Prerequisites to Machine Learning:

- Statistics
- **Probability**
- **Linear Algebra**
- **Calculus**

Probability means possibility. It is a branch of mathematics that deals with the occurrence of a random event. The value is expressed from zero to one. Probability has been introduced in Maths to predict how likely events are to happen. The meaning of probability is basically the extent to which something is likely to happen. This is the basic probability theory, which is also used in the probability distribution, where you will learn the possibility of outcomes for a random experiment. To find the probability of a single event to occur, first, we should know the total number of possible outcomes.

Probability Definition in Math

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)}.

Formula for Probability

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.

Probability of event to happen P(E) = Number of favourable outcomes/Total Number of outcomes

Sometimes students get mistaken for “favourable outcome” with “desirable outcome”. This is the basic formula. But there are some more formulas for different situations or events.

Solved Examples

1) There are 6 pillows in a bed, 3 are red, 2 are yellow and 1 is blue. What is the probability of picking a yellow pillow?

Ans: The probability is equal to the number of yellow pillows in the bed divided by the total number of pillows, i.e. 2/6 = 1/3.

2) There is a container full of coloured bottles, red, blue, green and orange. Some of the bottles are picked out and displaced. Sumit did this 1000 times and got the following results:

- No. of blue bottles picked out: 300
- No. of red bottles: 200
- No. of green bottles: 450
- No. of orange bottles: 50

a) What is the probability that Sumit will pick a green bottle?

Ans: For every 1000 bottles picked out, 450 are green.

Therefore, P(green) = 450/1000 = 0.45

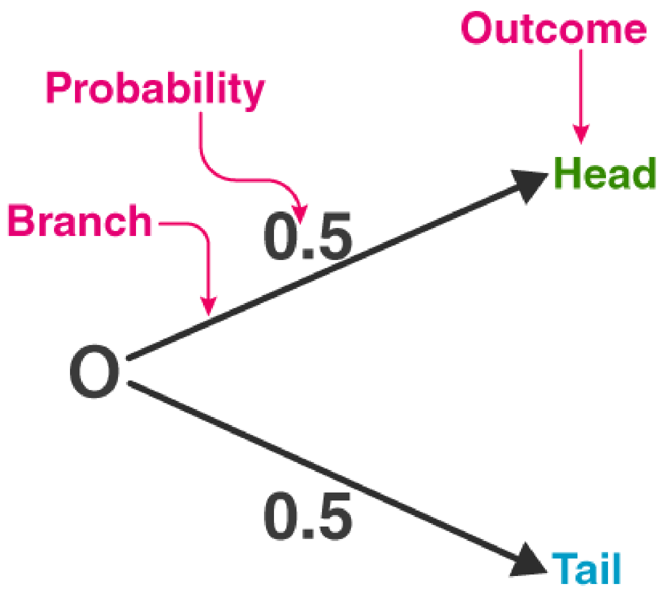
b) If there are 100 bottles in the container, how many of them are likely to be green?

Ans: The experiment implies that 450 out of 1000 bottles are green.

Therefore, out of 100 bottles, 45 are green.

Probability Tree

The **tree diagram** helps to organize and visualize the different possible outcomes. Branches and ends of the tree are two main positions. Probability of each branch is written on the branch, whereas the ends are containing the final outcome. Tree diagrams are used to figure out when to multiply and when to add. You can see below a tree diagram for the coin:



Types of Probability

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.

Conditional Probability is the likelihood of an event or outcome occurring based on the occurrence of a previous event or outcome.

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

This means that the total of all the probabilities in any random test or experiment is equal to 1.

What are Equally Likely Events?

When the events have the same theoretical probability of happening, then they are called equally likely events. The results of a sample space are called equally likely if all of them have the same probability of occurring. For example, if you throw a die, then the probability of getting 1 is 1/6. Similarly, the probability of getting all the numbers from 2,3,4,5 and 6, one at a time is 1/6. Hence, the following are some examples of equally likely events when throwing a die:

- Getting 3 and 5 on throwing a die
- Getting an even number and an odd number on a die
- Getting 1, 2 or 3 on rolling a die

are equally likely events, since the probabilities of each event are equal.

Complementary Events

The possibility that there will be only two outcomes which states that an event will occur or not. Like a person will come or not come to your house, getting a job or not getting a job, etc. are examples of complementary events. Basically, the complement of an event occurring in the exact opposite that the probability of it is not occurring. Some more examples are:

- It will rain or not rain today
- The student will pass the exam or not pass.
- You win the lottery or you don't.

Probability Theory

Probability theory had its root in the 16th century when J. Cardan, an Italian mathematician and physician, addressed the first work on the topic, The Book on Games of Chance. After its inception, the knowledge of probability has brought to the attention of great mathematicians. Thus, **Probability theory** is the branch of mathematics that deals with the possibility of the happening of events. Although there are many distinct probability interpretations, probability theory interprets the concept precisely by expressing it through a set of axioms or hypotheses. These hypotheses help form the probability in terms of a possibility space, which allows a measure holding values between 0 and 1. This is known as the probability measure, to a set of possible outcomes of the sample space.

Probability Density Function

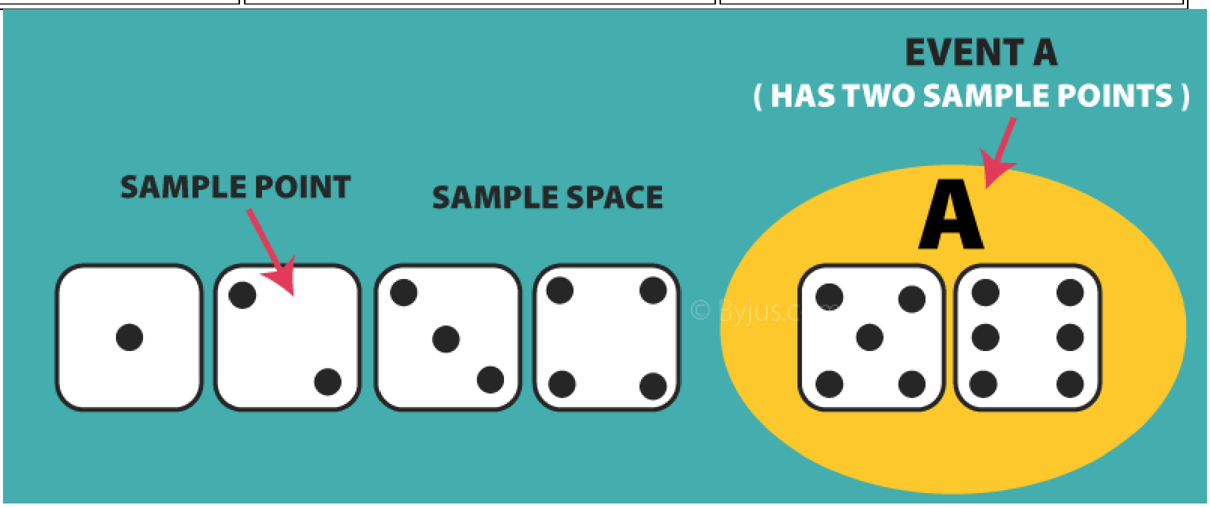
The Probability Density Function (PDF) is the probability function which is represented for the density of a continuous random variable lying between a certain range of values. Probability Density Function explains the normal distribution and how mean and deviation exists. The standard normal distribution is used to create a database or statistics, which are often used in science to represent the real-valued variables, whose distribution is not known.

Probability Terms and Definition

Some of the important probability terms are discussed here:

Term	**Definition**	**Example**
Sample Space	The set of all the possible outcomes to occur in any trial	1. Tossing a coin, Sample Space (S) = {H,T} 2. Rolling a die, Sample Space (S) = {1,2,3,4,5,6}

Term	Definition	Example
Sample Point	It is one of the possible results	In a deck of Cards: • 4 of hearts is a sample point. • The queen of clubs is a sample point.
Experiment or Trial	A series of actions where the outcomes are always uncertain.	The tossing of a coin, Selecting a card from a deck of cards, throwing a dice.
Event	It is a single outcome of an experiment.	Getting a Heads while tossing a coin is an event.
Outcome	Possible result of a trial/experiment	T (tail) is a possible outcome when a coin is tossed.
Complimentary event	The non-happening events. The complement of an event A is the event, not A (or A')	In a standard 52-card deck, A = Draw a heart, then A' = Don't draw a heart
Impossible Event	The event cannot happen	In tossing a coin, impossible to get both head and tail at the same time



Applications of Probability

Probability has a wide variety of applications in real life. Some of the common applications which we see in our everyday life while checking the results of the following events:

- Choosing a card from the deck of cards
- Flipping a coin
- Throwing a dice in the air
- Pulling a red ball out of a bucket of red and white balls
- Winning a lucky draw

Other Major Applications of Probability

- It is used for risk assessment and modelling in various industries
- Weather forecasting or prediction of weather changes
- Probability of a team winning in a sport based on players and strength of team
- In the share market, chances of getting the hike of share prices

Problems and Solutions on Probability

Question 1: Find the probability of 'getting 3 on rolling a die'.

Solution:

Sample Space = S = {1, 2, 3, 4, 5, 6}

Total number of outcomes = n(S) = 6

Let A be the event of getting 3.

Number of favourable outcomes = n(A) = 1

i.e. A = {3}

Probability, P(A) = n(A)/n(S) = 1/6

Hence, P(getting 3 on rolling a die) = 1/6

Question 2: Draw a random card from a pack of cards. What is the probability that the card drawn is a face card?

Solution:

A standard deck has 52 cards.

Total number of outcomes = n(S) = 52

Let E be the event of drawing a face card.

Number of favourable events = n(E) = 4 x 3 = 12 (considered Jack, Queen and King only)

Probability, P = Number of Favourable Outcomes/Total Number of Outcomes

P(E) = n(E)/n(S)

= 12/52

= 3/13

P(the card drawn is a face card) = 3/13

Question 3: A vessel contains 4 blue balls, 5 red balls and 11 white balls. If three balls are drawn from the vessel at random, what is the probability that the first ball is red, the second ball is blue, and the third ball is white?

Solution:

Given,

The probability to get the first ball is red or the first event is 5/20.

Since we have drawn a ball for the first event to occur, then the number of possibilities left for the second event to occur is 20 – 1 = 19.

Hence, the probability of getting the second ball as blue or the second event is 4/19.

Again with the first and second event occurring, the number of possibilities left for the third event to occur is 19 – 1 = 18.

And the probability of the third ball is white or the third event is 11/18.

Therefore, the probability is 5/20 x 4/19 x 11/18 = 44/1368 = 0.032.

Or we can express it as: P = 3.2%.

Question 4: Two dice are rolled, find the probability that the sum is:

1. **equal to 1**
2. **equal to 4**
3. **less than 13**

Solution:

To find the probability that the sum is equal to 1 we have to first determine the sample space S of two dice as shown below.

S = { (1,1),(1,2),(1,3),(1,4),(1,5),(1,6)

(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)

(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)

(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)

(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)

(6,1),(6,2),(6,3),(6,4),(6,5),(6,6) }

So, n(S) = 36

1) Let E be the event "sum equal to 1". Since, there are no outcomes which where a sum is equal to 1, hence,

P(E) = n(E) / n(S) = 0 / 36 = 0

2) Let A be the event of getting the sum of numbers on dice equal to 4.

Three possible outcomes give a sum equal to 4 they are:

A = {(1,3),(2,2),(3,1)}

n(A) = 3

Hence, P(A) = n(A) / n(S) = 3 / 36 = 1 / 12

3) Let B be the event of getting the sum of numbers on dice is less than 13.

From the sample space, we can see all possible outcomes for the event B, which gives a sum less than B. Like:

(1,1) or (1,6) or (2,6) or (6,6).

So you can see the limit of an event to occur is when both dies have number 6, i.e. (6,6).

Thus, n(B) = 36

Hence,

P(B) = n(B) / n(S) = 36 / 36 = 1

Probability Problems

1. Two dice are thrown together. Find the probability that the product of the numbers on the top of the dice is:
(i) 6 (ii) 12 (iii) 7
2. A bag contains 10 red, 5 blue and 7 green balls. A ball is drawn at random. Find the probability of this ball being a
(i) red ball (ii) green ball (iii) not a blue ball
3. All the jacks, queens and kings are removed from a deck of 52 playing cards. The remaining cards are well shuffled and then one card is drawn at random. Giving ace a value 1 similar value for other cards, find the probability that the card has a value
(i) 7 (ii) greater than 7 (iii) less than 7

4. A die has its six faces marked 0, 1, 1, 1, 6, 6. Two such dice are thrown together and the total score is recorded.
   (i) How many different scores are possible?
   (ii) What is the probability of getting a total of 7?

Probability quantifies the uncertainty of the outcomes of a random variable.

It is relatively easy to understand and compute the probability for a single variable. Nevertheless, in machine learning, we often have many random variables that interact in often complex and unknown ways.

There are specific techniques that can be used to quantify the probability for multiple random variables, such as the joint, marginal, and conditional probability. These techniques provide the basis for a probabilistic understanding of fitting a predictive model to data.

- Joint probability is the probability of two events occurring simultaneously.
- Marginal probability is the probability of an event irrespective of the outcome of another variable.
- Conditional probability is the probability of one event occurring in the presence of a second event.

This tutorial is divided into three parts; they are:

1. Probability of One Random Variable
2. Probability of Multiple Random Variables
3. Probability of Independence and Exclusivity

Probability of One Random Variable

Probability quantifies the likelihood of an event.

Specifically, it quantifies how likely a specific outcome is for a random variable, such as the flip of a coin, the roll of a dice, or drawing a playing card from a deck.

Probability gives a measure of how likely it is for something to happen.

For a random variable *x*, *P(x)* is a function that assigns a probability to all values of *x*.

- Probability Density of x = P(x)

The probability of a specific event *A* for a random variable x is denoted as *P(x=A)*, or simply as *P(A).*

- Probability of Event A = P(A)

Probability is calculated as the number of desired outcomes divided by the total possible outcomes, in the case where all outcomes are equally likely.

- Probability = (number of desired outcomes) / (total number of possible outcomes)

This is intuitive if we think about a discrete random variable such as the roll of a die. For example, the probability of a die rolling a 5 is calculated as one outcome of rolling a 5 (1) divided by the total number of discrete outcomes (6) or 1/6 or about 0.1666 or about 16.666%.

The sum of the probabilities of all outcomes must equal one. If not, we do not have valid probabilities.

- Sum of the Probabilities for All Outcomes = 1.0.

The probability of an impossible outcome is zero. For example, it is impossible to roll a 7 with a standard six-sided die.

- Probability of Impossible Outcome = 0.0

The probability of a certain outcome is one. For example, it is certain that a value between 1 and 6 will occur when rolling a six-sided die.

- Probability of Certain Outcome = 1.0

The probability of an event not occurring, called the complement.

This can be calculated by one minus the probability of the event, or *1 – P(A)*. For example, the probability of not rolling a 5 would be 1 – P(5) or 1 – 0.166 or about 0.833 or about 83.333%.

- Probability of Not Event A = 1 – P(A)

Now that we are familiar with the probability of one random variable, let’s consider probability for multiple random variables.

Probability of Multiple Random Variables

In machine learning, we are likely to work with many random variables.

For example, given a table of data, such as in excel, each row represents a separate observation or event, and each column represents a separate random variable.

Variables may be either discrete, meaning that they take on a finite set of values, or continuous, meaning they take on a real or numerical value.

As such, we are interested in the probability across two or more random variables.

This is complicated as there are many ways that random variables can interact, which, in turn, impacts their probabilities.

This can be simplified by reducing the discussion to just two random variables (*X, Y*), although the principles generalize to multiple variables.
And further, to discuss the probability of just two events, one for each variable (*X=A, Y=B*), although we could just as easily be discussing groups of events for each variable.
Therefore, we will introduce the probability of multiple random variables as the probability of event *A* and event *B*, which in shorthand is *X=A* and *Y=B*.
We assume that the two variables are related or dependent in some way.

As such, there are three main types of probability we might want to consider; they are:

- **Joint Probability**: Probability of events *A* and *B*.
- **Marginal Probability**: Probability of event X=*A* given variable *Y*.
- **Conditional Probability**: Probability of event *A* given event *B*.

These types of probability form the basis of much of predictive modeling with problems such as classification and regression. For example:

- The probability of a row of data is the joint probability across each input variable.
- The probability of a specific value of one input variable is the marginal probability across the values of the other input variables.
- The predictive model itself is an estimate of the conditional probability of an output given an input example.

Joint, marginal, and conditional probability are foundational in machine learning.

Let's take a closer look at each in turn.

Joint Probability of Two Variables

We may be interested in the probability of two simultaneous events, e.g. the outcomes of two different random variables.

The probability of two (or more) events is called the joint probability. The joint probability of two or more random variables is referred to as the joint probability distribution.
For example, the joint probability of event *A* and event *B* is written formally as:

- P(A and B)

The "*and*" or conjunction is denoted using the upside down capital "*U*" operator "^" or sometimes a comma ",".

- P(A ^ B)
- P(A, B)

The joint probability for events *A* and *B* is calculated as the probability of event *A* given event *B* multiplied by the probability of event *B*.
This can be stated formally as follows:

- P(A and B) = P(A given B) * P(B)

The calculation of the joint probability is sometimes called the fundamental rule of probability or the "*product rule*" of probability or the "chain rule" of probability.
Here, *P(A given B)* is the probability of event A given that event B has occurred, called the conditional probability, described below.
The joint probability is symmetrical, meaning that *P(A and B)* is the same as *P(B and A)*. The calculation using the conditional probability is also symmetrical, for example:

- P(A and B) = P(A given B) * P(B) = P(B given A) * P(A)

Marginal Probability

We may be interested in the probability of an event for one random variable, irrespective of the outcome of another random variable.

For example, the probability of *X=A* for all outcomes of *Y*.
The probability of one event in the presence of all (or a subset of) outcomes of the other random variable is called the marginal probability or the marginal distribution. The marginal probability of one random variable in the presence of additional random variables is referred to as the marginal probability distribution.
It is called the marginal probability because if all outcomes and probabilities for the two variables were laid out together in a table (*X* as columns, *Y* as rows), then the marginal probability of one variable (*X*) would be the sum of probabilities for the other variable (Y rows) on the margin of the table.
There is no special notation for the marginal probability; it is just the sum or union over all the probabilities of all events for the second variable for a given fixed event for the first variable.

- P(X=A) = sum P(X=A, Y=yi) for all y

This is another important foundational rule in probability, referred to as the "*sum rule*."
The marginal probability is different from the conditional probability (described next) because it considers the union of all events for the second variable rather than the probability of a single event.

Conditional Probability

We may be interested in the probability of an event given the occurrence of another event.

The probability of one event given the occurrence of another event is called the conditional probability. The conditional probability of one to one or more random variables is referred to as the conditional probability distribution.

For example, the conditional probability of event *A* given event *B* is written formally as:

- P(A given B)

The "*given*" is denoted using the pipe "|" operator; for example:

- P(A | B)

The conditional probability for events *A* given event *B* is calculated as follows:

- P(A given B) = P(A and B) / P(B)

This calculation assumes that the probability of event *B* is not zero, e.g. is not impossible.

The notion of event *A* given event *B* does not mean that event *B* has occurred (e.g. is certain); instead, it is the probability of event *A* occurring after or in the presence of event *B* for a given trial.

Probability of Independence and Exclusivity

When considering multiple random variables, it is possible that they do not interact.

We may know or assume that two variables are not dependent upon each other instead are independent.

Alternately, the variables may interact but their events may not occur simultaneously, referred to as exclusivity.

We will take a closer look at the probability of multiple random variables under these circumstances in this section.

Independence

If one variable is not dependent on a second variable, this is called independence or statistical independence.

This has an impact on calculating the probabilities of the two variables.

For example, we may be interested in the joint probability of independent events *A* and *B*, which is the same as the probability of *A* and the probability of *B.*

Probabilities are combined using multiplication, therefore the joint probability of independent events is calculated as the probability of event *A* multiplied by the probability of event *B*.

This can be stated formally as follows:

- **Joint Probability**: P(A and B) = P(A) * P(B)

As we might intuit, the marginal probability for an event for an independent random variable is simply the probability of the event.

It is the idea of probability of a single random variable that are familiar with:

- **Marginal Probability**: P(A)

We refer to the marginal probability of an independent probability as simply the probability.

Similarly, the conditional probability of *A* given *B* when the variables are independent is simply the probability of *A* as the probability of *B* has no effect. For example:

- **Conditional Probability**: P(A given B) = P(A)

We may be familiar with the notion of statistical independence from sampling. This assumes that one sample is unaffected by prior samples and does not affect future samples.

Many machine learning algorithms assume that samples from a domain are independent to each other and come from the same probability distribution, referred to as independent and identically distributed, or i.i.d. for short.

Exclusivity

If the occurrence of one event excludes the occurrence of other events, then the events are said to be mutually exclusive.

The probability of the events are said to be disjoint, meaning that they cannot interact, are strictly independent.

If the probability of event *A* is mutually exclusive with event *B*, then the joint probability of event *A* and event *B* is zero.

- P(A and B) = 0.0

Instead, the probability of an outcome can be described as event *A* or event *B*, stated formally as follows:

- P(A or B) = P(A) + P(B)

The "or" is also called a union and is denoted as a capital "*U*" letter; for example:

- P(A or B) = P(A U B)

If the events are not mutually exclusive, we may be interested in the outcome of either event.

The probability of non-mutually exclusive events is calculated as the probability of event *A* and the probability of event *B* minus the probability of both events occurring simultaneously.

This can be stated formally as follows:

- P(A or B) = P(A) + P(B) – P(A and B)

Bayes' theorem describes the probability of occurrence of an event related to any condition. It is also considered for the case of conditional probability. 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.

Conditional probability: Bayes' Theorem

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

Bayes Theorem Statement

Let $E_1, E_2, \ldots, E_n$ be a set of events associated with a sample space S, where all the events $E_1, E_2, \ldots, E_n$ have nonzero probability of occurrence and they form a partition of S. Let A be any event associated with S, then according to Bayes theorem,

||P(Ei|A) = P(Ei)P(A|Ei)∑k=1nP(Ek)P(A|Ek)

for any k = 1, 2, 3, …., n

Bayes Theorem Proof

According to the conditional probability formula,

|P(Ei|A) = P(Ei∩A)P(A)…(1)

Using the multiplication rule of probability,

|P(Ei∩A) = P(Ei)P(A|Ei)…(2)

Using total probability theorem,

P(A) = ∑k=1n P(Ek)P(A|Ek)…(3)

Putting the values from equations (2) and (3) in equation 1, we get

||P(Ei|A) = P(Ei)P(A|Ei)∑k=1n P(Ek)P(A|Ek)

Note:

The following terminologies are also used when the Bayes theorem is applied:

Hypotheses: The events E_1, E_2,…E_n is called the hypotheses

Priori Probability: The probability $P(E_i)$ is considered as the priori probability of hypothesis E_i

Posteriori Probability: The probability $P(E_i|A)$ is considered as the posteriori probability of hypothesis E_i
Bayes' theorem is also called the formula for the probability of "causes". Since the E_i's are a partition of the sample space S, one and only one of the events E_i occurs (i.e. one of the events E_i must occur and the only one can occur). Hence, the above formula gives us the probability of a particular E_i (i.e. a "Cause"), given that the event A has occurred.

Bayes Theorem Formula

If A and B are two events, then the **formula for the Bayes theorem** is given by:

P(A|B)=P(B|A)P(A)P(B)whereP(B)≠0

Where P(A|B) is the probability of condition when event A is occurring while event B has already occurred.

Bayes Theorem Derivation

Bayes Theorem can be derived for events and random variables separately using the definition of conditional probability and density.

From the definition of conditional probability, Bayes theorem can be derived for events as given below:

$P(A|B) = P(A \cap B)/ P(B)$, where $P(B) \neq 0$

$P(B|A) = P(B \cap A)/ P(A)$, where $P(A) \neq 0$

Here, the joint probability $P(A \cap B)$ of both events A and B being true such that,

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

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

$P(A|B) = [P(B|A)\, P(A)]/ P(B)$, where $P(B) \neq 0$

Similarly, from the definition of conditional density, Bayes theorem can be derived for two continuous random variables namely X and Y as given below:

fX|Y=y(x)=fX,Y(x,y)fY(y)fY|X=x(y)=fX,Y(x,y)fX(x)

Therefore,

fX|Y=y(x)=fY|X=x(y)fX(x)fY(y)

Examples and Solutions

Some illustrations will improve the understanding of the concept.

Example 1:

A bag I contains 4 white and 6 black balls while another Bag II contains 4 white and 3 black balls. One ball is drawn at random from one of the bags, and it is found to be black. Find the probability that it was drawn from Bag I.

Solution:

Let E_1 be the event of choosing bag I, E_2 the event of choosing bag II, and A be the event of drawing a black ball.

Then,

P(E1) = P(E2) = 12

Also, $P(A|E_1)$ = P(drawing a black ball from Bag I) = 6/10 = 3/5

$P(A|E_2)$ = P(drawing a black ball from Bag II) = 3/7

By using Bayes' theorem, the probability of drawing a black ball from bag I out of two bags,

|P(E1|A) = P(E1)P(A|E1)P(E1)P(A|E1)+P(E2)P(A|E2)

=12 × 3512 × 35 + 12 × 37

=712

Example 2:

A man is known to speak the truth 2 out of 3 times. He throws a die and reports that the number obtained is a four. Find the probability that the number obtained is actually a four.

Solution:

Let A be the event that the man reports that number four is obtained.

Let E_1 be the event that four is obtained and E_2 be its complementary event.

Then, $P(E_1)$ = Probability that four occurs = 1/6.

$P(E_2)$ = Probability that four does not occur = 1- $P(E_1)$ = 1 – (1/6) = 5/6.

Also, $P(A|E_1)$= Probability that man reports four and it is actually a four = 2/3

$P(A|E_2)$ = Probability that man reports four and it is not a four = 1/3.

By using Bayes' theorem, probability that number obtained is actually a four, $P(E_1|A)$

|=P(E1)P(A|E1)P(E1)P(A|E1) + P(E2)P(A|E2) = 16 × 2316 × 23 + 56 × 13

=27

Bayes Theorem Applications

One of the many applications of Bayes' theorem is Bayesian inference, a particular approach to statistical inference. Bayesian inference has found application in various activities, including medicine, science, philosophy, engineering, sports, law, etc. For example, we can use Bayes' theorem to define the accuracy of medical test results by considering how likely any given person is to have a disease and the test's overall accuracy. Bayes' theorem relies on consolidating prior probability distributions to generate posterior probabilities. In Bayesian statistical inference, prior probability is the probability of an event before new data is collected.

Practice Problems

Solve the following problems using Bayes Theorem.

1. A bag contains 5 red and 5 black balls. A ball is drawn at random, its colour is noted, and again the ball is returned to the bag. Also, 2 additional balls of the colour drawn are put in the bag. After that, the ball is drawn at random from the bag. What is the probability that the second ball drawn from the bag is red?
2. Of the students in the college, 60% of the students reside in the hostel and 40% of the students are day scholars. Previous year results report that 30% of all students who stay in the hostel scored A Grade and 20% of day scholars scored A grade. At the end of the year, one student is chosen at random and found that he/she has an A grade. What is the probability that the student is a hosteler?
3. From the pack of 52 cards, one card is lost. From the remaining cards of a pack, two cards are drawn and both are found to be diamond cards. What is the probability that the lost card is a diamond?

Continuous Random Variable

Continuous random variable is a random variable that can take on a continuum of values. In other words, a random variable is said to be continuous if it assumes a value that falls between a particular interval.

Continuous random variables are used to denote measurements such as height, weight, time, etc. The area under a density curve is used to represent a continuous random variable. In this article, we will learn about the definition of a continuous random variable, its mean, variance, types, and associated examples.

What is a Continuous Random Variable?

A continuous random variable and a discrete random variable are the two types of random variables. A random variable is a variable whose value depends on all the possible outcomes of an experiment. A continuous random variable is defined over a range of values while a discrete random variable is defined at an exact value.

Continuous Random Variable Definition

A continuous random variable can be defined as a random variable that can take on an infinite number of possible values. Due to this, the probability that a continuous random variable will take on an exact value is 0. The cumulative distribution function and the probability density function are used to describe the characteristics of a continuous random variable.

Continuous Random Variable Example

Suppose the probability density function of a continuous random variable, X, is given by $4x^3$, where $x \in [0, 1]$. The probability that X takes on a value between 1/2 and 1 needs to be determined. This can be done by integrating $4x^3$ between 1/2 and 1. Thus, the required probability is 15/16.

Continuous Random Variable Formulas

The probability density function (pdf) and the cumulative distribution function (CDF) are used to describe the probabilities associated with a continuous random variable. The continuous random variable formulas for these functions are given below.

PDF of Continuous Random Variable

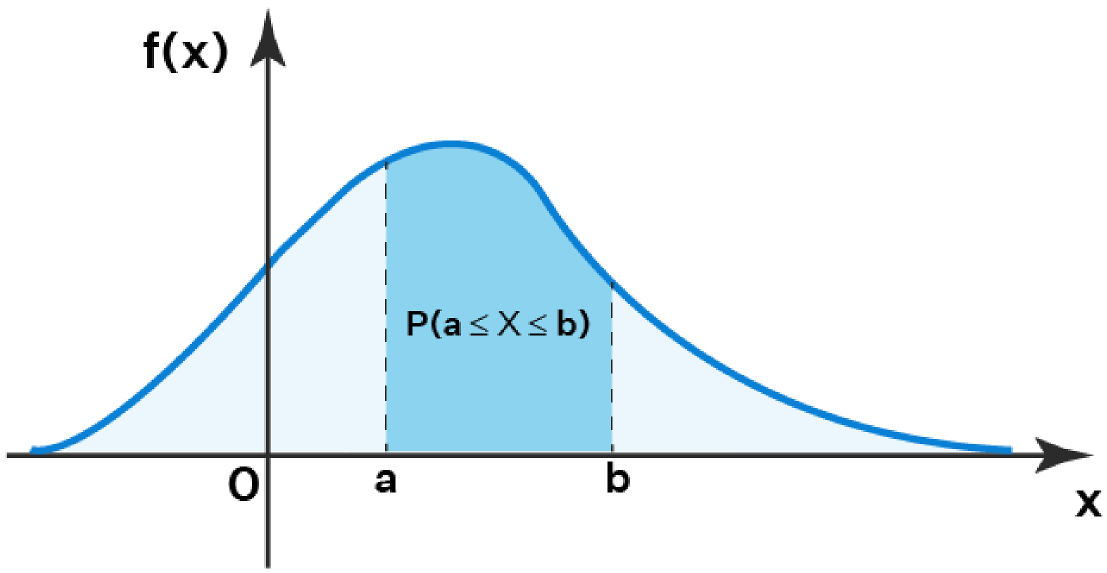
The probability density function of a continuous random variable can be defined as a function that gives the probability that the value of the random variable will fall between a range of values. Let X be the continuous random variable, then the formula for the pdf, f(x), is given as follows:

f(x) = dF(x)dxdF(x)dx = F'(x)

where, F(x) is the cumulative distribution function.

For the pdf of a continuous random variable to be valid, it must satisfy the following conditions:

- $\int_{-\infty}^{\infty} f(x)dx=1 \int_{-\infty}^{\infty} f(x)dx=1$. This means that the total area under the graph of the pdf must be equal to 1.
- $f(x) \geq 0$. This implies that the probability density function of a continuous random variable cannot be negative.



CDF of Continuous Random Variable

The cumulative distribution function of a continuous random variable can be determined by integrating the probability density function. It can be defined as the probability that the random variable, X, will take on a value that is lesser than or equal to a particular value, x. The formula for the cdf of a continuous random variable, evaluated between two points a and b, is given below:

$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx \int_a^b f(x)dx$

Mean and Variance of Continuous Random Variable

The mean and variance of a continuous random variable can be determined with the help of the probability density function, f(x).

Mean of Continuous Random Variable

The mean of a continuous random variable can be defined as the weighted average value of the random variable, X. It is also known as the expectation of the continuous random variable. The formula is given as follows:

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

Variance of Continuous Random Variable

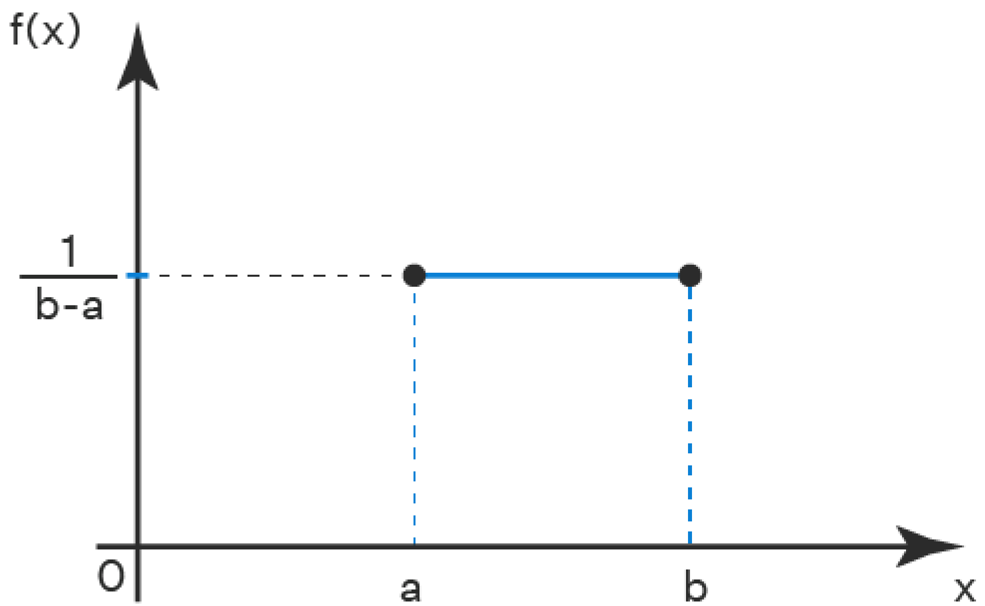
The variance of a continuous random variable can be defined as the expectation of the squared differences from the mean. It helps to determine the dispersion in the distribution of the continuous random variable with respect to the mean. The formula is given as follows:
Var(X) = σ2=∫∞−∞(x−μ)2f(x)dxσ2=∫−∞∞(x−μ)2f(x)dx

Continuous Random Variable Types

A continuous random variable is usually used to model situations that involve measurements. For example, the possible values of the temperature on any given day. As the temperature could be any real number in a given interval thus, a continuous random variable is required to describe it. Some important continuous random variables associated with certain probability distributions are given below.

Uniform Random Variable

A continuous random variable that is used to describe a uniform distribution is known as a uniform random variable. Such a distribution describes events that are equally likely to occur. The pdf of a uniform random variable is as follows:
f(x)={1b−aa≤x≤b0otherwisef(x)={1b−aa≤x≤b0otherwise

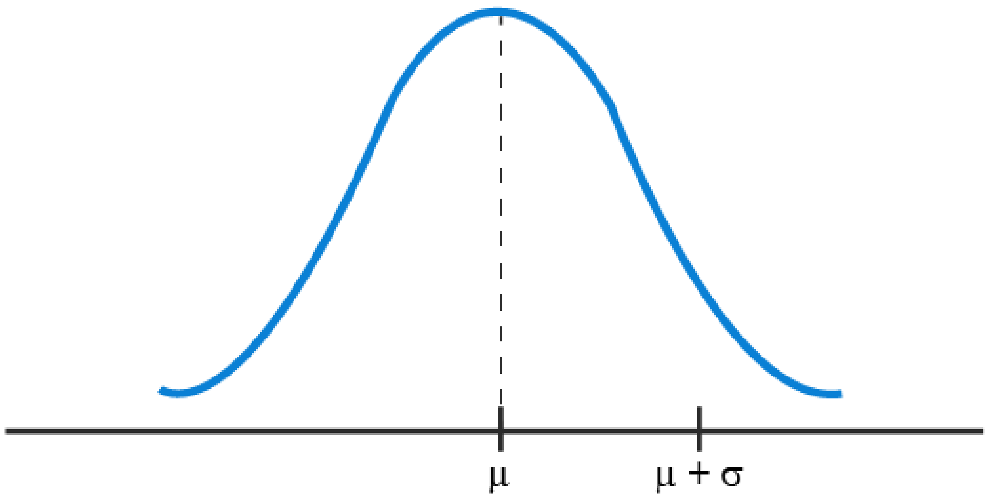


Normal Random Variable

A continuous random variable that is used to model a normal distribution is known as a normal random variable. If the parameters of a normal distribution are given as X∼N(μ,σ2)X∼N(μ,σ2) then the formula for the pdf is given as follows:
f(x) = 1σ√2Πe−12(x−μσ)21σ2Πe−12(x−μσ)2

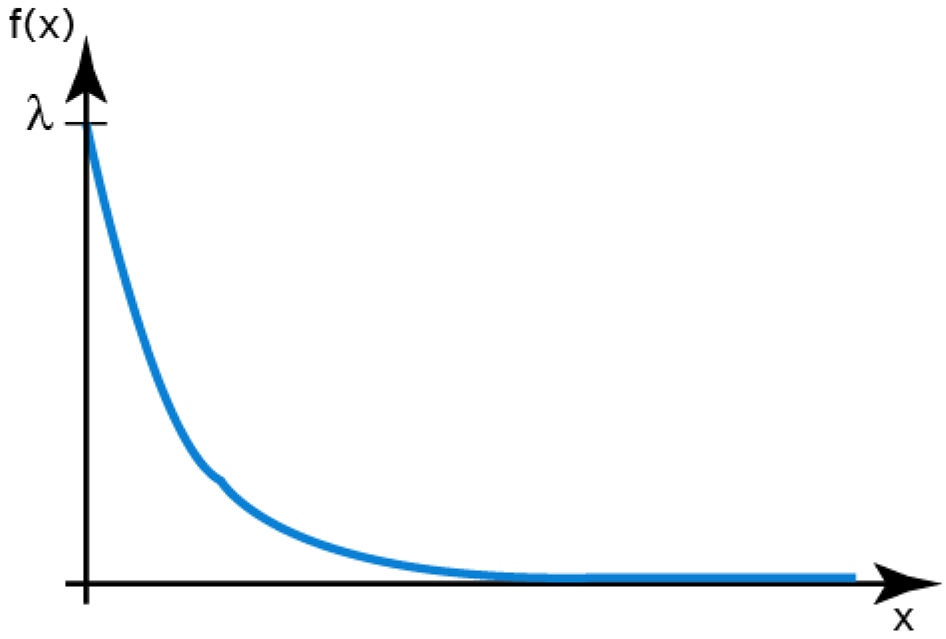
where,

$\mu\mu$ = mean
$\sigma\sigma$ = standard deviation
$\sigma\sigma^2$ = variance.
A normal distribution where $\mu\mu$ = 0 and $\sigma\sigma^2$ = 1 is known as a standard normal distribution. Thus, a standard normal random variable is a continuous random variable that is used to model a standard normal distribution. The pdf formula is as follows:
f(x) = 1√2Πe−x2212Πe−x22



Exponential Random Variable

Exponential distributions are continuous probability distributions that model processes where a certain number of events occur continuously at a constant average rate, $\lambda\geq0\lambda\geq0$. Thus, a continuous random variable used to describe such a distribution is called an exponential random variable. The pdf is given as follows:
f(x) = $\lambda e^{-\lambda x}$



Continuous Random Variable vs Discrete Random Variable

Both discrete and continuous random variables are used to model a random phenomenon. The differences between a continuous random variable and discrete random variable are given in the table below:

Continuous Random Variable	Discrete Random Variable
The value of a continuous random variable falls between a range of values.	The value of a discrete random variable is an exact value.
The probability density function is associated with a continuous random variable.	The probability mass function is used to describe a discrete random variable
A continuous random variable can take on an infinite number of values.	Such a variable can take on a finite number of distinct values.

Continuous Random Variable	Discrete Random Variable
Mean of a continuous random variable is E[X] = ∫∞−∞xf(x)dx∫−∞∞xf(x)dx	The mean of a discrete random variable is E[X] = ∑ x P(X = x), where P(X = x) is the probability mass function.
The variance of a continuous random variable is Var(X) = ∫∞−∞(x−μ)2f(x)dx∫−∞∞(x−μ)2f(x)dx	The variance of a discrete random variable is Var[X] = $\sum(x - \mu)^2 P(X = x)$
The examples of a continuous random variable are uniform random variable, exponential random variable, normal random variable, and standard normal random variable.	The examples of a discrete random variable are binomial random variable, geometric random variable, Bernoulli random variable, and Poisson random variable.

Important Notes on Continuous Random Variable

- A continuous random variable is a variable that is used to model continuous data and its value falls between an interval of values.
- The probability density function of a continuous random variable is given as f(x) = dF(x)dxdF(x)dx = F'(x).
- The cumulative distribution function is given by P(a < X ≤ b) = F(b) - F(a) = ∫baf(x)dx∫abf(x)dx.
- The mean of a continuous random variable is E[X] = μ=∫∞−∞xf(x)dxμ=∫−∞∞xf(x)dx and variance is Var(X) = σ2=∫∞−∞(x−μ)2f(x)dx
- Uniform random variable, exponential random variable, normal random variable, and standard normal random variable are examples of continuous random variables.

Linear Algebra

Linear algebra is a branch of mathematics that deals with linear equations and their representations in the vector space using matrices. In other words, linear algebra is the study of linear functions and vectors. It is one of the most central topics of mathematics. Most modern geometrical concepts are based on linear algebra.

Linear algebra facilitates the modelling of many natural phenomena and hence, is an integral part of engineering and physics. Linear equations, matrices, and vector spaces are the most important components of this subject.

What is Linear Algebra?

Linear algebra can be defined as a branch of mathematics that deals with the study of linear functions in vector spaces. When information related to linear functions is presented in an organized form then it results in a matrix. Thus, linear algebra is concerned with vector spaces, vectors, linear functions, the system of linear equations, and matrices. These concepts are a prerequisite for sister topics such as geometry and functional analysis.

Linear Algebra Definition

The branch of mathematics that deals with vectors, matrices, finite or infinite dimensions as well as a linear mapping between such spaces is defined as linear algebra. It is used in both pure and applied mathematics along with different technical forms such as physics, engineering, natural sciences, etc.

Branches of Linear Algebra

Linear algebra can be categorized into three branches depending upon the level of difficulty and the kind of topics that are encompassed within each. These are elementary, advanced, and applied linear algebra. Each branch covers different aspects of matrices, vectors, and linear functions.

Elementary Linear Algebra

Elementary linear algebra introduces students to the basics of linear algebra. This includes simple matrix operations, various computations that can be done on a system of linear equations, and certain aspects of vectors. Some important terms associated with elementary linear algebra are given below:

Scalars - A scalar is a quantity that only has magnitude and not direction. It is an element that is used to define a vector space. In linear algebra, scalars are usually real numbers.

Vectors - A vector is an element in a vector space. It is a quantity that can describe both the direction and magnitude of an element.

Vector Space - The vector space consists of vectors that may be added together and multiplied by scalars.

Matrix - A matrix is a rectangular array wherein the information is organized in the form of rows and columns. Most linear algebra properties can be expressed in terms of a matrix.

Matrix Operations - These are simple arithmetic operations such as addition, subtraction, and multiplication that can be conducted on matrices.

Advanced Linear Algebra

Once the basics of linear algebra have been introduced to students the focus shifts on more advanced concepts related to linear equations, vectors, and matrices. Certain important terms that are used in advanced linear algebra are as follows:

Linear Transformations - The transformation of a function from one vector space to another by preserving the linear structure of each vector space.

Inverse of a Matrix - When an inverse of a matrix is multiplied with the given original matrix then the resultant will be the identity matrix. Thus, $A^{-1}A = I$.

Eigenvector - An eigenvector is a non-zero vector that changes by a scalar factor (eigenvalue) when a linear transformation is applied to it.

Linear Map - It is a type of mapping that preserves vector addition and vector multiplication.

Applied Linear Algebra

Applied linear algebra is usually introduced to students at a graduate level in fields of applied mathematics, engineering, and physics. This branch of algebra is driven towards integrating the concepts of elementary and advanced linear algebra with their practical implications. Topics such as the norm of a vector, QR factorization, Schur's complement of a matrix, etc., fall under this branch of linear algebra.

Linear Algebra Topics

The topics that come under linear algebra can be classified into three broad categories. These are linear equations, matrices, and vectors. All these three categories are interlinked and need to be understood well in order to master linear algebra. The topics that fall under each category are given below.

Linear Equations

A linear equation is an equation that has the standard form a1x1+a2x2+...+anxna1x1+a2x2+...+anxn. It is the fundamental component of linear algebra.

Vectors

In linear algebra, there can be several operations that can be performed on vectors such as multiplication, addition, etc. Vectors can be used to describe quantities such as the velocity of moving objects. Some crucial topics encompassed under vectors are as follows:

Matrices

A matrix is used to organize data in the form of a rectangular array. It can be represented as Am×nAm×n. Here, m represents the number of rows and n denotes the number of columns in the matrix. In linear algebra, a matrix can be used to express linear equations in a more compact manner.

Linear Algebra Formula

Formulas form an important part of linear algebra as they help to simplify computations. The key to solving any problem in linear algebra is to understand the formulas and associated concepts rather than memorize them. The important linear algebra formulas can be broken down into 3 categories, namely, linear equations, vectors, and matrices.

Linear Equations: The important linear equation formulas are listed as follows:

- General form: ax + by = c
- Slope Intercept Form: y = mx + b
- a + b = b + a
- a + 0 = 0 + a = a

Vectors: If there are two vectors →uu→ = (u1u1, u2u2, u3u3) and →vv→ = (v1v1, v2v2, v3v3) then the important vector formulas associated with linear algebra are given below.

- →u+→v=(u1+v1,u2+v2,u3+v3)u→+v→=(u1+v1,u2+v2,u3+v3)
- →u−→v=(u1−v1,u2−v2,u3−v3)u→−v→=(u1−v1,u2−v2,u3−v3)
- ‖u‖=√u21+u22+u23||u||=u12+u22+u32
- →u.→v=u1v1+u2v2+u3v3u→.v→=u1v1+u2v2+u3v3
- →u×→v=(u2v3−u3v2,u3v1−u1v3,u1v2−u2v1)u→×v→=(u2v3−u3v2,u3v1−u1v3,u1v2−u2v1)

Matrix: If there are two square matrices given by A and B where the elements are aijaij and bijbij respectively, then the following important formulas are used in linear algebra:

$$\begin{array}{cc} & \overbrace{\begin{array}{cccc} 1 & 2 & \cdots & n \end{array}}^{\text{Columns}} \\ \text{Rows}\left\{\begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array}\right. & \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \end{array} = A_{m \times n}$$

- $A^{-1}A = I$
- C = A + B, where cijcij = aijaij + bij
- C = A - B, where cijcij = aijaij - bij
- kA = kaijaij
- C = AB = ∑nk=1aikbkj∑k=1naikbkj

Linear Algebra and its Applications

Linear algebra is used in almost every field. Simple algorithms also make use of linear algebra topics such as matrices. Some of the applications of linear algebra are given as follows:

- Signal Processing - Linear algebra is used in encoding and manipulating signals such as audio and video signals. Furthermore, it is required in the analysis of such signals.
- Linear Programming - It is an optimizing technique that is used to determine the best outcome of a linear function.
- Computer Science - Data scientists use several linear algebra algorithms to solve complicated problems.
- Prediction Algorithms - Prediction algorithms use linear models that are developed using concepts of linear algebra.

Important Notes on Linear Algebra

- Linear algebra is concerned with the study of three broad subtopics - linear functions, vectors, and matrices
- Linear algebra can be classified into 3 categories. These are elementary, advanced, and applied linear algebra.
- Elementary linear algebra is concerned with the introduction to linear algebra. Advanced linear algebra builds on these concepts. Applied linear algebra applies these concepts to real-life situations.

What is a Tensor?

Tensors are simply mathematical objects that can be used to describe physical properties, just like scalars and vectors. In fact tensors are merely a generalisation of scalars and vectors; a scalar is a zero rank tensor, and a vector is a first rank tensor.

The rank (or order) of a tensor is defined by the number of directions (and hence the dimensionality of the array) required to describe it. For example, properties that require one direction (first rank) can be fully described by a 3×1 column vector, and properties that require two directions (second rank tensors), can be described by 9 numbers, as a 3×3 matrix. As such, in general an n^{th} rank tensor can be described by 3^n coefficients.

The need for second rank tensors comes when we need to consider more than one direction to describe one of these physical properties. A good example of this is if we need to describe the electrical conductivity of a general, anisotropic crystal. We know that in general for isotropic conductors that obey Ohm's law:

$$\boldsymbol{j} = \sigma \boldsymbol{E}$$

Which means that the current density **j** is parallel to the applied electric field, **E** and that each component of **j** is linearly proportional to each component of **E**. (e.g. $\boldsymbol{j}_1 = \sigma \boldsymbol{E}_1$).

However in an anisotropic material, the current density induced will not necessarily be parallel to the applied electric field due to preferred directions of current flow within the crystal (a good example of this is in graphite). This means that in general each component of the current density vector can depend on all the components of the electric field:

$$j_1 = \sigma_{11}E_1 + \sigma_{12}E_2 + \sigma_{13}E_3$$
$$j_2 = \sigma_{21}E_1 + \sigma_{22}E_2 + \sigma_{23}E_3$$
$$j_3 = \sigma_{31}E_1 + \sigma_{32}E_2 + \sigma_{33}E_3$$

So in general, electrical conductivity is a second rank tensor and can be specified by 9 independent coefficients, which can be represented in a 3×3 matrix as shown below:

$$\sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

Other examples of second rank tensors include electric susceptibility, thermal conductivity, stress and strain. They typically relate a vector to another vector, or another second rank tensor to a scalar. Tensors of higher rank are required to fully describe properties that relate two second rank tensors (e.g. Stiffness (4th rank): stress and strain) or a second rank tensor and a vector (e.g. Piezoelectricity (3rd rank): stress and polarisation).

Tensor Rank

The total number of contravariant and covariant indices of a tensor. The rank of a tensor isindependent of the number of dimensions of the underlying space.
An intuitive way to think of the rank of a tensor is as follows: First, consider intuitively that a tensor represents a physical entity which may be characterized by magnitude and multiple directions simultaneously. Therefore, the number of simultaneous directions is denoted andis called the rank of the tensor in question. In -dimensional space, it follows that a rank-0 tensor (i.e., a scalar) can be represented by number since scalars represent quantities with magnitude and no direction; similarly, a rank-1 tensor (i.e., a vector) in -dimensional space can be represented by numbers and a general tensor by numbers.

Calculus

Calculus is the branch of mathematics that deals with continuous change. Calculus is also called infinitesimal calculus or "the calculus of infinitesimals". The meaning of classical calculus is the study of continuous change of functions. Most of these quantities are the functions of time such as velocity is equal to change in distance with respect to time. The two major concepts of calculus are:

- Derivatives

- Integrals

The derivative is the measure of the rate of change of a function whereas integral is the measure of the area under the curve. The derivative explains the function at a specific point while the integral accumulates the discrete values of a function over a range of values.

Calculus Definition

Calculus, a branch of Mathematics, developed by Newton and Leibniz, deals with the study of the rate of change. Calculus Math is generally used in Mathematical models to obtain optimal solutions. It helps us to understand the changes between the values which are related by a function. Calculus Math mainly focused on some important topics such as differentiation, integration, limits, functions, and so on.

Calculus Mathematics is broadly classified into two different such:

- Differential Calculus
- Integral Calculus

Both the differential and integral calculus deals with the impact on the function of a slight change in the independent variable as it leads to zero. Both differential and integral calculus serves as a foundation for the higher branch of Mathematics known as "Analysis". Calculus Mathematics plays a vital role in modern Physics as well as in Science and technology.

What is the Meaning of Calculus?

Calculus means the part of maths that deals with the properties of derivatives and integrals of quantities such as area, volume, velocity, acceleration, etc., by processes initially dependent on the summation of infinitesimal differences. It helps in determining the changes between the values that are related to the functions.

Basic Calculus

Basic Calculus is the study of differentiation and integration. Both concepts are based on the idea of limits and functions. Some concepts, like continuity, exponents, are the foundation of advanced calculus. Basic calculus explains about the two different types of calculus called "Differential Calculus" and "Integral Calculus". Differential Calculus helps to find the rate of change of a quantity, whereas integral calculus helps to find the quantity when the rate of change is known.

Differential Calculus Basics

Differential Calculus is concerned with the problems of finding the rate of change of a function with respect to the other variables. To get the optimal solution, derivatives are used to find the maxima and minima values of a function. Differential calculus arises from the study of the limit of a quotient. It deals with variables such as x and y, functions f(x), and the corresponding changes in the variables x and y. The symbol dy and dx are called differentials. The process of finding the derivatives is called differentiation. The derivative of a function is represented by dy/dx or f’ (x). It means that the function is the derivative of y with respect to the variable x. Let us discuss some of the important topics covered in the basic differential calculus.

Limits

The degree of closeness to any value or the approaching term. A limit is normally expressed using the limit formula as-

limx→cf(x)=A

It is read as “the limit of f of x as x approaches c equals A”.

Derivatives

Instantaneous rate of change of a quantity with respect to the other. The derivative of a function is represented as:

limx→hf(x+h)−f(x)h=A

Continuity

A function f(x) is said to be continuous at a particular point x = a, if the following three conditions are satisfied –

- f(a) is defined
- limx→af(x) exists
- limx→a−f(x)=limx→a+f(x)=f(a)

Continuity and Differentiability

A Function is always continuous if it is differentiable at any point, whereas the vice-versa condition is not always true.

Quotient Rule

The Quotient rule is a method for determining the derivative (differentiation) of a function that is in fractional form.

Chain Rule

The rule applied for finding the derivative of the composition of a function is basically known as the chain rule.

Integral Calculus Basics

Integral calculus is the study of integrals and their properties. It is mostly useful for the following two purposes:

- To calculate f from f’ (i.e. from its derivative). If a function f is differentiable in the interval of consideration, then f’ is defined in that interval.
- To calculate the area under a curve.

Integration

Integration is the reciprocal of differentiation. As differentiation can be understood as dividing a part into many small parts, integration can be said as a collection of small parts in order to form a whole. It is generally used for calculating areas.

Definite Integral

A definite integral has a specific boundary within which function needs to be calculated. The lower limit and upper limit of the independent variable of a function is specified; its integration is described using definite integrals. A definite integral is denoted as:

$\int abf(x).dx=F(x)$

Indefinite Integral

An indefinite integral does not have a specific boundary, i.e. no upper and lower limit is defined. Thus the integration value is always accompanied by a constant value (C). It is denoted as:

$\int f(x).dx=F(x)+C$

Advanced Calculus

Advanced Calculus includes some topics such as infinite series, power series, and so on which are all just the application of the principles of some basic calculus topics such as differentiation, derivatives, rate of change and o on. The important areas which are necessary for advanced calculus are vector spaces, matrices, linear transformation. Advanced Calculus helps us to gain knowledge on a few important concepts such as

- Quadratic forms
- Generalized Stokes theorem

- Vector fields as derivatives
- Integration of forms
- Multilinear algebra
- Continuous differentiability
- Tangent space and normal space via gradients
- Dual space and dual basis
- Critical point analysis for multivariate functions, etc.

Applications of Calculus

Calculus is a Mathematical model, that helps us to analyze a system to find an optimal solution to predict the future. In real life, concepts of calculus play a major role either it is related to solve the area of complicated shapes, safety of vehicles, evaluating survey data for business planning, credit card payment records, or finding the changing conditions of a system affect us, etc. Calculus is a language of physicians, economists, biologists, architects, medical experts, statisticians and it is often used by them. For example, Architects and engineers use concepts of calculus to determine the size and shape of the curves to design bridges, roads and tunnels, etc. Using Calculus, some of the concepts are beautifully modelled, such as birth and death rates, radioactive decay, reaction rates, heat and light, motion, electricity, etc.

Problems and Solutions

Go through the below calculus problems to understand the process of differentiation and integration.

Problem 1: Let $f(y) = e^y$ and $g(y) = 10y$. Use the chain rule to calculate $h'(y)$ where $h(y) = f(g(y))$.

Solution: Given,

$f(y) = e^y$ and

$g(y) = 10y$

First derivative above functions are

$f'(y) = e^y$ and

$g'(y) = 10$

To find: $h'(y)$

Now, $h(y) = f(g(y))$

$h'(y) = f'(g(y))g'(y)$

$h'(y) = f'(10y)10$

By substituting the values.

$h'(y) = e^{10y} \times 10$

or $h'(y) = 10\, e^{10y}$

Problem 2: Integrate sin 3x + 2x with respect to x.

Solution: Given instructions can be written as:

$\int \sin 3x + 2x\, dx$

Use the sum rule, which implies

$\int \sin 3x\, dx + \int 2x\, dx$ ……… Equation 1

Solve $\int \sin 3x\, dx$ first.

use substitution method,

let 3x = u => 3 dx = du (after derivation)

or dx = 1/3 du

=> $\int \sin 3x\, dx$ turned as $\int \sin u \times 1/3\, du$

or $1/3 \int \sin u\, du$

which is 1/3 (-cos u) + C, where C= constant of integration

Substituting values again, we get

$\int \sin 3x\, dx = -\cos(3x)/3 + C$ ……… Equation 2

Solve $\int 2x\, dx$

$\int 2x\, dx = 2\int x\, dx = 2 * x^2/2 + C = x^2 + C$ ……. Equation 3

Equation (1) => $\int \sin 3x\, dx + \int 2x\, dx$

$= -\cos(3x)/3 + x^2 + C$

Practice Questions

- Differentiate $f(x) = 6x^3 – 9x + 4$
- Differentiate $f(x) = x^3 – 2x^2 + x – 1$
- Find: $\int 6x^5 – 18x^2 + 7\, dx$
- Find the area under the curve for $y = x^2 + 2$, $y = \sin x$, $x = -1$ and $x = 2$

Why Is Calculus Important to Machine Learning?

So, why is calculus used so much to describe machine learning algorithms?

Machine learning is based on finding the optimal way to describe data, so we can use that same way to predict data we haven't seen. To consider what is optimal and what is not, calculus is the tool we use.

Calculus is the mathematics of change.

It provides useful tools to check how things will change caused by perturbation in something else. It also help us understand the cause and effect of an algorithm.

Calculus is not obscure. It is the language for modelingbehaviors. Without calculus, we would not be able to fully understand techniques such as:

- Backpropagation in neural networks
- Regression using optimal least square
- Expectation maximization in fitting probability models

Areas of Calculus to Focus On

You don't need to know all of calculus

The three key areas of calculus that I recommend you focus on are:

1. Differentiation

Differentiation is the key thing in calculus. It describes the rate of change or the cause and effect of tuning parameters.

Algorithms described in books, papers, and on websites usually involves differentiations. You need to know what it means and its notations.

2. Vector calculus

This brings differentiation to a higher dimension. Usually, machine learning algorithms involve more than one parameter. Sometimes, there are multiple outputs from a single model. We typically describe such machine learning algorithms with vector functions and use multivariate calculus to describe their behavior.

You need to know how to do differentiation on a vector function and how to present it as a vector of a matrix. This is the tool behind backpropagation algorithms in neural network training.

3. Calculus as a Tool for Optimization

One important use of calculus, differentiation in particular, is to find the optimum value of a function.

While we may find the maximum or minimum of a function algorithmically, using calculus can give you the answer in one shot. More importantly, it is also the tool to find the maximum or minimum under constraints.

One example is the support vector machine classifier. This is essentially finding the maximum separation of two classes. You need to understand calculus in order to know how support vector machine gives its solution.