# Project Document

Topic: - Probability

Created by



Under The Guidance of  
Prof. Tejal Nagar

Table of Content: -

1. Introduction to Probability
2. Basic Probability Concepts
3. Probability Rules
4. Probability Distributions
5. Conditional Probability
6. Bayes' Theorem
7. Expected Value
8. Variance and Standard Deviation
9. Probability Models
10. Law of Large Numbers
11. Central Limit Theorem

# Introduction to Probability

Probability is a measure of the likelihood of an event to occur. Many events cannot be predicted with total certainty. We can predict only the chance of an event to occur i.e., how likely they are going to happen, using it.

Probability can range from 0 to 1, where 0 means the event to be an impossible one and 1 indicates a certain event. Probability for Class 10 is an important topic for the students which explains all the basic concepts of this topic. **The probability of all the events in a sample space adds up to 1.**

**For example**, when we toss a coin, either we get Head OR Tail, only two possible outcomes are possible (H, T). But when two coins are tossed then there will be four possible outcomes,  i.e {(H, H), (H, T), (T, H), (T, T)}.

The probability formula is defined as the possibility of an event to happen is equal to the ratio of the number of favourable outcomes and the total number of outcomes.

Formality for Probability :-   
**Probability of event to happen P(E) = Number of favourable outcomes/Total Number of outcomes**

There are three major types of probabilities:

* Theoretical Probability
* Experimental Probability
* Axiomatic Probability

**Theoretical Probability: -**It is based on the possible chances of something to happen. The theoretical probability is mainly based on the reasoning behind probability. For example, if a coin is tossed, the theoretical probability of getting a head will be ½.

**Experimental Probability: -**It is based on the basis of the observations of an experiment. The experimental probability can be calculated based on the number of possible outcomes by the total number of trials. For example, if a coin is tossed 10 times and head is recorded 6 times then, the experimental probability for heads is 6/10 or, 3/5.

**Axiomatic Probability: -**In axiomatic probability, a set of rules or axioms are set which applies to all types. These axioms are set by Kolmogorov and are known as **Kolmogorov’s three axioms.**With the axiomatic approach to probability, the chances of occurrence or non-occurrence of the events can be quantified. The axiomatic probability lesson covers this concept in detail with Kolmogorov’s three rules (axioms) along with various examples.

**Probability of an Event**  
Assume an event E can occur in r ways out of a sum of n probable or possible **equally likely ways.** Then the probability of happening of the event or its success is expressed as;

P(E) = r/n  
The probability that the event will not occur or known as its failure is expressed as:

P(E’) = (n-r)/n = 1-(r/n)  
E’ represents that the event will not occur.

Therefore, now we can say;  
**P(E) + P(E’) = 1**

# Basic Probability Concepts

A probability is a number that reflects the chance or likelihood that a particular event will occur. Probabilities can be expressed as proportions that range from 0 to 1, and they can also be expressed as percentages ranging from 0% to 100%. A probability of 0 indicates that there is no chance that a particular event will occur, whereas a probability of 1 indicates that an event is certain to occur. A probability of 0.45 (45%) indicates that there are 45 chances out of 100 of the event occurring.

The concept of probability can be illustrated in the context of a study of obesity in children 5-10 years of age who are seeking medical care at a particular paediatric practice. The population (sampling frame) includes all children who were seen in the practice in the past 12 months and is summarized below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Age (years)** | | | | | | |
|  | 5 | 6 | 7 | 8 | 9 | 10 | total |
| **Boys** | 432 | 379 | 501 | 410 | 420 | 418 | 2,560 |
| **Girls** | 408 | 513 | 412 | 436 | 461 | 500 | 2,730 |
| **Totals** | 840 | 892 | 913 | 846 | 881 | 918 | 5,290 |

**There are two types of probability**

1. **Unconditional Probability**
2. **Conditional Probability**

**Unconditional Probability**

If we select a child at random (by simple random sampling), then each child has the same probability (equal chance) of being selected, and the probability is 1/N, where N=the population size. Thus, the probability that any child is selected is 1/5,290 = 0.0002. In most sampling situations we are generally not concerned with sampling a specific individual but instead we concern ourselves with the probability of sampling certain types of individuals. For example, what is the probability of selecting a boy or a child 7 years of age? The following formula can be used to compute probabilities of selecting individuals with specific attributes or characteristics.

**P(characteristic) = # persons with characteristic / N**

**Conditional Probability**

Each of the probabilities computed in the previous section (e.g., P(boy), P(7 years of age)) is an unconditional probability, because the denominator for each is the total population size (N=5,290) reflecting the fact that everyone in the entire population is eligible to be selected. However, sometimes it is of interest to focus on a particular subset of the population (e.g., a sub-population). For example, suppose we are interested just in the girls and ask the question, what is the probability of selecting a 9 year old from the sub-population of girls? There is a total of NG=2,730 girls (here NG refers to the population of girls), and the probability of selecting a 9 year old from the sub-population of girls is written as follows:

**P(9 year old | girls) = # persons with characteristic / N**

# Probability Rules

1) Possible values for probabilities range from 0 to 1

0 = impossible event  
1 = certain event

2) The sum of all the probabilities for all possible outcomes is equal to 1. Note the connection to the **complement rule.**

3) **Addition Rule** –

The probability that one or both events occur mutually exclusive events:   
P(A or B) = P(A) + P(B)

not mutually exclusive events:   
P(A or B) = P(A) + P(B) - P(A and B)

4) **Multiplication Rule** - the probability that **both**events occur together independent events:   
  
P(A and B) = P(A) \* P(B)  
P(A and B) = P(A) \* P(B|A)

5) **Conditional Probability** - the probability of an event happening **given** that another event has already happened

P(A|B) = P(A and B) / P(B)  
\*Note the line | means "given" while the slash / means divide

# Probability Distributions

**What is Probability Distribution?**

Probability distribution yields the possible outcomes for any random event. It is also defined based on the underlying sample space as a set of possible outcomes of any random experiment. These settings could be a set of real numbers or a set of vectors or a set of any entities. It is a part of probability and statistics.

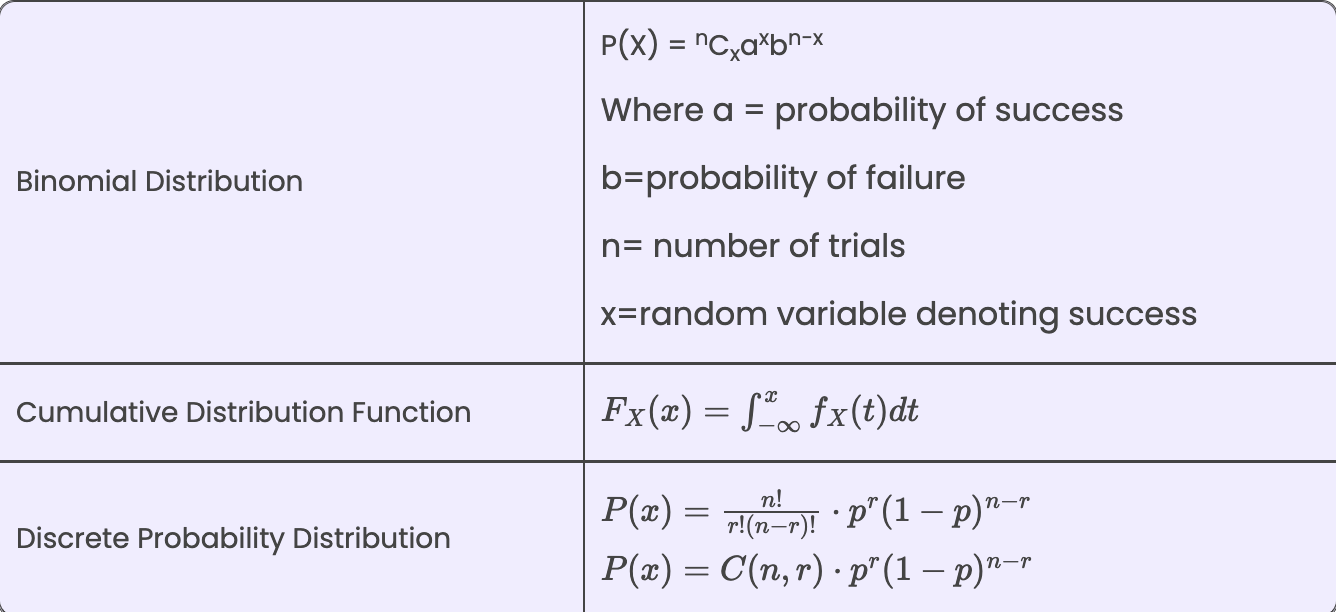
Random experiments are defined as the result of an experiment, whose outcome cannot be predicted. Suppose, if we toss a coin, we cannot predict, what outcome it will appear either it will come as Head or as Tail. The possible result of a random experiment is called an outcome. And the set of outcomes is called a sample point. With the help of these experiments or events, we can always create a probability pattern table in terms of variables and probabilities.

**Probability Distribution of Random Variables**

A random variable has a probability distribution, which defines the probability of its unknown values. Random variables can be discrete (not constant) or continuous or both. That means it takes any of a designated finite or countable list of values, provided with a probability mass function feature of the random variable’s probability distribution or can take any numerical value in an interval or set of intervals. Through a probability density function that is representative of the random variable’s probability distribution or it can be a combination of both discrete and continuous.

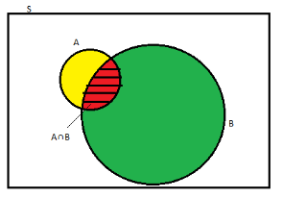
Two random variables with equal probability distribution can yet vary with respect to their relationships with other random variables or whether they are independent of these. The recognition of a random variable, which means, the outcomes of randomly choosing values as per the variable’s probability distribution function, are called **random variates.**

**Probability Distribution Formulas**



# Conditional Probability

The probability of occurrence of any event A when another event B in relation to A has already occurred is known as conditional probability. It is depicted by P(A|B).



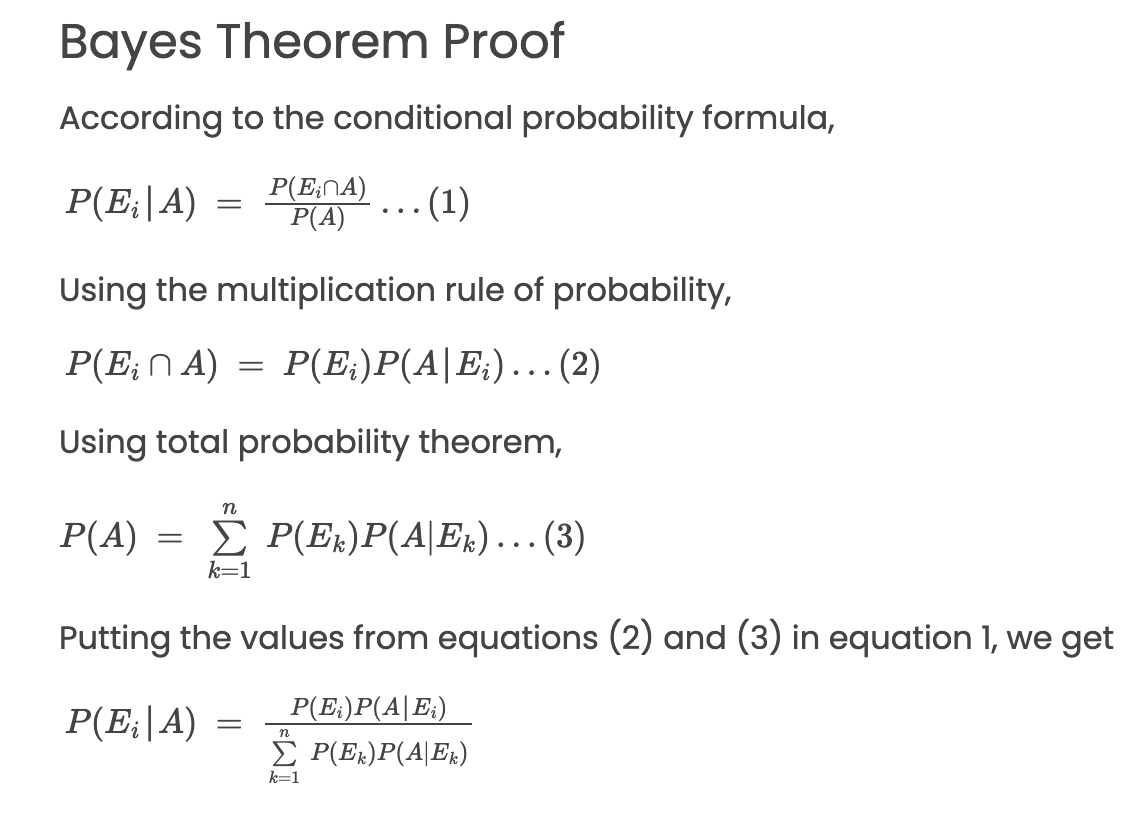
As depicted by the above diagram, sample space is given by S, and there are two events A and B. In a situation where event B has already occurred, then our sample space S naturally gets reduced to B because now the chances of occurrence of an event will lie inside B.

As we have to figure out the chances of occurrence of event A, only a portion common to both A and B is enough to represent the probability of occurrence of A, when B has already occurred. The common portion of the events is depicted by the intersection of both the events A and B, i.e. A ∩ B.

This explains the concept of conditional probability problems, i.e. occurrence of any event when another event in relation to has already occurred.

# Bayes' Theorem

**Bayes’ theorem** describes the probability of occurrence of an event related to any condition. It is also considered for the case of [conditional probability](https://byjus.com/maths/conditional-probability-and-conditional-probability-examples/). Bayes theorem is also known as the formula for the probability of “causes”. For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability. In this article, let us discuss the statement and proof for Bayes theorem, its derivation, formula, and many solved examples.

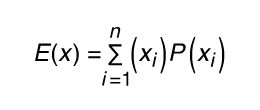


# Expected Value

**Expected value**, in general, the [value](https://www.britannica.com/dictionary/value) that is most likely the result of the next repeated trial of a statistical experiment. The probability of all possible outcomes is factored into the calculations for expected value in order to determine the expected outcome in a random trial of an experiment. Expected value uses all possible outcomes and their probabilities of occurring to find the weighted average of the data in the data set.

For a game spinner that has 8 equal segments made up of 1 yellow segment, 2 blue segments, 1 red segment, and 4 green segments, the “expected value” would correspond to which colour is statistically most likely to be the result of a single spin. Since 4 of the 8 segments are green, there is a 50 percent [chance](https://www.britannica.com/science/likelihood) the spinner will land on green, which is the colour most likely to occur on a single spin.

When the experiment involves numerical [data](https://www.britannica.com/dictionary/data), the expected value is found by calculating the weighted value from the data using the formula



in which *E(x)* represents the expected value, *xi* represents the event, and *P(xi)* represents the probability of the event. The sigma sign in this formula represents the sum of all events multiplied by their individual probabilities. It is necessary that the sum of all probabilities be equal to 1. It should also be noted that if the probability of each event is the same, the weighted value is the equivalent of the mean value of the data.

# Variance and Standard Deviation

**Variance**

According to layman’s words, the variance is a measure of how far a set of data are dispersed out from their mean or average value. It is denoted as ‘σ2’.

**Properties of Variance**

* It is always non-negative since each term in the variance sum is squared and therefore the result is either positive or zero.
* Variance always has squared units. For example, the variance of a set of weights estimated in kilograms will be given in kg squared. Since the population variance is squared, we cannot compare it directly with the mean or the data themselves.

**Standard Deviation**

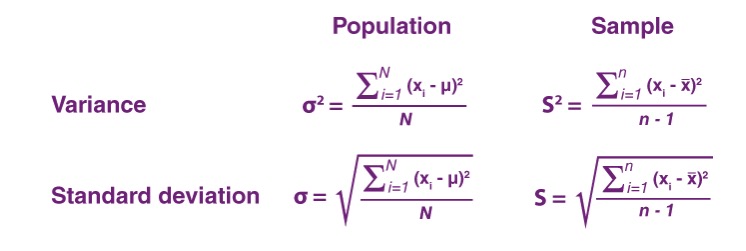
The spread of statistical data is measured by the standard deviation. Distribution measures the deviation of data from its mean or average position. The degree of dispersion is computed by the method of estimating the deviation of data points. It is denoted by the symbol, ‘σ’.

**Properties of Standard Deviation**

* It describes the square root of the mean of the squares of all values in a data set and is also called the root-mean-square deviation.
* The smallest value of the standard deviation is 0 since it cannot be negative.
* When the data values of a group are similar, then the standard deviation will be very low or close to zero. But when the data values vary with each other, then the standard variation is high or far from zero.

**Variance and Standard Deviation Formula**

As discussed, the variance of the data set is the average square distance between the mean value and each data value. And standard deviation defines the spread of data values around the mean.



# Probability Models

Building a probability model involves a few simple steps.

First, you identify the *random variables* of interest in your system. A random variable is just *a numerical summary of an uncertain outcome.*

* In our airline example, we could have any possible combination of passengers fail to show up (seat 2C, 14G, etc). But at the end of the day, if we want to know whether any passengers are likely to get bumped to the next flight, all we care about is *how many* ticketed passengers are no-shows, not their specific identities or seat numbers. So that’s our numerical summary, i.e. our random variable: X� = the number of no-shows.
* Or in the soccer game between Arsenal and Man City, there are two obvious numerical summaries: X1�1 = the number of goals scored by Arsenal, and X2�2 = the number of goals scored by Man City.

Second, you identify the set of possible outcomes for your random variable, which we refer to as the *sample space*. In our airline example, the sample space is the set of whole numbers between 0 and 140 (the maximum number of no-shows possible, because that’s how many tickets were sold).

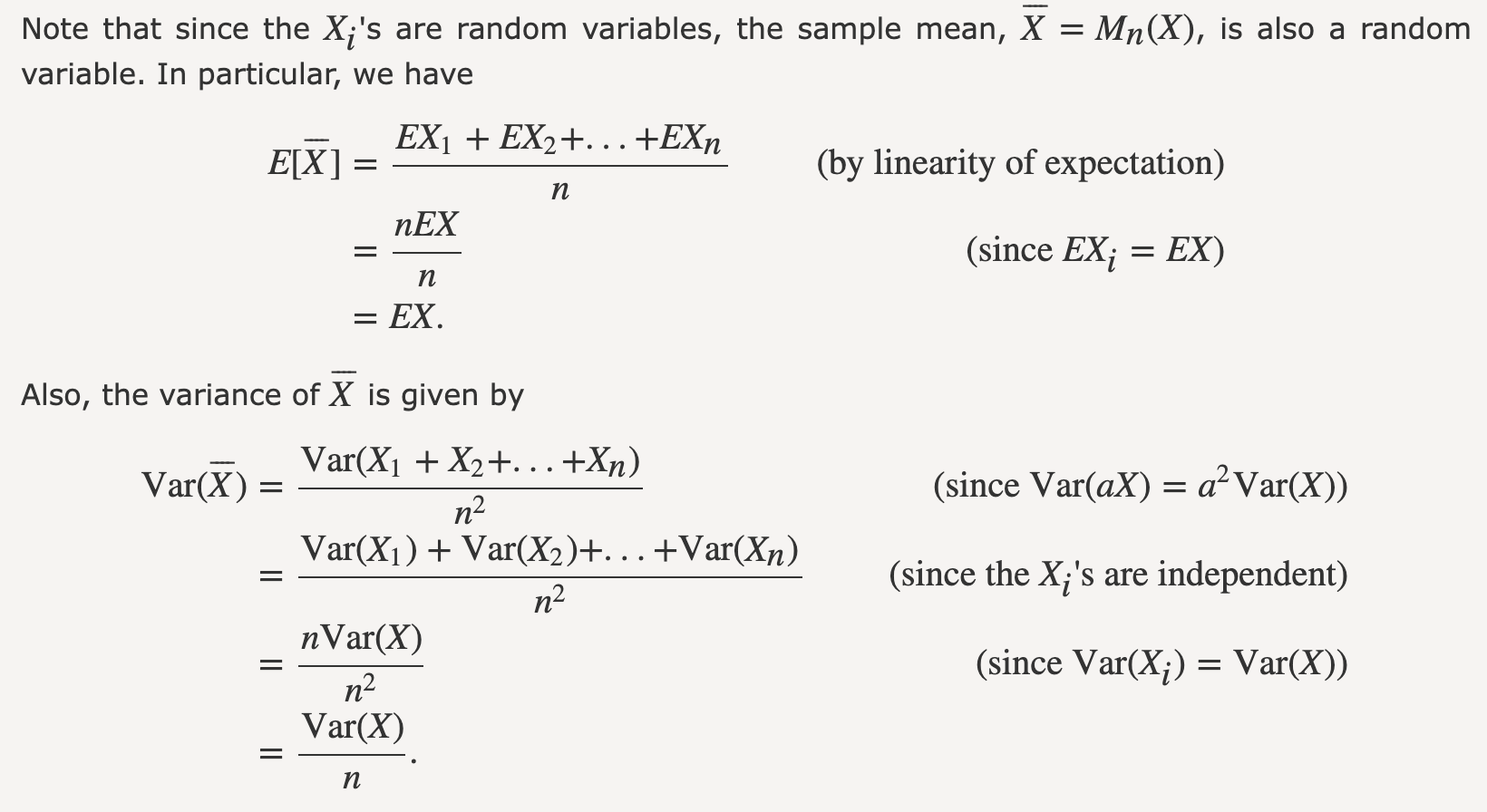
Finally, you provide a *probability distribution*, which is a rule for calculating probabilities associated with each possible outcome in the sample space. In the airline example, this distribution might be described using a simple lookup table based on historical data, e.g. 1% of all flights have 1 no-show, 1.2% have 2 no-shows, 1.7% have 3 no-shows, and so forth. In building a probability model, this final step is usually where the action is, and it’s what we’ll discuss extensively in this lesson.

There are two common types of random variables, corresponding to two common types of outcomes.

* **Discrete**: the sample space consists of whole numbers (0, 1, 2, 3, etc.). Both the number of airline no-shows and the score of a soccer game are discrete random variables: you can’t have 2.4 no-shows or 3.7 goals.
* **Continuous**: the random variable could be anything within a continuous range of numbers, like the price of Apple stock tomorrow, or the volume of a subsurface oil reservoir.

# Law of Large Numbers

The **law of large numbers** has a very central role in probability and statistics. It states that if you repeat an experiment independently a large number of times and average the result, what you obtain should be close to the expected value. There are two main versions of the law of large numbers. They are called the **weak** and **strong** laws of the large numbers. The difference between them is mostly theoretical. In this section, we state and prove the weak law of large numbers (WLLN). The strong law of large numbers is discussed in Section [7.2](https://www.probabilitycourse.com/chapter7/7_2_0_convergence_of_random_variables.php). Before discussing the WLLN, let us define the *sample mean*.



# Central Limit Theorem

The central limit theorem relies on the concept of a **sampling distribution**, which is the [probability distribution](https://www.scribbr.com/statistics/probability-distributions/) of a **statistic**for a large number of [samples](https://www.scribbr.com/methodology/population-vs-sample/) taken from a population.

Imagining an experiment may help you to understand sampling distributions:

* Suppose that you draw a [random sample](https://www.scribbr.com/methodology/simple-random-sampling/) from a population and calculate a [statistic](https://www.scribbr.com/statistics/parameter-vs-statistic/) for the sample, such as the mean.
* Now you draw another random sample of the same size, and again calculate the [mean](https://www.scribbr.com/statistics/mean/).
* You repeat this process many times, and end up with a large number of means, one for each sample.

The distribution of the sample means is an example of a **sampling distribution.**

The central limit theorem says that the sampling distribution of the mean will always be **normally distributed**, as long as the sample size is large enough. Regardless of whether the population has a normal, Poisson, binomial, or any other distribution, the sampling distribution of the mean will be normal.

A normal distribution is a symmetrical, bell-shaped distribution, with increasingly fewer observations the further from the center of the distribution.

**Central limit theorem formula**

Fortunately, you don’t need to actually repeatedly sample a population to know the shape of the sampling distribution. The [parameters](https://www.scribbr.com/statistics/parameter-vs-statistic/) of the sampling distribution of the mean are determined by the parameters of the population:

* The [mean](https://www.scribbr.com/statistics/mean/) of the sampling distribution is the mean of the population.

\begin{equation*}\mu_{\bar{x}}=\mu\end{equation*}

* The [standard deviation](https://www.scribbr.com/statistics/standard-deviation/) of the sampling distribution is the standard deviation of the population divided by the square root of the sample size.

\begin{equation*}\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}\end{equation*}

We can describe the sampling distribution of the mean using this notation:

\begin{equation*}\bar{X}\sim N (\mu,\dfrac{\sigma}{\sqrt{n}})\end{equation*}

Where:

* X̄ is the sampling distribution of the sample means
* ~ means “follows the distribution”
* N is the [normal distribution](https://www.scribbr.com/statistics/normal-distribution/)
* µ is the mean of the population
* σ is the standard deviation of the population
* n is the sample size

**References Links:-**

1. <https://www.scribbr.com/statistics/>
2. <https://byjus.com/>
3. <https://www.investopedia.com/>
4. <https://www.geeksforgeeks.org/>
5. <https://www.cuemath.com/data/probability-rules/>
6. <https://online.stat.psu.edu/stat800/lesson/general-probability-rules>
7. <https://resources.nu.edu/statsresources/BasicProbabilityRules>
8. <https://stattrek.com/probability/probability-rules#google_vignette>
9. <https://www.khanacademy.org/math/statistics-probability/probability-library/conditional-probability-independence/v/calculating-conditional-probability#:~:text=Conditional%20probabilities%20are%20written%20like,P(B%7CA)>.
10. <https://study.com/academy/lesson/basic-probability-theory-rules-formulas.html>
11. <https://www.youtube.com/>
12. <https://chat.openai.com/>