received in a minute period

b) 2 or more calls will be received in 1-minute period

c) No call will be received in a 2-minute period

d) 4 or more calls will be received in a 2-minute period

e) What is the maximum number of calls received 99% of the time in a 1-minute period?

*Solution:* Let $X$ denote the number of phone calls per minute.

It is a Poisson distribution with $\mu = \frac{48}{60} = 0.8$

(a) $P$(no calls in 1-minute period)

$$= p(0) = e^{-0.8} = 0.4493$$

(b) $P(X \geq 2) = 1 - p(0) - p(1)$

$$= 1 - e^{-0.8} - e^{-0.8}(0.8)$$
$$= 1 - 0.4493 - 0.4493(0.8)$$
$$= 1 - 0.8088$$
$$= 0.1912$$

(c) Let $Y$ denote the number of calls in a 2-minute period. Then $Y$ follows a Poisson distribution with $\mu = 1.6$

$P$(no calls in a 2-minute period)

$$= p(0)$$
$$= e^{-1.6}$$
$$= 0.2019$$

(d) $P(X \geq 4) = 1 - p(0) - p(1) - p(2) - p(3)$

$$= 1 - e^{-1.6}\left(1 + 1.6 + \frac{(1.6)^2}{2!} + \frac{(1.6)^3}{3!}\right)$$
$$= 1 - e^{-1.6}\ (4.5627)$$
$$= 1 - 0.9211$$
$$= 0.0789$$

For finding the maximum number of calls in a 1-minute period with probability 0.99, let us find $p(0)$, $p(1)$, $p(2)$, etc. We add these probabilities until the sum exceeds 0.99

$$p(0) = 0.4493 \quad p(1) = 0.3595 \quad p(2) = 0.1438 \quad p(3) = 0.0383$$
$$p(0) + p(1) + p(2) = 0.4493 + 0.3595 + 0.1438 = 0.9526$$
$$p(0) + p(1) + p(2) + p(3) = 0.9526 + 0.0383 = 0.9909$$

So, we receive up to 3 calls with probability 0.9909. Thus, the maximum number of calls in a 1-minute period with probability 0.99 is 3 (*Note*, we can receive up to 2 calls only with probability 0.9526 which is less than 0.99)

# Exercise 3

- The number of misprints on a page of the Daily Mercury has Poisson Distribution with mean 1.2. Find the probability that the number of errors

a) On page four is 2
b) On page three is less than 3
c) On the first ten pages totals 5
d) On all forty pages up to at least 3

# Exercise 4

- A shop sells a particular make of video recorder.

a) Assuming that the weekly demand for the video recorder is a Poisson variable with mean 3, find the probability that the shop sells

   i. At least 3 in a week
   ii. At most 7 in a week
   iii. More than 20 in a month (4 weeks)

   Stocks are replenished at the beginning of each month.

b) Find the minimum number that should be in stock at the beginning of a month so that the shop can be at least 95% sure of being able to meet the demands during the month.

# Application of Poisson Distribution

- The number of arrivals of customers in a queue for bus.
- The number of patches requiring for repair on a stretch of 5 km on a highway.
- The number of persons who died of horse kick (a classical example) or in a road accident (in modern days).
- Number of breakdowns of computer terminals in a railway reservation counter in a day
- Number of successes in a Bernoulli process having large number of trials.
- The number of mistakes in a page of a book published by a reputed publishing company.

# Continuous Random Variables

- A random variable *X* which can take a value from an infinite set of values is called a continuous random variable
- Examples of continuous random variables are listed below:
  - Market share of a company (which take any value from an infinite set of values between 0 and 100%).
  - Percentage of attrition among employees of an organization.
  - Time to failure of engineering systems.
  - Time taken to complete an order placed at an e-commerce portal.
  - Time taken to resolve a customer complaint at call and service centers.

# Probability Density Function (PDF)

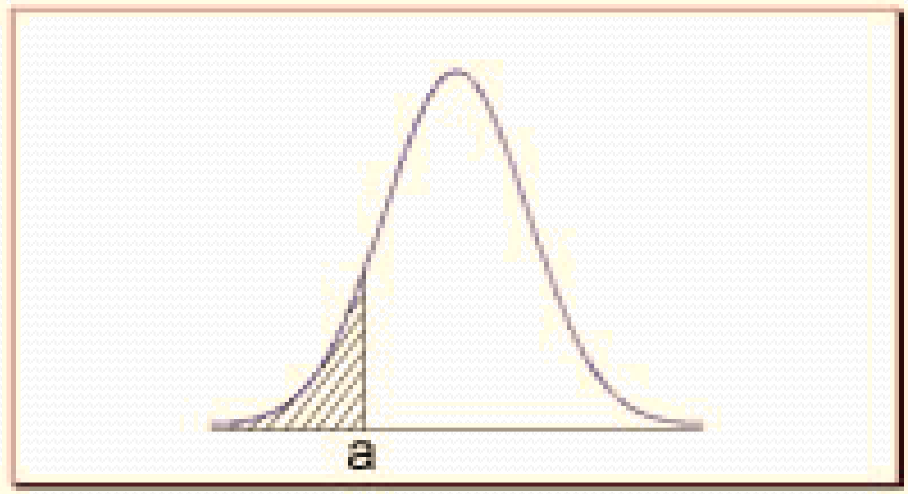
- The probability density function, $f(x_i)$, is defined as probability that the value of random variable $X$ lies between an infinitesimally small interval defined by $x_i$ and $x_i + \delta x$

$$f(x) = \lim_{\delta x \to 0} \frac{P(x_i \le X \le x_i + \delta x)}{\delta x}$$

# Cumulative Distribution Function (CDF)

- The cumulative distribution function (CDF) of a continuous random variable is defined by

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




Cumulative distribution function

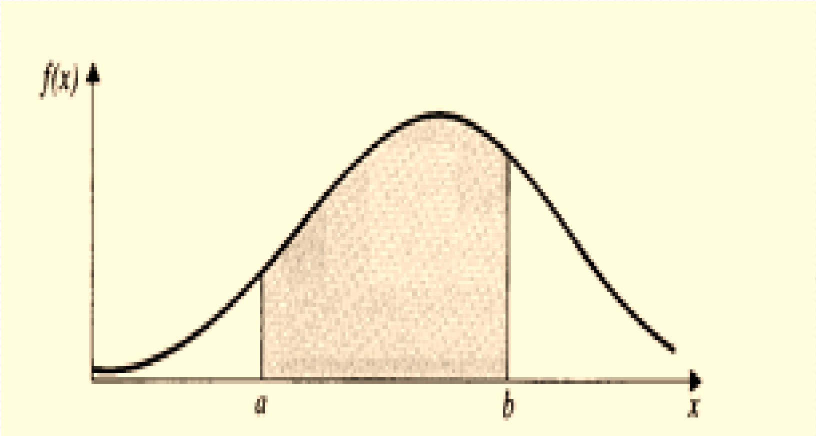
# PDF and CDF

- Probability density function and cumulative distribution function of a continuous random variable satisfy the following properties
- $f(x) \geq 0$

$$F(\infty) = \int_{-\infty}^{+\infty} f(x)dx = 1$$

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

- The probability between two values a and b, $P(a \leq X \leq b)$, is the area between the values a and b under the probability density function



# Expected Value and Variance

- The expected value of a continuous random variable, $E(X)$, is given by

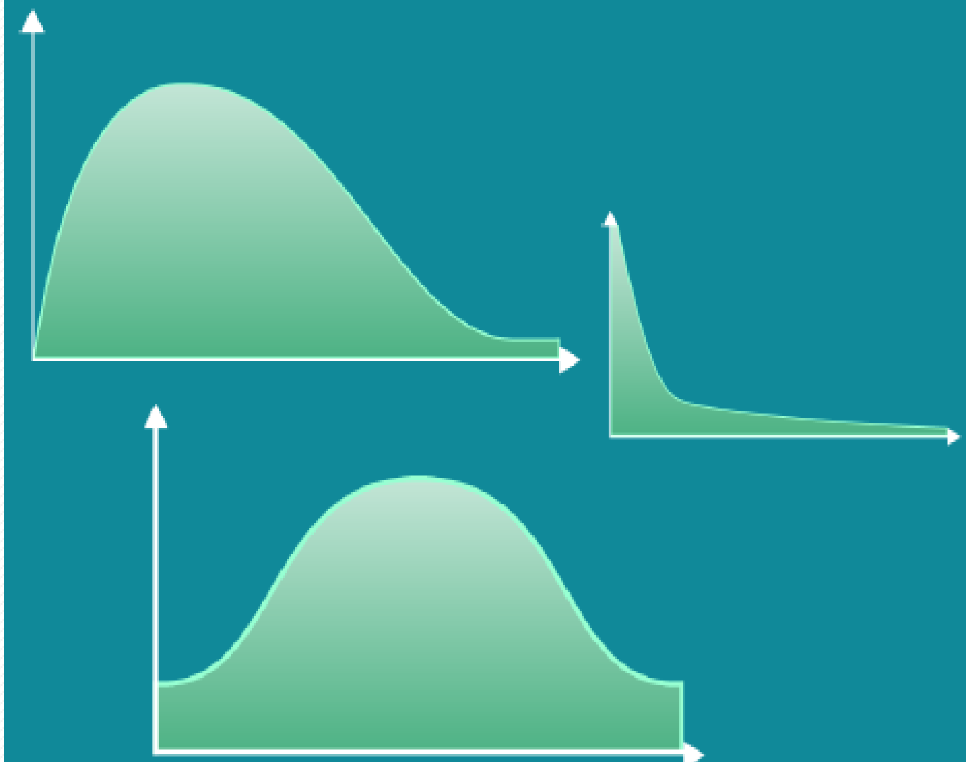
$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

- The variance of a continuous random variable, Var(X), is given by

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} [x - E(x)]^2 f(x) dx$$

# Continuous Distribution

If the possible values a random variable can take are a sequence of infinitely many consecutive values, we are dealing with a **continuous distribution**.



**Key characteristics**

- Have infinitely many consecutive possible values.
- Cannot add up the individual values that make up an interval because there are **infinitely many** of them.
- Can be expressed with a graph or a continuous function. Cannot use a table, be
- Graph consists of a smooth curve.
- To calculate the likelihood of an interval, we need integrals.
- They have important CDFs.
- $P(Y = y) = 0$ for any individual value y.
- $P(Y < y) = P(Y \leq y)$

# Normal Distribution

- The **normal distribution** represents the behavior of most of the situations in the universe (That is why it's called a "normal" distribution. I guess!). The large sum of (small) random variables often turns out to be normally distributed, contributing to its widespread application. Any distribution is known as Normal distribution if it has the following characteristics:
  - The mean, median, and mode of the distribution coincide.
  - The curve of the distribution is bell-shaped and symmetrical about the line $x=\mu$.
  - The total area under the curve is 1.
  - Exactly half of the values are to the left of the center, and the other half to the right.
- A normal distribution is highly different from Binomial Distribution. However, if the number of trials approaches infinity, then the shapes will be quite similar.

- The PDF of a random variable X, following a normal distribution, is given by:

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

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt, \qquad -\infty < x < +\infty$$

- The mean and variance of a random variable X, which is said to be normally distributed, is given by:
- **Mean -> E(X) = μ**
- **Variance -> Var(X) = σ^2**
- Here, μ (mean) and σ (standard deviation) are the parameters.

# Properties of Normal Distribution

1. The normal curve is bell-shaped (see Figure 4.1) and the area under the normal curve is one [by Eq. (4.8)]

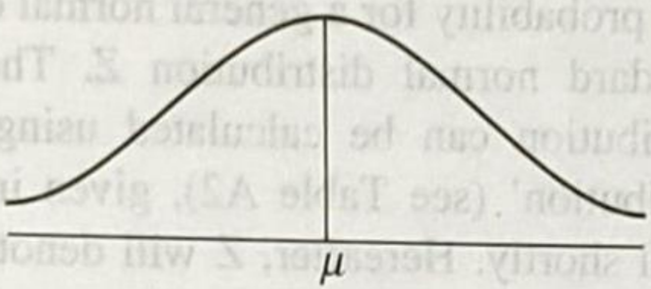



**Figure 4.1** General normal curve.

2. The tails of a normal distribution extend indefinitely on either side. They are coming closer to the $x$-axis, but never intersect the $x$-axis. [This follows since $f(x) > 0$ and $e^{-((x-\mu)/\sigma)^2}$ is very nearly zero for values with large modulus.]
3. The normal curve is symmetric about the vertical line $x = \mu$; the mean, median and mode of a normal distribution are all equal to $\mu$. This means the normal curve has a single peak at $\mu$ and the vertical line $x = \mu$ divides the area under the curve into two halves.

# Properties of Normal Distribution

4. If the means of two normal distributions are the same, then the normal curve with a larger $\sigma$ is flatter (see Figure 4.2).

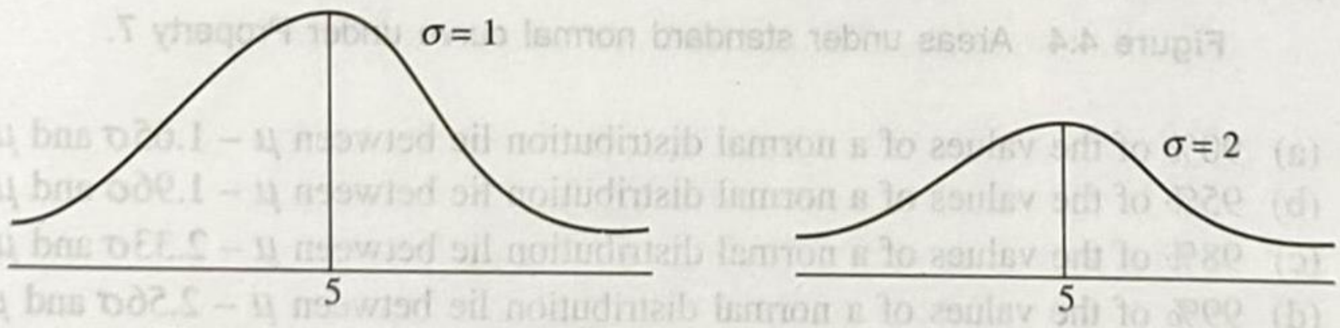



**Figure 4.2** Normal curves with different standard deviations.

5. If two normal distributions have different means $\mu_1$ and $\mu_2$, but the same $\sigma$, then the normal curve with mean $\mu_2$ can be obtained by displacing the normal curve with mean $\mu_1$ horizontally by a distance $\mu_2 - \mu_1$ (see Figure 4.3).

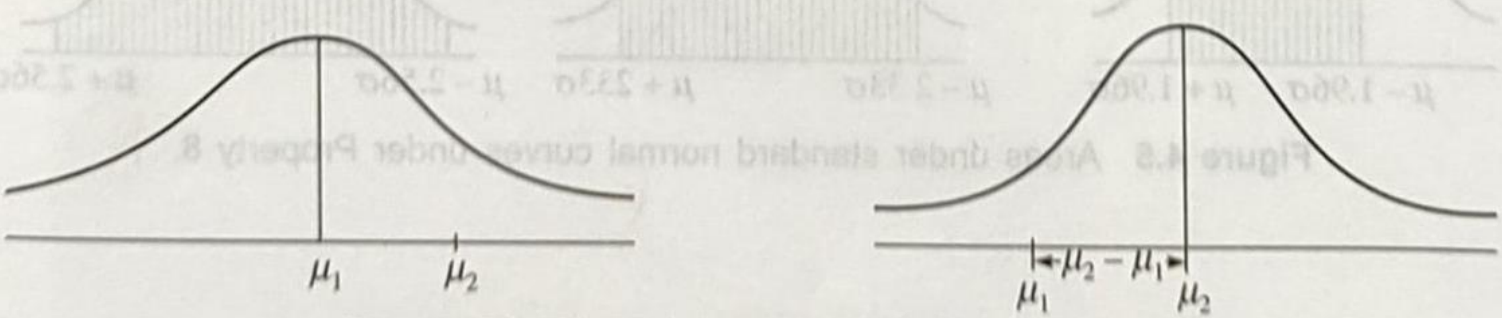



**Figure 4.3** Normal curves with different means.

# Properties of Normal Distribution

6. The normal distribution with $\mu = 0$ and $\sigma = 1$ is called the *standard normal distribution*. Using properties 4 and 5, any normal distribution can be reduced to the standard normal distribution by 'horizontal displacement' (property 4) and 'flattening' (property 5). Mathematically, if $X$ is a normal distribution with mean $\mu$ and standard deviation $\sigma$, then $Z = \dfrac{X-\mu}{\sigma}$ is the standard normal distribution.

*Note:* The simple transformation of a normal variate $X$ into $Z = \dfrac{X-\mu}{\sigma}$ enables us to calculate the probability for a general normal distribution from the probability for the standard normal distribution $Z$. The probability for the standard normal distribution can be calculated using the 'Area table for standard normal distribution' (see Table A2), given in Chapter 6. This will be dealt with in detail shortly. Hereafter, Z will denote the standard normal variable.

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

# Properties of Normal Distribution

7. (a) 68.26% of the values of a normal distribution lie between $\mu - \sigma$ and $\mu + \sigma$ (between –1 and 1 in the case of standard normal distribution).

(b) 95.94% of the values of a normal distribution lie between $\mu - 2\sigma$ and $\mu + 2\sigma$.

(c) 99.72% of the values of a normal distribution lie between $\mu - 3\sigma$ and $\mu + 3\sigma$ (see Figure 4.4).

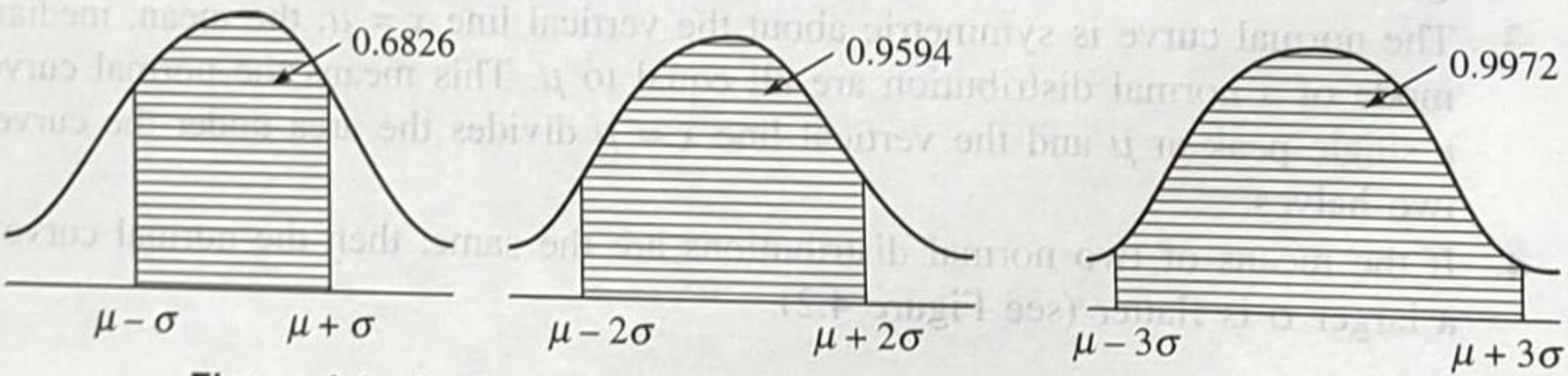



**Figure 4.4** Areas under standard normal curve under Property 7.

# Properties of Normal Distribution

8. (a) 90% of the values of a normal distribution lie between $\mu - 1.65\sigma$ and $\mu + 1.65\sigma$.
   (b) 95% of the values of a normal distribution lie between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$.
   (c) 98% of the values of a normal distribution lie between $\mu - 2.33\sigma$ and $\mu + 2.33\sigma$.
   (d) 99% of the values of a normal distribution lie between $\mu - 2.56\sigma$ and $\mu + 2.56\sigma$ (see Figure 4.5).

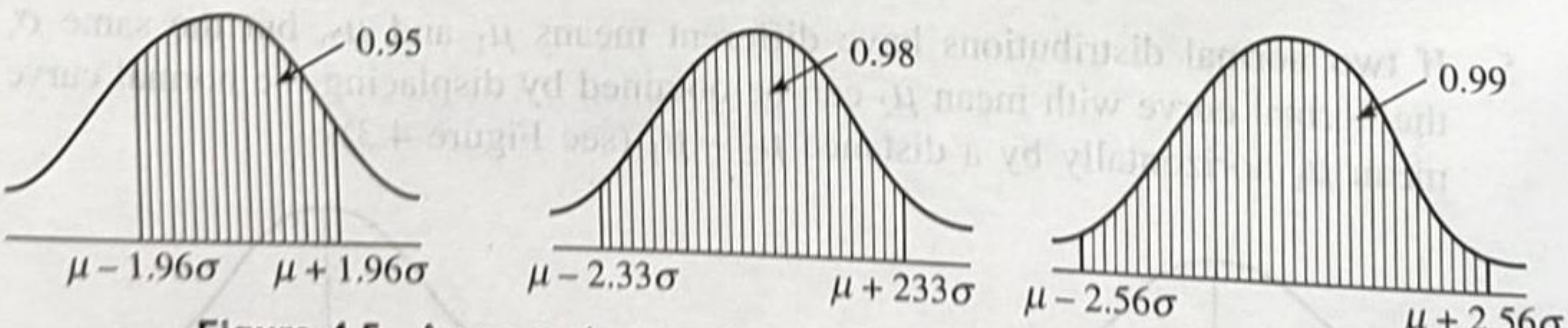
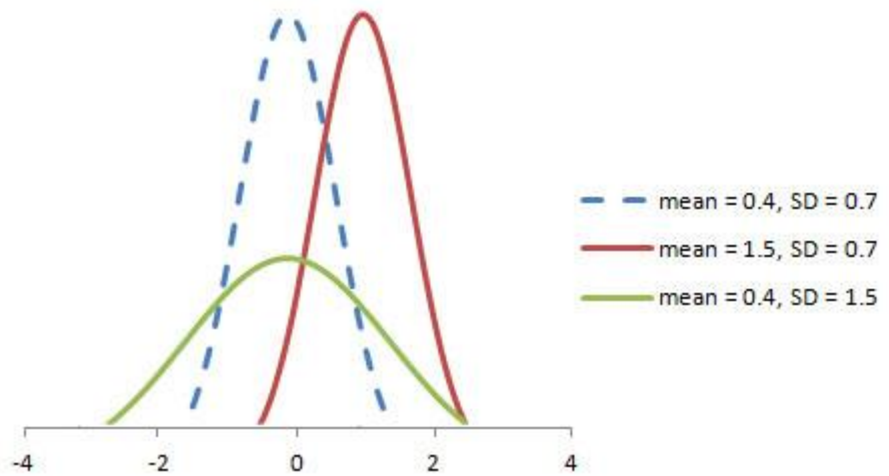
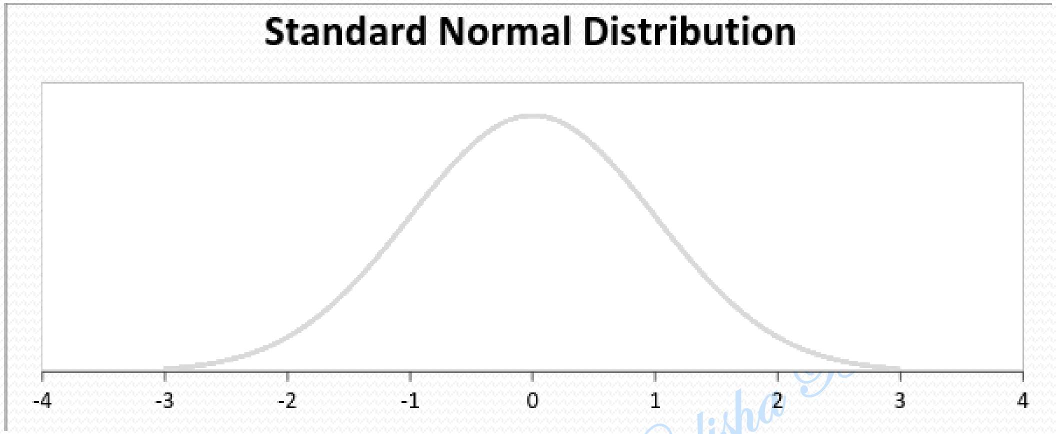



**Figure 4.5** Areas under standard normal curves under Property 8.

- The graph of a random variable X ~ N (μ, σ) is shown below.

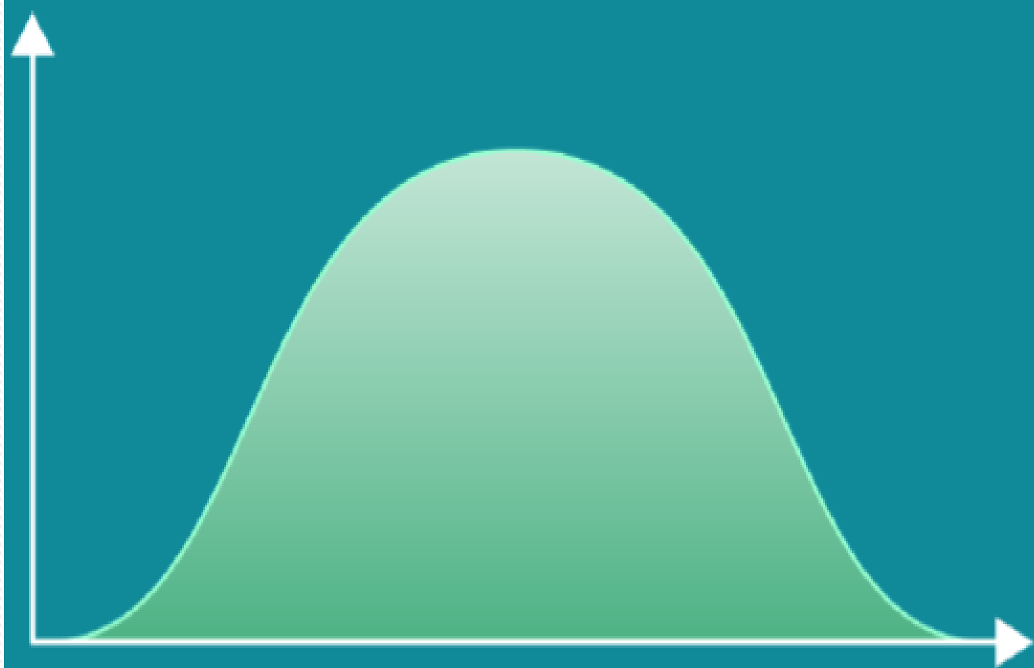



- A standard normal distribution is defined as a distribution with a mean of 0 and a standard deviation of 1. For such a case, the PDF becomes:




$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad for\ -\infty < x < \infty$$

A **Normal Distribution** represents a distribution that most natural events follow.



**Notation:**

- $Y \sim N(\mu, \sigma^2)$

**Key characteristics**

- Its graph is bell-shaped curve, symmetric and has thin tails.
- $E(Y) = \mu$
- $Var(Y) = \sigma^2$
- 68% of all its values should fall in the interval:
  - $(\mu - \sigma,\ \mu + \sigma)$

**Example and uses:**

- Often observed in the size of animals in the wilderness.
- Could be standardized to use the Z-table.

# Standardizing a Normal Distribution

To standardize any normal distribution we need to transform it so that the mean is 0 and the variance and standard deviation are 1.

Ensures mean is 0.

Using a transformation to create a new random variable z.

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

Ensures standard deviation is 1.

Importance of the Standard Normal Distribution.

- The new variable z, represents how many standard deviations away from the mean, each corresponding value is.
- We can transform any Normal Distribution into a Standard Normal Distribution using the transformation shown above.
- Convenient to use because of a table of known values for its CDF, called the Z-score table, or simply the Z-table.

# Example 1

- According to a survey on use of smart phones in India, the smart phone users spend 68 minutes in a day on average in sending messages and the corresponding standard deviation is 12 minutes. Assume that the time spent in sending messages follows a normal distribution.

(a) What proportion of the smart phone users are spending more than 90 minutes in sending messages daily?

(b) What proportion of customers are spending less than 20 minutes?

(c) What proportion of customers are spending between 50 minutes and 100 minutes?

# Solution

It is given that $\mu = 68$ minutes and $\sigma = 12$ minutes.

(a) Proportion of customers spending more than 90 minutes is given by $P(X \geq 90) = 1 - P(X \leq 90) = 1 - F(90)$

The standard normal random variable value for X = 90 is given by

$$Z = \frac{x - \mu}{\sigma} = \frac{90 - 68}{12} = 1.8333$$

That is, F(X = 90) = F(Z = 1.8333). From standard normal distribution table, we get for Z = 1.8333. The area under the standard normal distribution curve is 0.9666. Thus , $P(X \geq 90) = 1 - P(X \leq 90) = 1 - F(90) = 1 - 0.9666 = 0.0334$

# Solution

b) Proportion of customers spending less than 20 minutes is

$P(X \leq 20) = F(20)$

Z = (20-68)/12 = (-48)/12 = -4

That is, $P(X \leq 20)$ = F(Z ≤ -4). From standard normal distribution table, we get for Z = -4.

The area under the standard normal distribution curve is 0.0003.

c) Proportion of customers spending between 50 and 100 minutes is given by P(50 ≤ X ≤100) = F(100)-F(50)

Z1 = (50-68)/12 = (-18)/12 = -1.5

Z2 = (100-68)/12 = 32/12 = 2.67

From standard normal distribution table, we get for Z1 = -1.5 and z2 = 2.67. The area under the standard normal distribution curve is 0.06681 and 0.99609.

F(100)-F(50) = 0.99609 - 0.06681 = 0.9293

# Example 2

- A software company is recruiting fresh engineering graduates during the placement season and training them for providing skills for undertaking projects independently by themselves. The training period for the trainees varies due to their varied educational background and eagerness to learn the programming. From historical data, the company finds that the time of completion of the training program has an average of 4 months and a variance of 2.25 months.

a) What is the probability that a randomly selected graduate completes the training program within 4 months, 5 months and 6 months?

b) Find the probability that the completion time for a random selected graduate varies from 3 to 5 months.

***Solution:*** Let $X$ denote the time of completion of the training programme. We can assume that $X$ follows a normal distribution. For $X$,

$$\mu = 4 \quad \sigma^2 = 2.25 \quad \text{and} \quad \sigma = 1.5$$

(a) We have to find $P(X \leq 4)$, $P(X \leq 5)$ and $P(X \leq 6)$

$$P(X \leq 4) = P\left(Z \leq \frac{4-4}{1.5}\right) = P(Z \leq 0) = 0.5$$

$$P(X \leq 5) = P\left(Z \leq \frac{5-4}{1.5}\right) = P(Z \leq 0.67)$$

The table value for $Z = 0.67$ is 0.2486

So, $P(X \leq 5) = 0.5 + 0.2486 = 0.7486$

$$P(X \leq 6) = P\left(Z \leq \frac{6-4}{1.5}\right) = P(Z \leq 1.33)$$

The table value for $Z = 1.33$ is 0.4082

So, $P(X \leq 6) = 0.5 + 0.4082 = 0.9082$

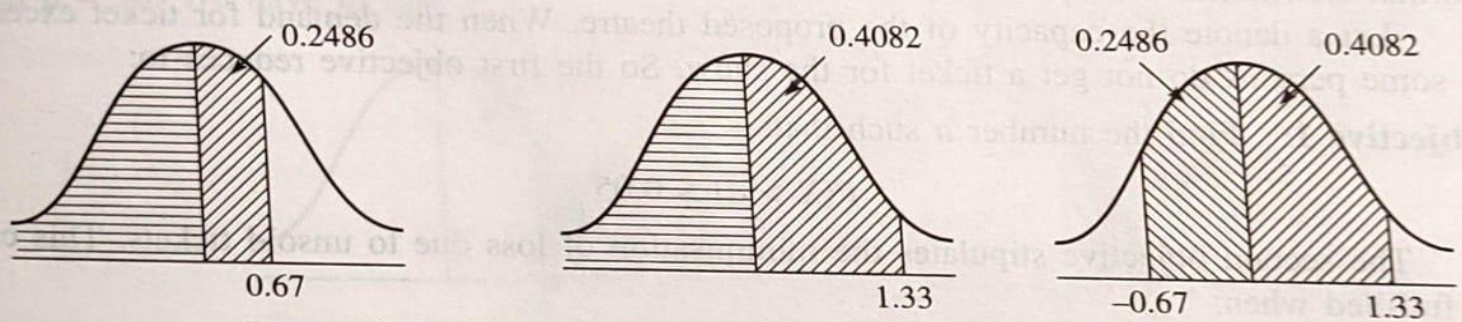



**Figure 4.11** Calculation of areas for Example 4.17(a), and (b)

(b) $P(3 \leq X \leq 5) = P\left(\frac{3-4}{1.5} \leq Z \leq \frac{5-4}{1.5}\right)$

$$= P(-0.67 \leq Z \leq 1.33)$$

$$= 0.2486 + 0.4082 = 0.7568$$

# Example 3

$$P(X = k) = P\left(k - \frac{1}{2} \le Y \le k + \frac{1}{2}\right) \tag{4.15}$$

where, $X$ is binomial with parameters $n$ and $p$ and $Y$ is normal with mean $np$ and standard deviation $\sqrt{npq}$ (when $np, nq \ge 5$)

- A business school knows (from past records) that 10% of their students get placements in MNCs. In the current year, the business school has 400 students. Find the probability that:

a) 35 to 45 students get placement in MNCs
b) Less than 30 students get placement in MNCs
c) More than 48 students get placement in MNCs

# Solution

***Solution:*** Let $X$ denote the number of students among 400 students of the business school who get placement in MNCs. It is a binomial distribution with $n = 400$ and $p = 0.10$.

As $np = 40$ and $nq = 360$ and both are greater than or equal to 5, we can approximate $X$ by a normal distribution $Y$ with $\mu = 400\ (0.10) = 40$ and

$$\sigma = \sqrt{400(0.10)(0.90)} = \sqrt{36} = 6$$

(a) By Eq. (4.15)

$$P(35 \le X \le 45) = P(34.5 \le Y \le 45.5)$$

$$= P\left(\frac{34.5 - 40}{6} \le Z \le \frac{45.5 - 40}{6}\right) = P(-0.92 \le Z \le 0.92)$$

$= 0.3212 + 0.3212$ (see Figure 4.20)

$= 0.6424$

# Solution

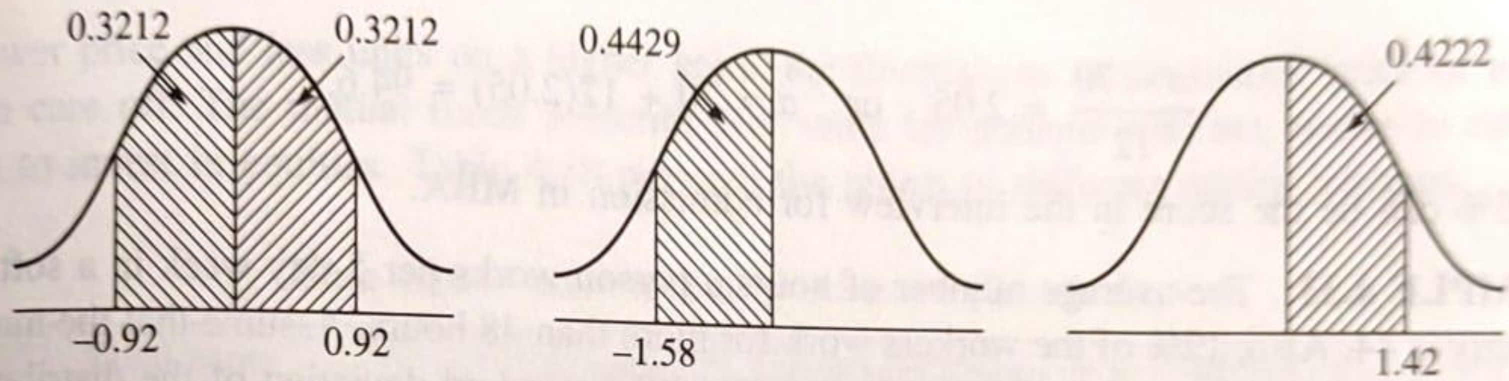



**Figure 4.20** Calculation of probabilities for Example 4.31(a), (b) and (c).

(b) $P(X \leq 30) = P(Y \leq 30.5) = P\left(Z \leq \dfrac{30.5 - 40}{6}\right)$

$= P(Z \leq -1.58) = 0.5 - 0.4429 = 0.0571$

(c) $P(X \geq 48) = P(Y \geq 48.5)$

$= P\left(Z \leq \dfrac{48.5 - 40}{6}\right) = P(Z \leq 1.42)$

$= 0.5 + 0.4222 = 0.9222$

# Exercise 5

- A business school wants to admit only top 2% of the students who were called for the interview and GD. It is found that the marks in the interview follow a normal distribution with a mean of 70 and standard deviation of 12. What marks should a student get in interview so that he gets admission in MBA?

# Exercise 6

- The average number of hours a person works per 5 day week in a software company is 44. Also, 12% of the employees work for more than 48 hours. Assume that the number of hours of work follows a normal distribution. Find the standard deviation of the distribution. What percentage of employees work for less than 40 hours?

# Chi-Squared Distribution

- Chi-square distribution with $k$ degrees of freedom [denoted as $\chi^2(k)$ distribution] is a non-parametric distribution which is obtained by adding square of $k$ independent standard normal random variables.
- Consider a normal random variable $X_1$ with mean $\mu_1$ and standard deviation $\sigma_1$. Then we can define $Z_1$ (the standard normal random variable) as $Z_1 = \frac{X_1 - \mu_1}{\sigma_1}$
- Then,

$$Z_1^2 = \left(\frac{X_1 - \mu_1}{\sigma_1}\right)^2$$

is a chi-square distribution with one degree of freedom [$\chi 2(1)$]

The probability density function of $\chi^2(k)$ is given by

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$$

where $\Gamma(k/2)$ is a Gamma function given by

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

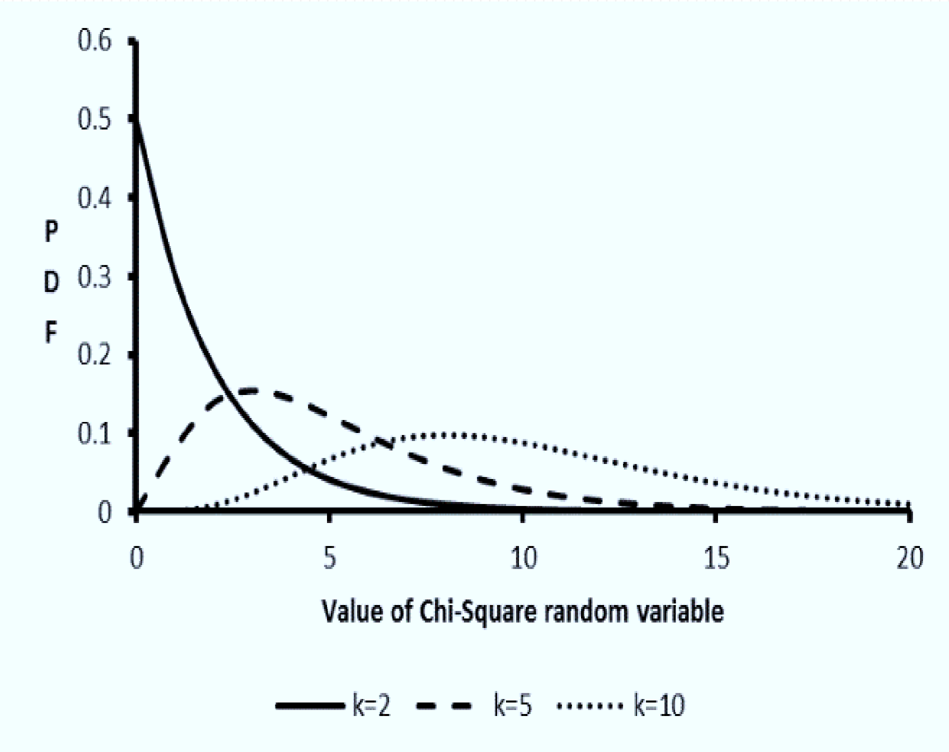
- The cumulative distribution function of a chi-square distribution with *k* degrees of freedom is given by

$$F(x) = \frac{\gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$$

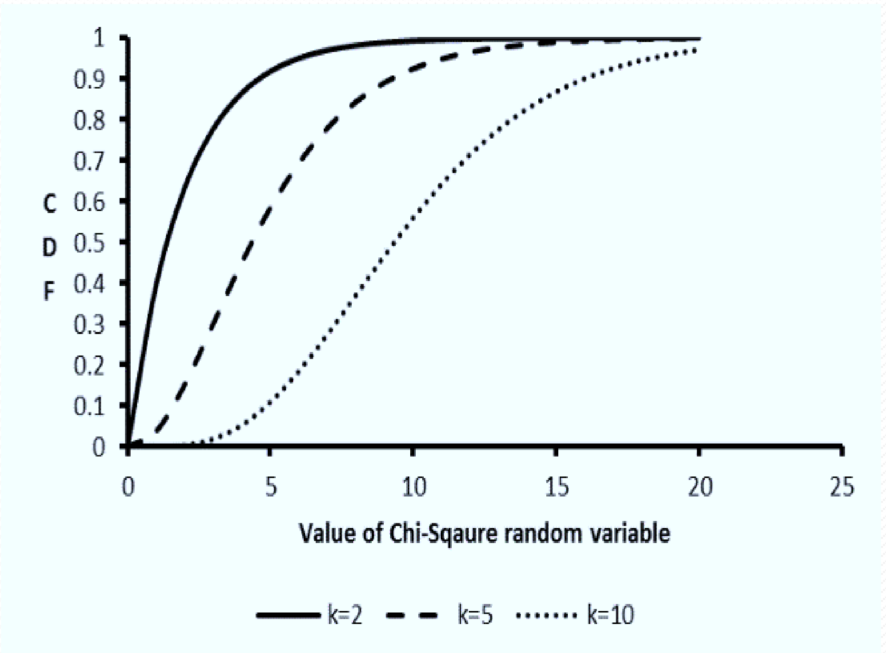
- Where $\gamma\left(\frac{k}{2}, \frac{x}{2}\right)$ is the lower incomplete Gamma function. It is given by

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

## Probability density function of chi-square distribution for different values of *k*




## Cumulative distribution of chi-square distribution with k degrees of freedom



## Properties of chi-square distribution

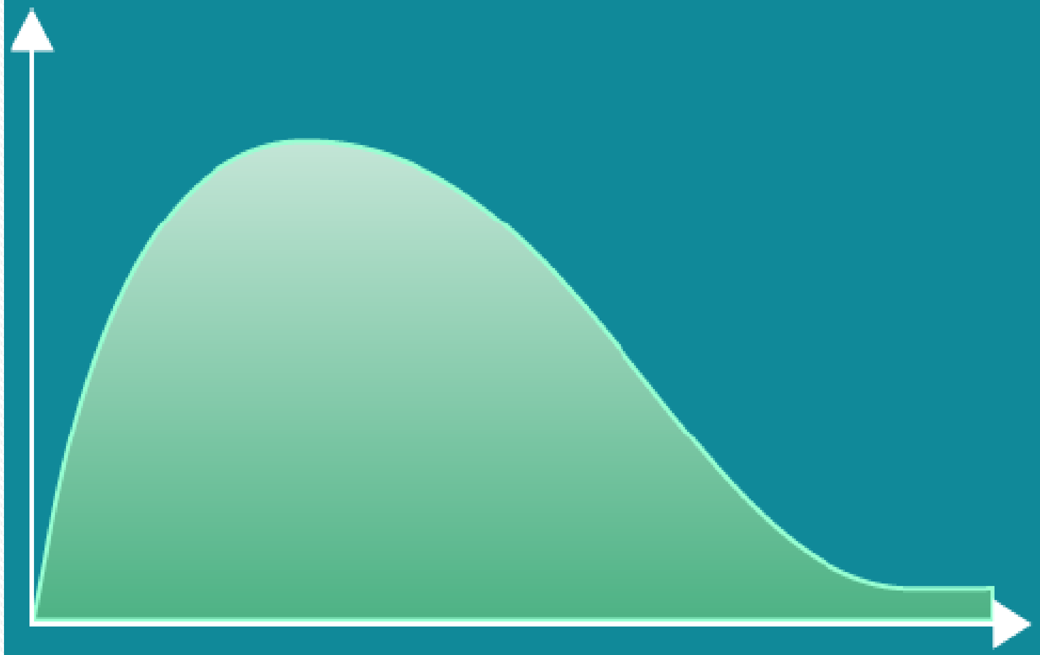
- The mean and standard deviation of a chi-square distribution are k and $\sqrt{2k}$ where k is the degrees of freedom

- As the degrees of freedom k increases the probability density function of a chi-square distribution approaches normal distribution.

- Chi-square goodness of fit test is one of the popular tests for checking whether a data follows a specific probability distribution.

# Chi-Squared Distribution

A Chi-Squared distribution is often used.



Notation:

- $Y \sim \chi^2(k)$

Key characteristics

- Its graph is asymmetric and skewed to the right.
- $E(Y) = k$
- $Var(Y) = 2k$
- The Chi-Squared distribution is the square of the t-distribution.

Example and uses:

- Often used to test goodness of fit.
- Contains a table of known values for its CDF called the $\chi^2$-table. The only difference is the table shows what part of the table

# Exponential Distribution

- Let's consider the call center example one more time. What about the interval of time between the calls? Here, the exponential distribution comes to our rescue. Exponential distribution models the interval of time between the calls.
- Other examples are:
  - Length of time between metro arrivals
  - Length of time between arrivals at a gas station
  - The life of an air conditioner
  - The exponential distribution is widely used for survival analysis. From the expected life of a machine to the expected life of a human, exponential distribution successfully delivers the result.
- A random variable X is said to have an **exponential distribution** with PDF:
- $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$ (PDF), $F(x) = 1 - e^{-\lambda x}$ (CDF)

# Exponential Distribution

- And parameter **$\lambda>0$,** which is also called the rate.
- For survival analysis, $\lambda$ is called the failure rate of a device at any time t, given that it has survived up to t.
- Mean and Variance of a random variable X following an exponential distribution:
- **Mean -> $E(X) = 1/\lambda$**
- **Variance -> $Var(X) = (1/\lambda)^2$**
- Also, the greater the rate, the faster the curve drops, and the lower the rate, the flatter the curve. This is explained better with the graph shown below.
- To ease the computation, there are some formulas given below:
  - $P\{X\leq x\} = 1 - e^{-\lambda x}$ corresponds to the area under the density curve to the left of x
  - $P\{X>x\} = e^{-\lambda x}$ corresponds to the area under the density curve to the right of x
  - $P\{x_1<X\leq x_2\} = e^{-\lambda x_1} - e^{-\lambda x_2}$, corresponds to the area under the density curve between x1 and x2.

The **Exponential Distribution** is usually observed in events which significantly change early on.


**Notation:**

- $Y \sim Exp\ (\lambda)$

**Key characteristics**

- Both the PDF and the CDF plateau after a certain point.
- $E(Y) = \frac{1}{\lambda}$
- $Var(Y) = \frac{1}{\lambda^2}$
- We often use the natural logarithm to transform the values of such distributions since we do not have a table of known values like the Normal or Chi-Squared.

**Example and uses:**

- Often used with dynamically changing variables, like online website traffic or radioactive decay.

# Example

The time to failure of an avionic system follows an exponential distribution with a mean time between failures (MTBF) of 1000 hours.

(a) Calculate the probability that the system will fail before 1000 hours.

(b) Calculate the probability that it will not fail up to 2000 hours.

(c) Calculate the time by which 10% of the systems will fail (that is calculate $P_{10}$ life)

# Solution

(a) The probability that the system will fail by 1000 hours is

In this case $F(1000) = 1 - e^{\lambda t}$ so , $F(1000) = 1 - e^{-\frac{1}{1000}\times 1000} = 1 - e^{-1} = 0.6321$

(b) The probability that the system will not fail up to 2000 hours is

$$P(X > 2000) = 1 - P(X \le 2000) = 1 - F(t) = e^{-\lambda t} = e^{-\frac{1}{1000}\times 2000} = e^{-2} = 0.1353$$

(c) The time by which 10% of the systems will fail is

$$F(t) = 0.10 \Rightarrow 1 - e^{-\lambda t} = 0.1 \rightarrow e^{-\lambda t} = 0.9$$

So ,

$$t = -\left(\frac{1}{\lambda}\right)\ln(0.9) = -1000 \times \ln(0.9) = 105.61$$ hours

That is, by 105.61 hours, 10% of items will fail.