

# Probability Theory For Machine Learning (Part 1)

Related tags: [mathematical-algorithms](#) [math-for-computer-science](#)

Data is one of the essential ingredients for building the best machine learning models. The more you know about the data, the better your machine learning model will be, as you will be able to depict the reason behind your model's performance. Probability is one of the most important mathematical tools that help in understanding different data patterns. Famous algorithms in machine learning like Naive Bayes are completely derived from the probability theory. Hence, knowing probability basics will always be considered to be the best to start the ML journey.

In this article, we will not be only describing the theoretical aspects of probability, but we will give you a sense of **where those theoretical aspects will be used in ML**. So, let's start without any further delay.

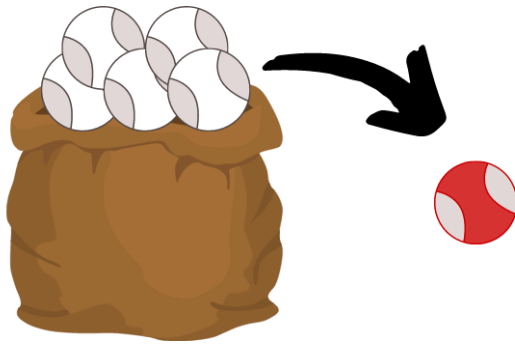
If we have to define "probability",

Originated from the "*Games of Chance*," probability in itself is a branch of mathematics concerned about how likely it is that a proposition is true.

In a more layman manner, *Probability is simply a Possibility of occurring a random event*. For example, *what is the possibility of having rain tomorrow?* The values of probability can only lie between 0 and 1, with 0 and 1 inclusive.

If we notice carefully, every daily-life phenomenon can only be of two types:

**Deterministic:** Phenomena that will always be true. For example, picking a white ball from a bag of white balls. Here probability will be either 0 or 1.



A bag full of  
only white balls

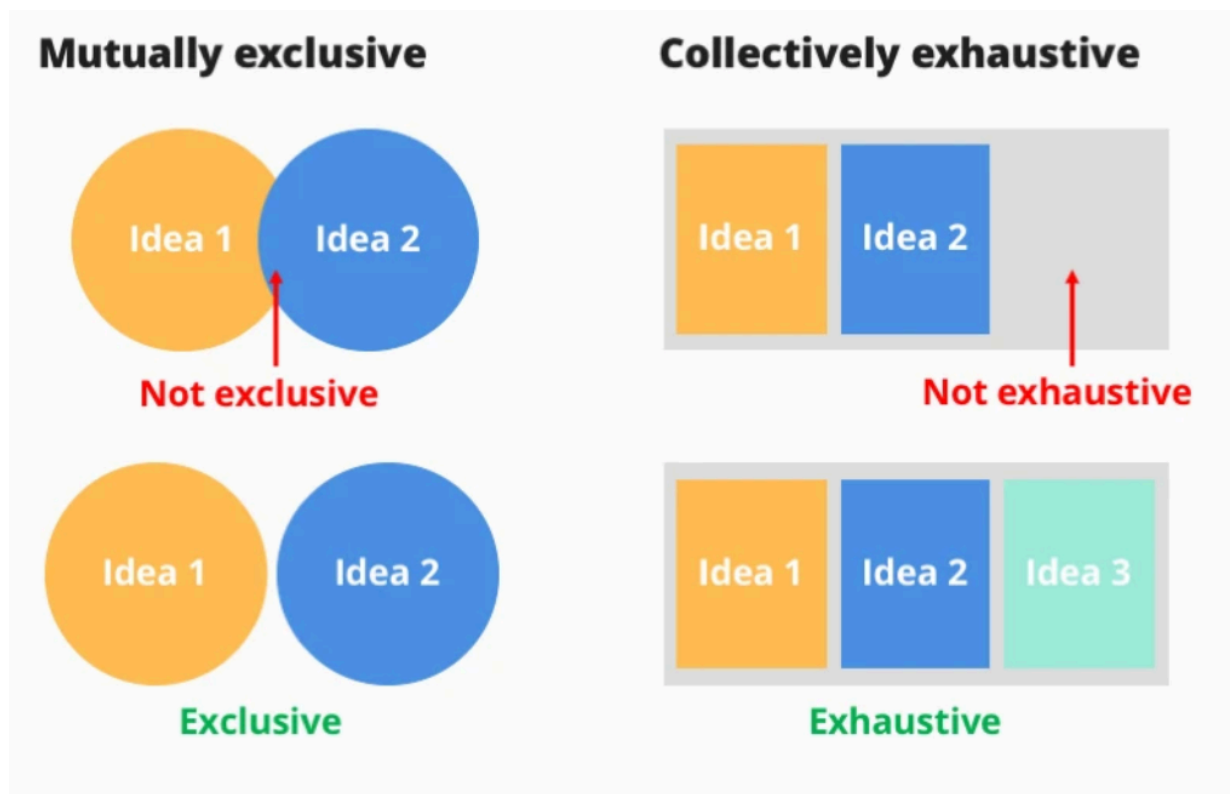
What is the probability  
of extracting red ball  
from the bag containing  
only white balls?

**Indeterministic:** Phenomena for which we are not sure. For example, if a fair dice is rolled, the probability of the number being 1 is indeterministic.

### Relationship between events

There are two famous terms when we define the relationship between events from the same experiment.

1. **Exclusive:** If the two events can never happen at the same time. For example, flipping a fair coin once can not lead to the scenario where Heads and Tails both come up.
2. **Exhaustive:** If the probability of happening different events collectively captures all possible events that can happen. For example, in an experiment of flipping the coin, the probability of getting Heads and the probability of getting Tails capture all possible outcomes from that experiment.



### IGotAnOffer

**Mathematically, probability can be defined as:**

If a random experiment has  $n > 0$  mutually exclusive, exhaustive, and equally likely events and, if out of this  $n$ ,  $m$  such events are favorable ( $m \geq 0$  and  $n \geq m$ ), then the probability of occurrence of any event  $E$  can be defined as

$$P(E) = \frac{\text{Number of favourable events}(m)}{\text{Number of exhaustive events}(n)} ; 0 \leq P(E) \leq 1$$

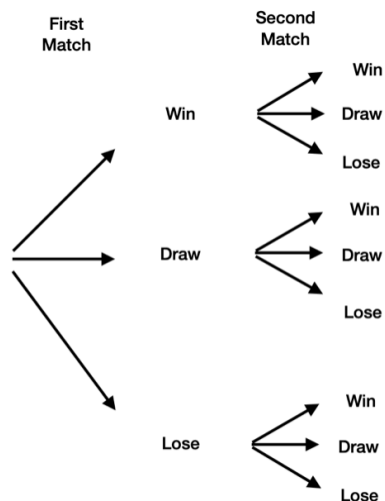
**Some common terms:**

**Random experiment:** Experiments or processes where the outcome is not predicted with certainty, like throwing a die (we can get any one of these 1,2,3,4,5,6).

**Trials and Events:** The occurrence of any event in a random experiment is a trial and, the outcomes of the random experiment are events. Throwing dice is a trial, and if it shows any number from 1 to 6, then that number is an event.

**Multiplication Rule:** If an experiment has multiple components and 1st component has **K1** possible outcomes, the 2nd component has **K2** possible

outcomes, . . . , and the **Kth** component has **Kr** possible outcomes, then overall there are **K1\*K2 \* . . . \*Kr** possibilities for the whole experiment.



**Sampling:** A technique in which samples are chosen from a larger population. The sample will be considered as a probability sample if it has been chosen from a random selection.

**Exhaustive Event:** Total number of all possible events of a random experiment. Like the tossing of a die once has 6 exhaustive events. If the die is tossed for **n** number of times, there will be **6^n** exhaustive events.

**Mutually Exclusive Events:** Events where the probability of any event's occurrence does not depend upon the probability of occurrence of another event in the same or different trails. Like the occurrence of 5 and 6 when two dice are rolled simultaneously.

**Equally likely events:** The events where one cannot expect in preference of another event in a random experiment like rolling of a fair die twice.

**Independent Events:** The events A and B are said to be independent events if the probability of occurrence of A doesn't depend upon the probability of occurrence of B. Mathematically, we can say that independent events must follow:

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

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

**Conditional Independent Events:** Based on some event  $C$ ,  $A$  &  $B$  events are conditional independent if  $P(A \cap B | C) = P(A | C) P(B | C)$ .

Conditional independence does not imply independence, and independence does not imply conditional independence. Try to think over this.

**De Morgan's Law:**  $\bar{A}$  is the complement of  $A$ , i.e.,  $A$  and  $\bar{A}$  are exhaustive events.  $\cup$  is the union of events and  $\cap$  is the intersection of events.

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

### Some Basic results:

Let  $A$  and  $B$  are two events;  $\bar{A}$  is the complement of  $A$ , then.

1.  $P(\bar{A}) = 1 - P(A)$
2.  $P(\bar{A} \cap B) = P(B) - P(A \cap B)$
3.  $P(A \cap \bar{B}) = P(A) - P(A \cap B)$
4.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5.  $P(\emptyset) = 0$

### Probability Under Statistical Independence:

Suppose two coins are to be tossed, then the probability of occurrences of the head or tail can be classified as:

1. **Marginal Probability:** The simple probability of occurrence of head or tail on tossing of a coin. (Simple probabilities of occurrence of any event).
2. **Joint Probability:** Probability of occurrence of the head with the first coin and probability of occurrence of the tail with the second coin when both the coins are tossed simultaneously. **OR** When a single coin is tossed consecutively, the probability of occurrence of the tail in the first chance and head in the second chance. (Probability of occurrence of joint events occurring together or in succession).

3. **Conditional Probability:** Probability of occurrence of a head in tossing a coin when the tail has already occurred. (Probability of occurrence of any event **A** when **B** has already occurred).

### Probability under Statistical Dependence:

When the probability of one event's occurrence depends on the probability of another event's occurrence, that scenario comes under **statistical dependence**.

If we have two events, **A** and **B**, then:

1. **Conditional Probability** is the probability of occurrence of an event **A** if event **B** has already occurred.

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} ; P(B) \neq 0$$

2. **Joint Probability** is the measure of two or more events happening at the same time. It can only be applied to situations where more than one observation can occur simultaneously, i.e., the probability of occurrence of event **B** at the same time when event **A** occurs.

$$P(A \cap B) = P\left(\frac{B}{A}\right) \cdot P(A) ; P(A) \neq 0$$

3. **Marginal Probability** is obtained by summing up probabilities of all the joint events in which a simple event is involved.

### Law of Total Probability

If  $B_1, B_2, \dots, B_n$  are disjoint events and their union completes the entire sample space (i.e., they are mutually exhaustive), then the probability of occurrence of an event **A** will be

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_n)$$

### Baye's Theorem:

This is one of the most famous theorems in probability and lies in the heart of the Naive Bayes algorithm in Machine Learning.



## BAYES THEOREM

$$p(A|B) = \frac{\overset{\text{Likelihood}}{p(B|A)} \overset{\text{Prior probability}}{p(A)}}{p(B)}$$

Posterior probability



## BAE'S THEOREM

$$P(\text{chill}|\text{Netflix}) = \frac{P(\text{Netflix}|\text{chill})P(\text{chill})}{P(\text{Netflix})}$$

i:funny.co

### Source

**In the context of the above image,**

$P(\text{chill})$ := Probability that you are chilling out.

$P(\text{Netflix})$ := Probability that you are watching Netflix.

$P(\text{chill}/\text{Netflix})$ := Probability that you are chilling while watching Netflix.

$P(\text{Netflix}/\text{chill})$ := Probability that you will watch Netflix while chilling out.

**More formally,** Let  $\mathbf{S}$  be a sample space such that  $B_1, B_2, B_3 \dots B_n$  form the partitions of  $\mathbf{S}$  and let  $\mathbf{A}$  be an arbitrary event then,

$$P\left(\frac{B_k}{A}\right) = \frac{P\left(\frac{A}{B_k}\right)P(B_k)}{\sum_{i=1}^n P\left(\frac{A}{B_i}\right)P(B_i)}$$

$P(B_i), i = 1, 2, \dots, n$  are called the prior probabilities of occurrence of events.

$P\left(\frac{A}{B_i}\right), i = 1, 2, \dots, n$  is the conditional probability of  $A$  when  $B_i$  has already occurred.

$P(Bk/A)$  is the posterior probability of  $Bk$  when  $A$  has already occurred.

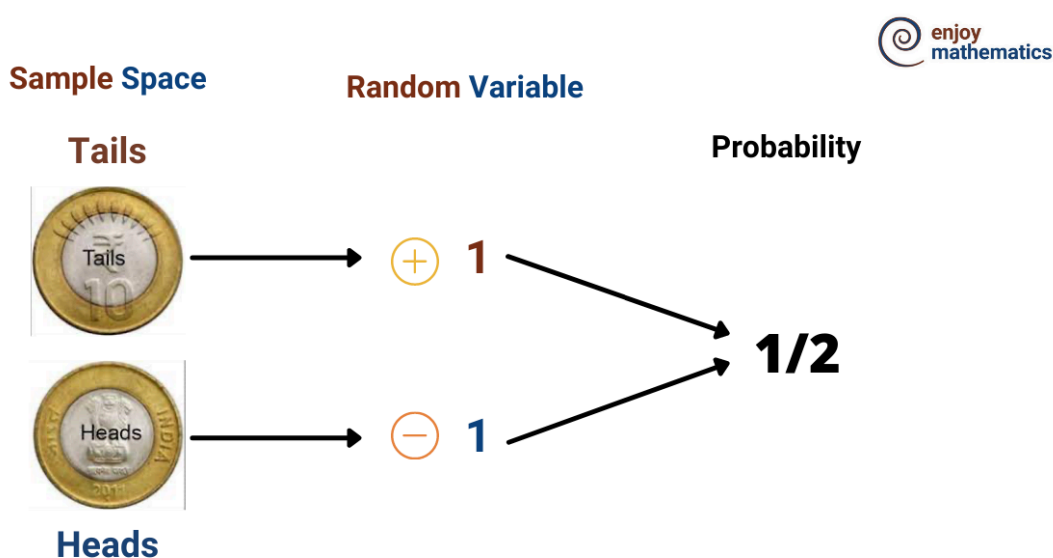
### Random Variables:

Unlike algebraic variables, where the variable in an algebraic equation is unknown and calculated, **random variables** take on different values based on the outcomes of any random experiment. It is just a rule that assigns a number to each possible outcome of an experiment.

Mathematically, a **random variable** is defined as a real function (X or Y or Z) of the elements of a sample space S to a measurable space E, i.e.,

$$X:S \rightarrow E$$

In more layman language,



### Random variables are of two types:

1. **Discrete random variables:** one which has finite numbers of distinct values, basically as count. **Ex.-** Number of times the head occurs if a coin is tossed thrice.
2. **Continuous random variables:** one which is defined in range. **Ex.-** Amount of sugar in 10ml orange juice.



That's it for this article. There are some important concepts related to probability distribution function, mathematical expectations, and famous distribution functions about which we will discuss in part 2 of the probability theory blog.

## Conclusion

In this article, we discussed the basics and most commonly used terminologies in probability and machine learning. We discussed things like deterministic and indeterministic probabilities, exclusive and exhaustive events, the definition of some famous terms, marginal, joint, and conditional probabilities, the famous Baye's theorem, and in the last, we talked about the random variables. We hope you have enjoyed the article.

## Enjoy Learning! Enjoy Mathematics!

Author: [Ravish Raj](#).

Reviewer: [Enjoy Mathematics](#)

Share on social media: [!\[\]\(17413706fd4997a1a4bdf85c6864eee1\_img.jpg\)](#) [!\[\]\(f419710cbe076aa30a9c6c031b5cbe84\_img.jpg\)](#) [!\[\]\(2726020a4107bdc9042b257034f90eb3\_img.jpg\)](#) [!\[\]\(9459655bf14a84f4d775e8d814cca8c9\_img.jpg\)](#)

Share Your Feedback

[Latest](#)

[Math](#)

[Puzzles](#)

[Algorithms](#)

## More blogs to explore

### Find GCD of Two Numbers: Euclidean Algorithm

Given two non-negative integers,  $m$  and  $n$ , we have to find their greatest common divisor or HCF. It is the largest number, a...

[Read More](#)

### Find prime numbers in a range: Sieve of Eratosthenes

The sieve of Eratosthenes is an ancient and efficient algorithm for finding all primes numbers from 2 to  $n$ . This algorithm finds al...

[Read More](#)

### Find the Number of Small Painted Cubes

A cube is painted with some color on all faces. Now, we cut it into 1000 small cubes of equal size. How many small cubes are...

[Read More](#)

### Bridge Crossing at Night

A group of four people, who have one torch, need to cross a bridge at night. A maximum of two people can cross the bridge at one...

[Read More](#)

### For $i = 1$ to $n$ , Sum of Cubes is Equal to Square of the Sum

Prove that the sum of the cubes of the first  $n$  positive integers is equal to the square of the sum of the first  $n$  positive integers.

[Read More](#)

### N-Bulb in a Circle Puzzle

You have given  $n$ -bulbs connected in a circle with a switch for each bulb. The connection is such that if you change the state of anyone...

[Read More](#)

[Coding interview blogs](#)

[System design blogs](#)

[Machine learning blogs](#)

[OOPS design blogs](#)

Follow us on:   

© 2020 Code Algorithms Pvt. Ltd.  
All rights reserved.