Math Refresher

Probability and Statistics

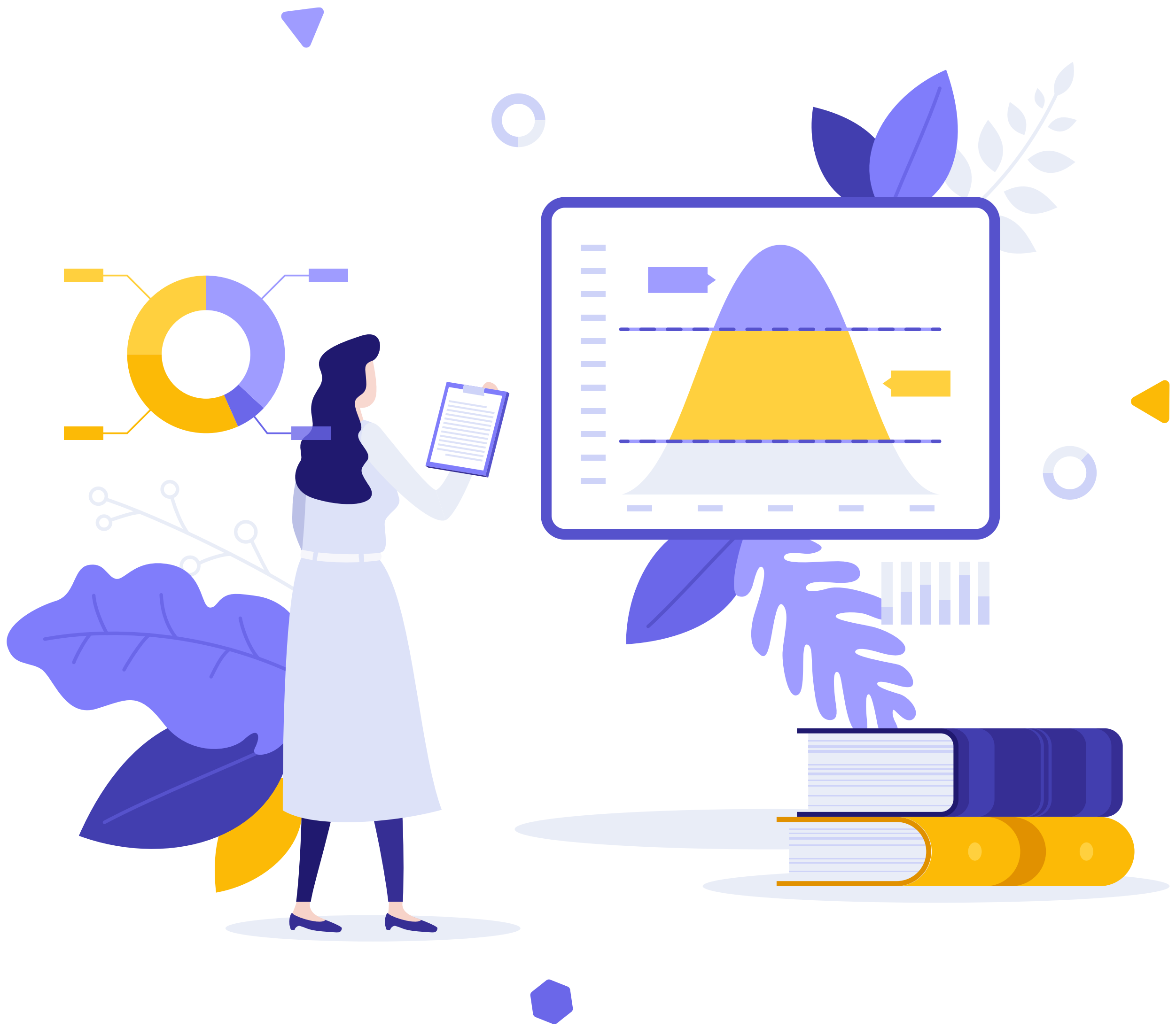
Learning Objectives

By the end of this lesson, you will be able to:

- Explain the concepts of probability and statistics
- Discuss types of data
- Discuss measures of central tendency, asymmetry, variability
- Explain different types of probability
- Explain the difference between mean and expectation

Basics of Probability and Statistics

Probability and Statistics



- Data science heavily relies on estimates and predictions.
- Evaluations and forecasts constitute a significant portion of the field.
- For data analysis, data scientists use statistical methods to make estimates.

Probability and Statistics

- Probability theory helps in making predictions.
- Statistical methods heavily rely on probability theory, and both probability and statistics rely on data.

Applications of Probability and Statistics

They have wide applications across various sectors. Some of them are listed below:

Data analysis	They provide the foundation for analyzing and understanding data.
Statistical modeling	These models can be used for prediction, forecasting, and understanding the underlying mechanisms of various phenomena.
Experimental design	They provide techniques for sample selection, hypothesis testing, and controlling for confounding factors, ensuring reliable and unbiased results.
Machine learning	They use techniques such as Bayesian inference, regression analysis, and hypothesis testing are used to train models, evaluate their performance, and make predictions.

Applications of Probability and Statistics

Data visualization	➡	Understanding statistical concepts allows for accurate representation and interpretation of data through charts, graphs, and other visual formats.
Decision-making and risk assessment	➡	Probability and statistics quantify uncertainty and risk and can help weigh different options for making optimal choices.
Anomaly detection and quality control	➡	Probability and statistical techniques are vital in identifying anomalies and outliers in data, which can indicate errors, fraud, or other unusual patterns.

Data in Probability and Statistics

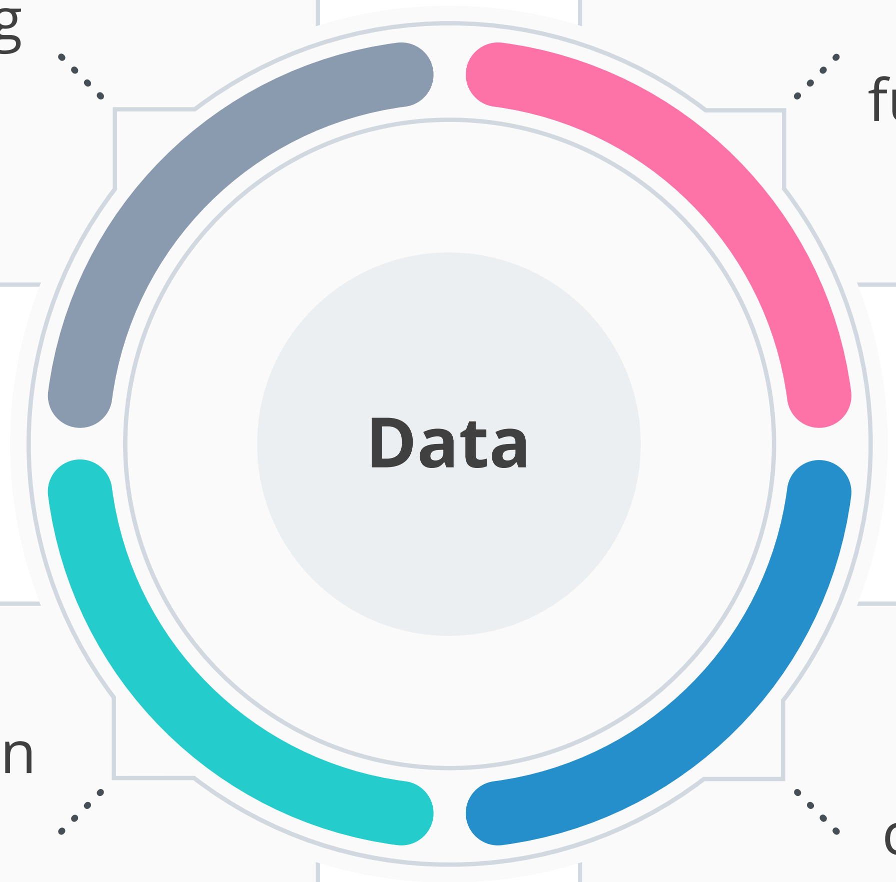
Data refers to information obtained through observations, facts, and measurements for reference or research purposes.

Data is a collection of facts, including numbers, words, estimates, and perspectives, organized in a format that computers can interpret.

Importance of Data

Data facilitates a deeper understanding of information by identifying possible connections between two features.

Considering previous information patterns, data can be used to forecast future events or understand an ongoing situation.

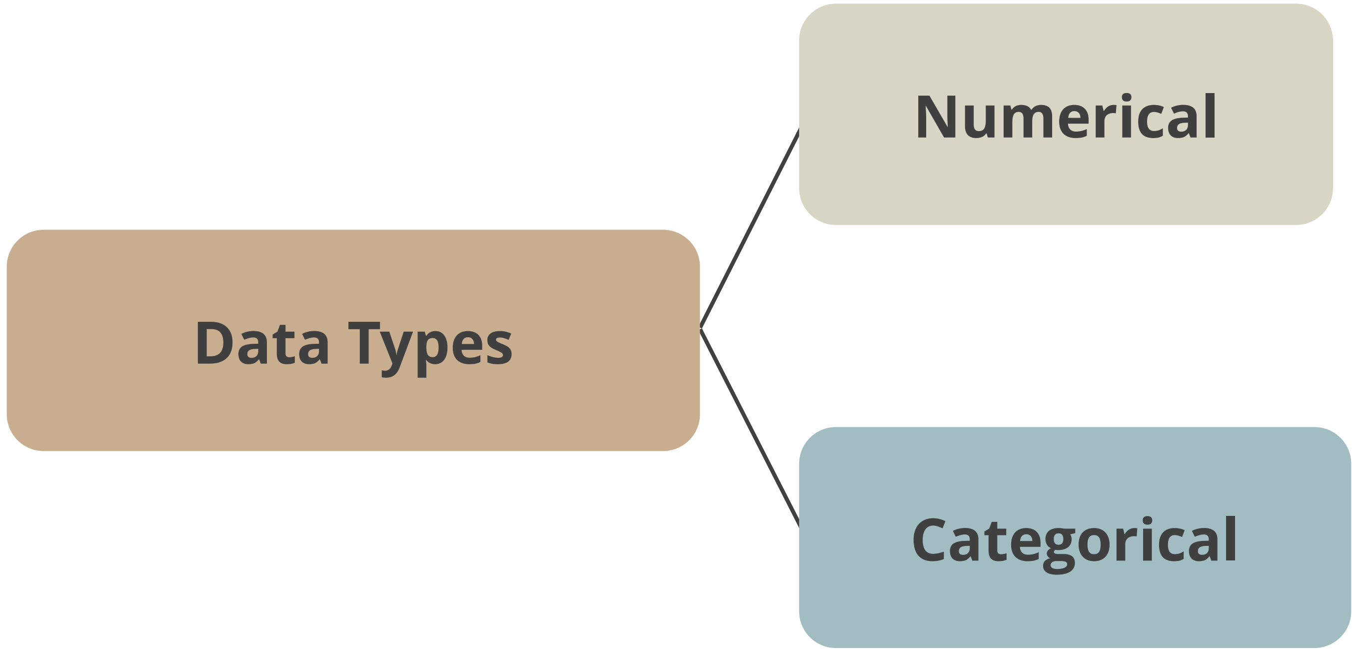


Data helps uncover hidden patterns, enabling the detection of anomalies or distortions.

Data helps determine the common patterns between two pieces of information.

Types of Data

Data might be numerical (such as age) or categorical (such as gender).



Numerical Data

Numerical data consists of values that are expressed as numbers and can be of the following two types:

Continuous numerical data

- It can take any numerical value within a range.
- It includes measurements such as height, weight, temperature, time.

Discrete numerical data

- It consists of whole numbers or counts that can only take specific values.
- For example, the number of students in a class, the number of items sold, or the number of cars in a parking lot.

Categorical Data

Categorical data represents characteristics or attributes that fall into distinct categories.

Nominal categorical data

- It consists of categories or labels that do not have any inherent order or numerical value.
- It includes gender, colors, or categories like yes or no.

Ordinal categorical data

- It possesses a natural order or ranking among categories.
- It includes survey ratings, educational levels, or satisfaction levels.

Types of Data: Example

A person's bank data may be categorized into numerical and category data.

RowNumber	CustomerID	Surname	Geography	Gender	Age	Balance	HasCrCard	IsActiveMember
1	15634602	Hargrave	France	Female	42	0	1	1
2	15647311	Hill	Spain	Female	41	83807.86	0	1
3	15619304	Onio	France	Female	42	159660.8	1	0
4	15701354	Boni	France	Female	39	0	0	0
5	15737888	Mitchell	Spain	Female	43	125510.8	1	1
6	15574012	Chu	Spain	Male	44	113755.8	1	0
7	15592531	Bartlett	France	Male	50	0	1	1
8	15656148	Obinna	Germany	Female	29	115046.7	1	0
9	15792365	He	France	Male	44	142051.1	0	1
10	15592389	H?	France	Male	27	134603.9	1	1
11	15767821	Bearce	France	Male	31	102016.7	0	0
12	15737173	Andrews	Spain	Male	24	0	1	0
13	15632264	Kay	France	Female	34	0	1	0
14	15691483	Chin	France	Female	25	0	0	0
15	15600882	Scott	Spain	Female	35	0	1	1

Numerical columns include the following:

CustomerID, Age, Balance

Categorical columns include the following:

Geography, Gender, HasCrCard, IsActiveMember

Scale of Measurement

Scale of Measurement

The scale of measurement determines the mathematical operations that can be applied to the data and the suitable statistical analysis.

Data measurement levels

- Qualitative
  - Nominal
  - Ordinal
- Quantitative
  - Interval
  - Ratio

Qualitative Scale of Measurement

Nominal

- Data is categorized using names, labels, or qualities.
- Data can be represented using text, codes, or symbols.
- Example: Brand name, zip code, gender

Ordinal

- Data can also be arranged in an ordered or ranked manner, allowing for comparison.
- The magnitude of differences between categories may not be quantifiable or uniform.
- Example: Grades, star reviews, position in a race, date

Quantitative Scale of Measurement

Quantitative scale of measurement includes the following:

Interval

- Data can be ordered in a range of values where meaningful differences between the data points can be calculated.
- Interval data is measured on a numerical scale where the intervals between values are equal.
- Example: Temperature in Celsius, year of birth

Ratio

- Data at this level is similar to that at the interval level with the added property of inherent zero.
- At this level, mathematical calculations can be performed on the data points.
- Example: Height, age, weight

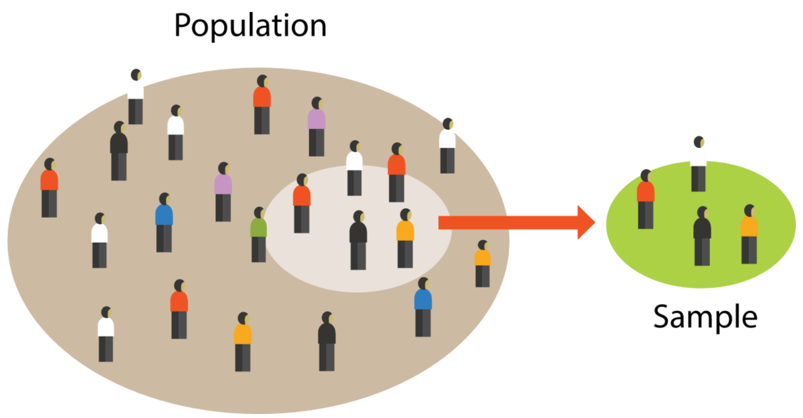
Population vs. Sample

Population vs. Sample

Before analyzing any data, it is crucial to determine whether it is derived from a population or a sample.

Population

- A population is a collection of all available items (N), as well as each unit in the study.
- Population data is used when the data pool is very small and can provide all the required information.



Sample

- A sample is a subset of the population (n) that contains only a few units of the population.
- Samples are collected randomly and represent the population.

Population vs. Sample: Example

Consider a nationwide research study focusing on students recruits 1000 students:

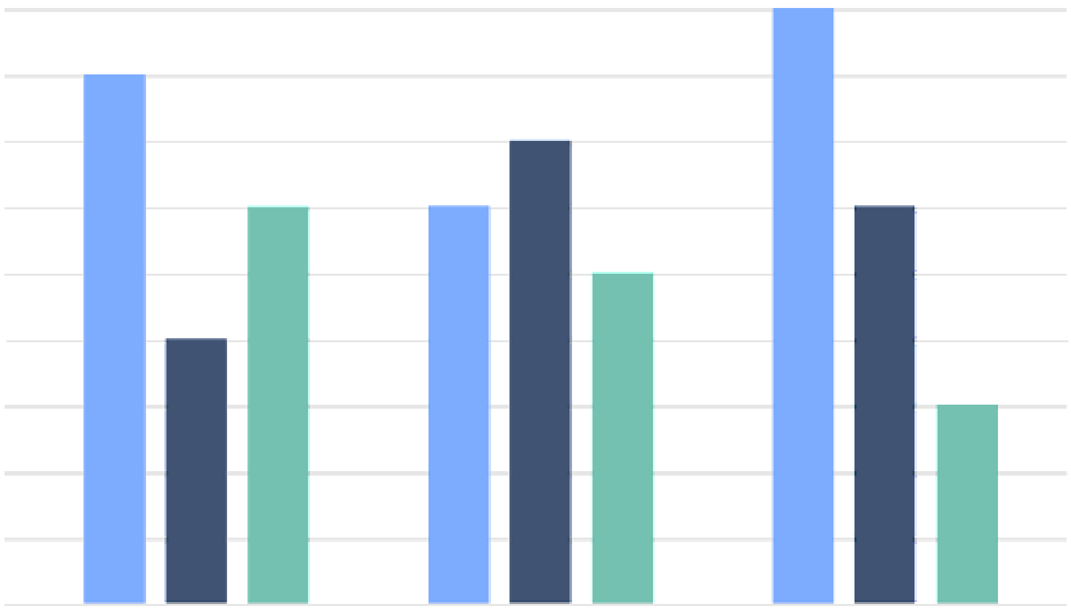
Population: Every student across the country

Sample: The set of students picked for the study

Introduction to Descriptive Statistics

Descriptive Statistics

Descriptive statistics is a branch of statistics that focuses on summarizing and describing the main characteristics of a dataset.



- It provides methods and tools to organize, analyze, and present data in a meaningful and concise manner.
- It aims to provide a clear and concise summary of the data.
- It enables researchers, analysts, and decision-makers to gain insights and make informed interpretations.

Descriptive Statistics

The main objectives of descriptive statistics are as follows:

Data description	To summarize the essential features of a dataset
Data visualization	To identify patterns, trends, and relationships
Data organization	To organize and structure the data for easier analysis
Data comparison	To compare different datasets or subgroups within a dataset
Data interpretation	To provide meaningful interpretations and summaries of the data

Measures of Central Tendencies

Measures of Central Tendencies

Central tendency refers to a single value that helps describe the center position of a dataset.

Measures of central tendency, also called summary statistics, describe the center position of the data.

The most used measures of central tendency are the mean, median, and mode.

Measures of Central Tendencies: Mean

Mean

- It is calculated by dividing the sum of all data values by the total number of data values.
- It is affected by unusual or extreme values.

Measures of Central Tendencies: Mean

The formula for calculating mean is given below:

$$\text{Arithmetic mean} = \frac{X_1 + X_2 + X_2 + \ldots\ldots\ldots\ldots + Xn}{n}$$

Example

Data values: 7, 3, 4, 1, 6, 7

Mean = 7+3+4+1+6+7
= 28 / 6
= 4.66

Measures of Central Tendencies: Median

The median is the middle value of a sorted list of numbers.

Median

It is less affected by outliers and skewness, making it a robust measure.

Measures of Central Tendencies: Median

The formulas for calculating median are given below:

If the total number of values is odd, then:

$$median = (\frac{n+1}{2})^{th} term$$

Example

7, 3, 4, 1, 6
After sorting:
1, 3, 4, 6, 7
Median = 4

If the total number of values is even, then:

$$median = (\frac{(\frac{n}{2})^{th} term + (\frac{n}{2} + 1)^{th} term}{2})^{th} term$$

Example

7, 3, 4, 1, 7, 6
After sorting:
1, 3, 4, 6, 7, 7
Median = 4+6 / 2 = 5

Measures of Central Tendencies: Mode

Mode

The mode represents the most common value in the dataset.

It is unaffected by extreme observations.

It is the preferred measure of central tendency for highly skewed or non-normal distributions.

Measures of Central Tendencies: Mode

The formula for calculating mode is given below:

Mode formula = $L + \frac{(f_m - f_1)h}{(f_m - f_1) + (fm - f_2)}$

Where

- h is the size of the class interval
- L is the lower limit of the class interval of the modal class
- f_1 is the modal class frequency
- f_m is the preceding class frequency
- f_2 is the succeeding class frequency

Example

7, 3, 4, 1, 6, 7
Mode = 7

Mean vs. Expectation

Mean vs. Expectation

Mean and expectation can be distinguished based on their definitions.

μ

The mean refers to the average of values in a dataset.

Σ

The expectation refers to the average value of a random variable.

Mean vs. Expectation: Example

Mean

- If the data is [2, 4, 6, 8, 10]
- Mean = (2 + 4 + 6 + 8 + 10) / 5 = 30 / 5 = 6

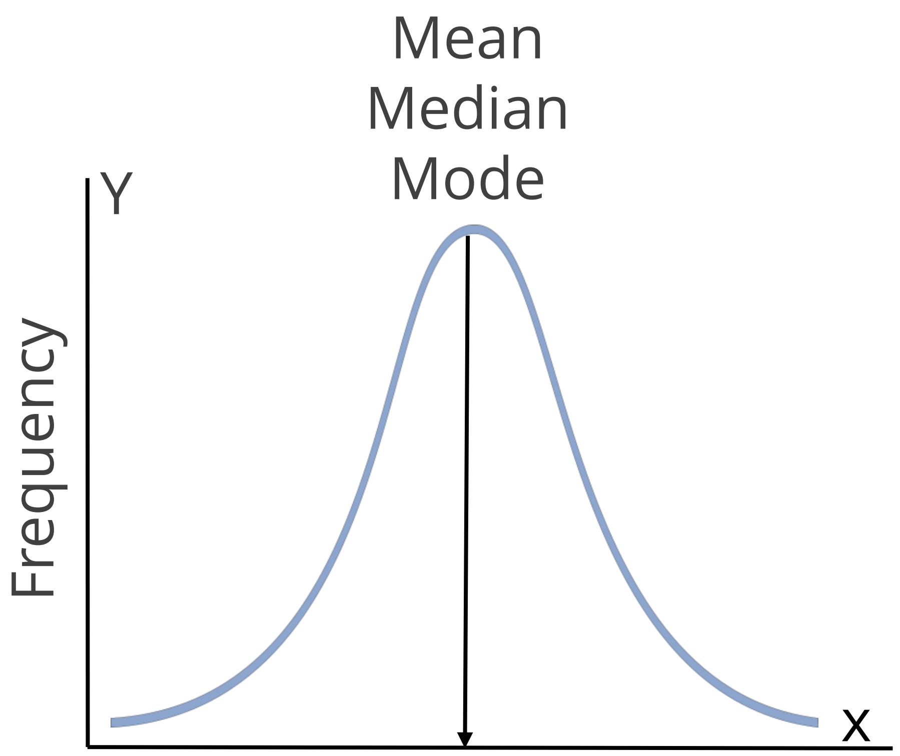
Expectation

- Consider a fair six-sided die with numbers 1 to 6.
- The expectation of rolling the die:
- E[X] = (1/6) * 1 + (1/6) * 2 + (1/6) * 3 + (1/6) * 4 + (1/6) * 5 + (1/6) * 6 = 3.5

Measures of Asymmetry

Measures of Asymmetry: Skewness

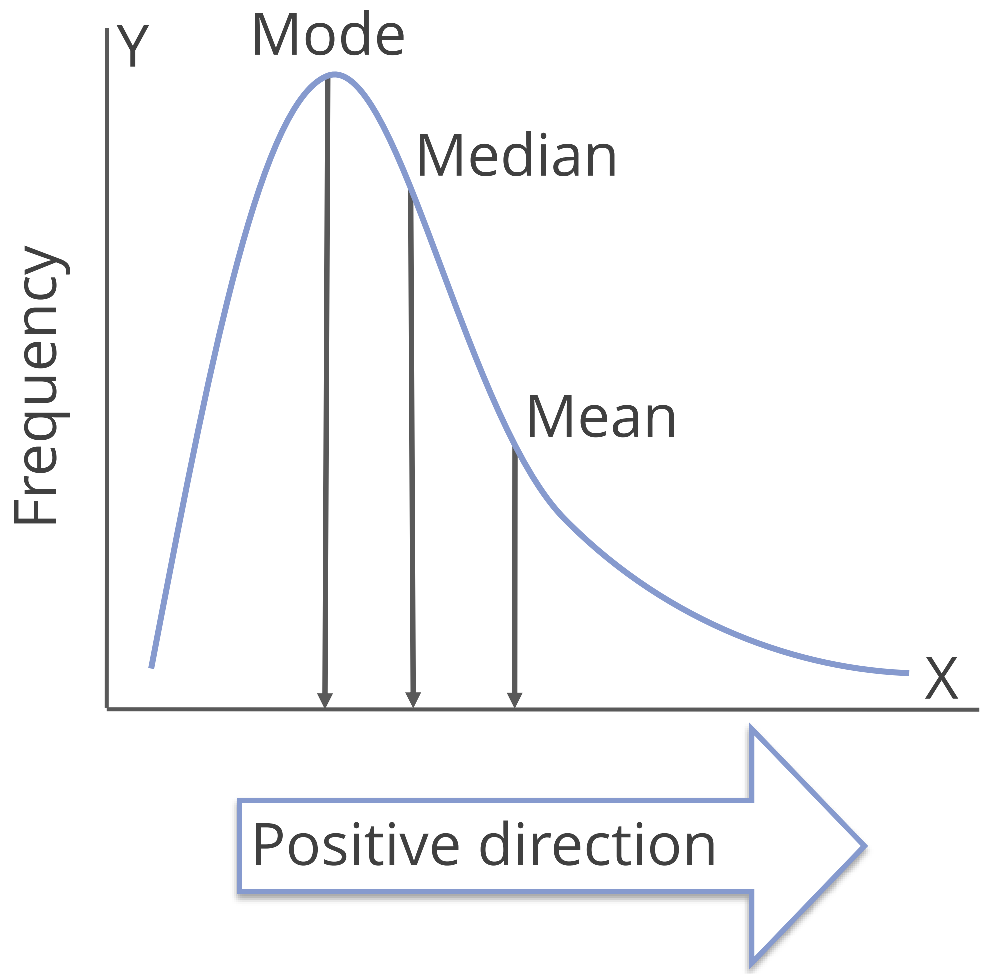
Skewness refers to a type of asymmetry in the distribution of statistical data. It occurs when curves are distorted or skewed to the right or left side.



A normal distribution curve, also known as a symmetrical bell curve, exhibits no skewness.

Skewness is crucial for data interpretation as it provides insights into the distribution of the data.

Measures of Asymmetry: Positive Skewness

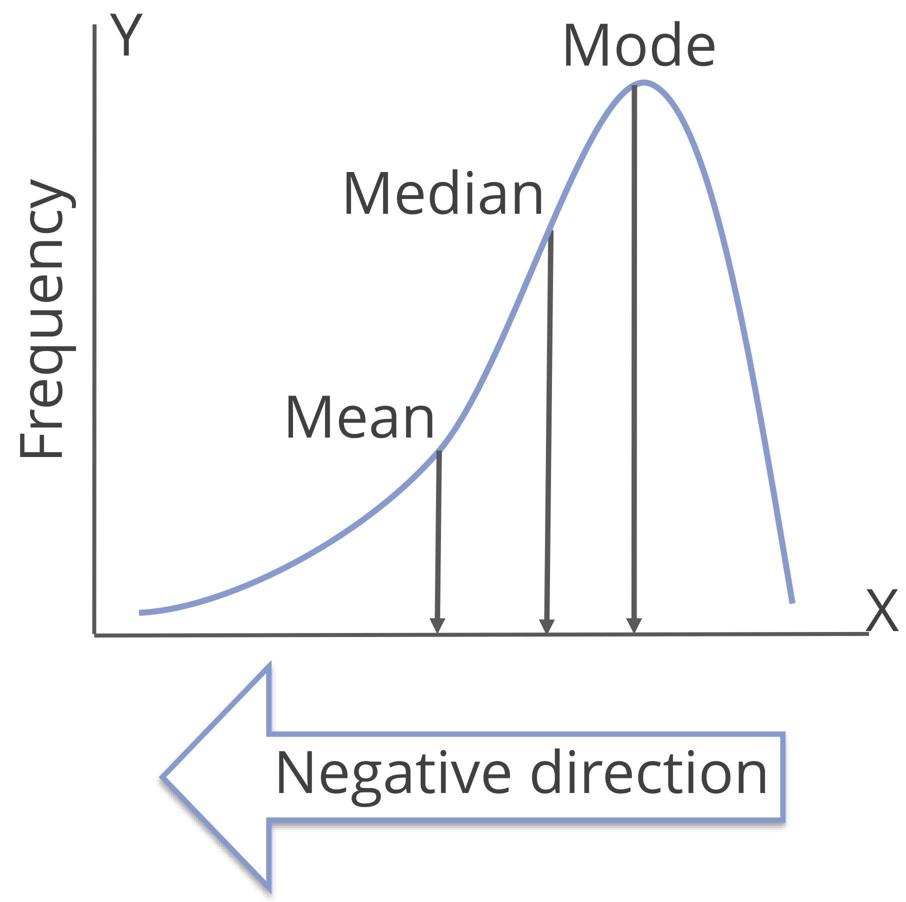


In positive skewness,
mean > median > mode

The tail of the distribution is skewed to the right. That is, the outliers are also skewed to the right.

Most of the points are concentrated on the left slide of the curve.

Measures of Asymmetry: Negative Skewness



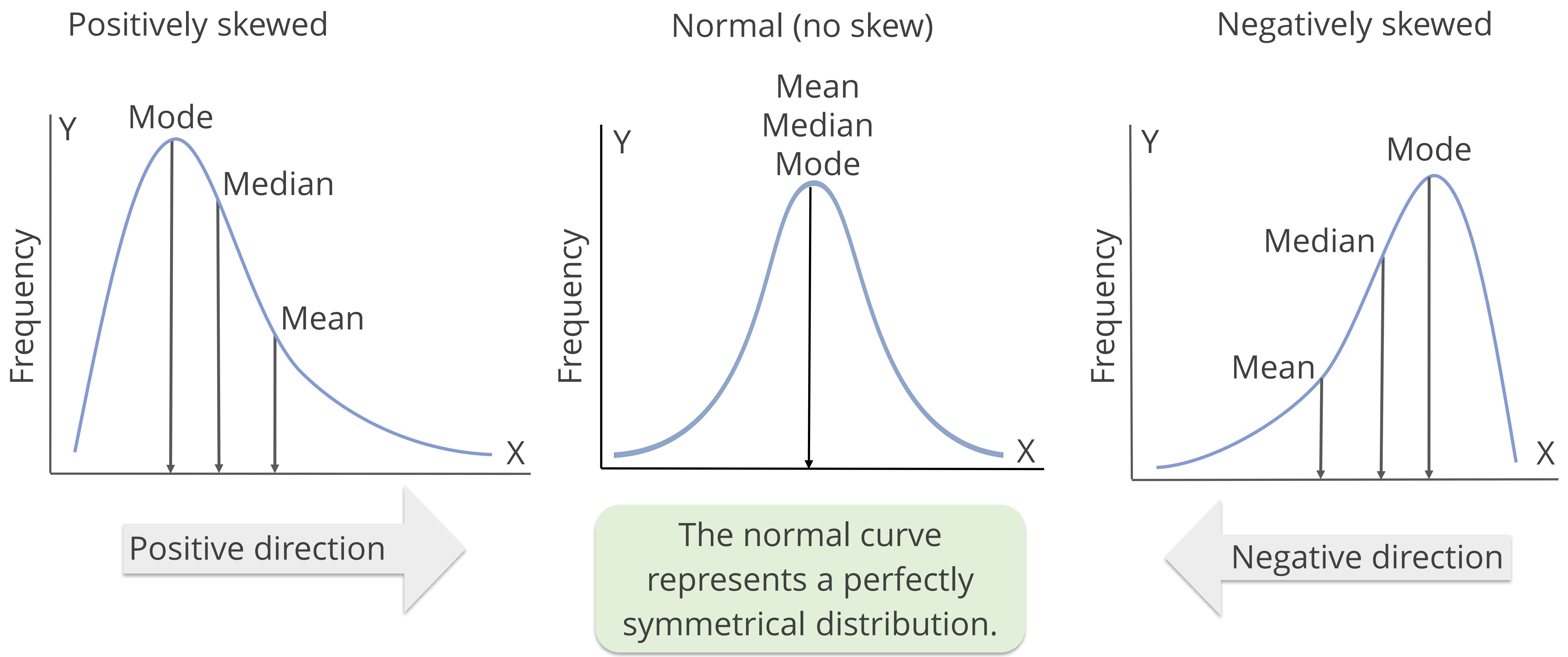
Most of the values are concentrated on the right side of the curve.

In negative skewness,
mean < median < mode

The left tail of the distribution is skewed, that is, the outliers are skewed to the left.

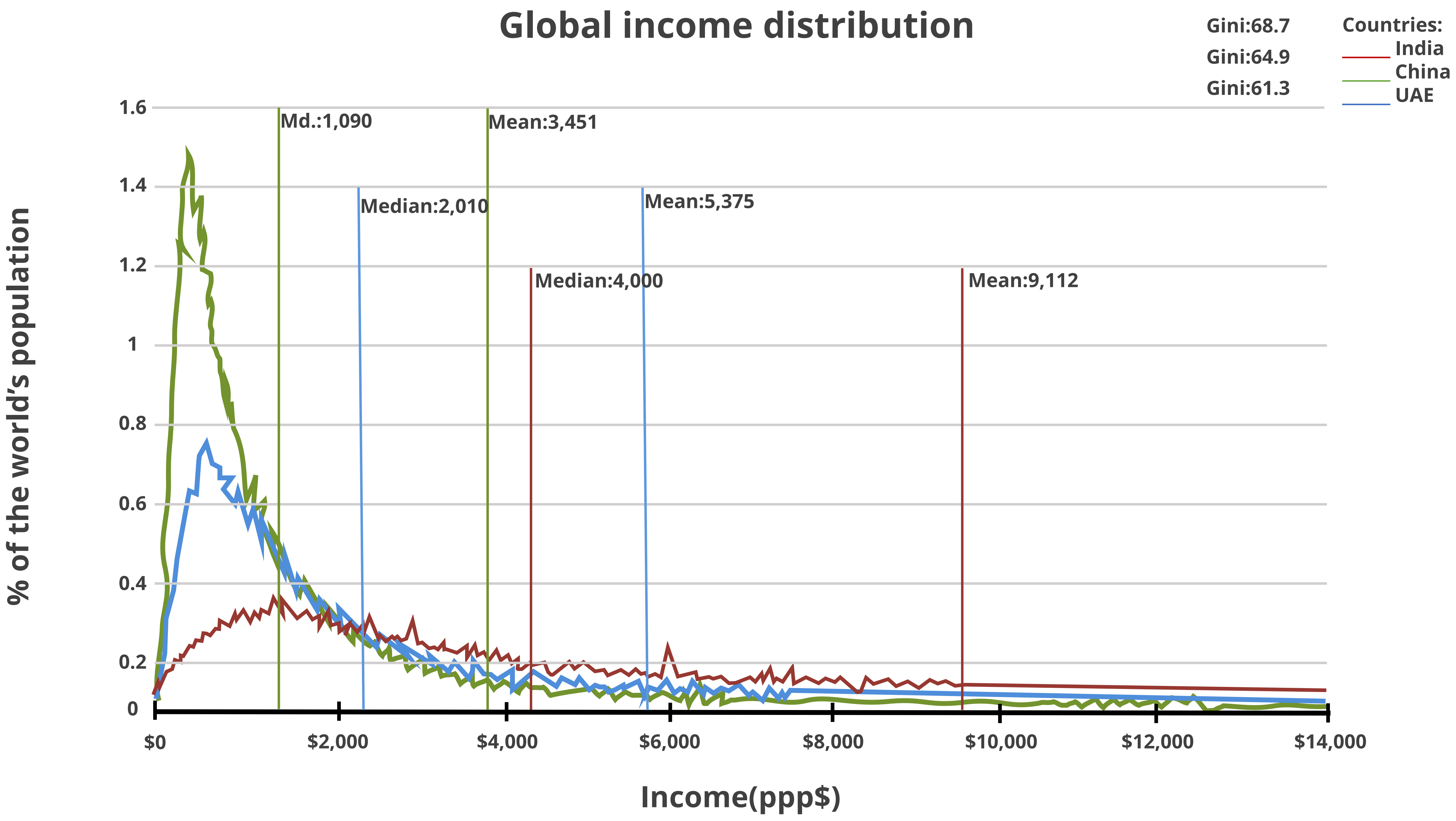
Measures of Asymmetry

Following are three different types of graphs:



Measures of Asymmetry: Example

The global income distribution shows a highly right-skewed pattern for different countries.



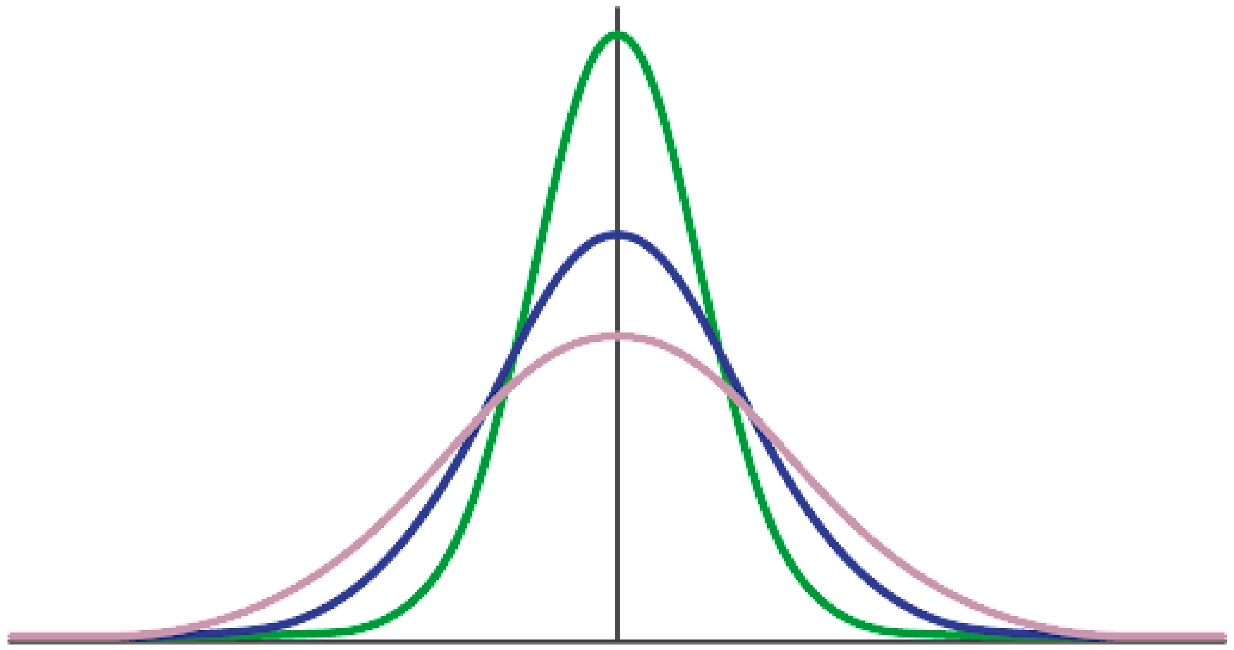
Measures of Asymmetry: Example

Key observations from the previous graph are as follows:

- The mean income of $3,451 is higher than the median income of $1,090.
- The global income is not evenly distributed.
- The majority of the population earns less than $2,000 annually.
- Only a small percentage of the population earns more than $14,000.

Measures of Asymmetry: Kurtosis

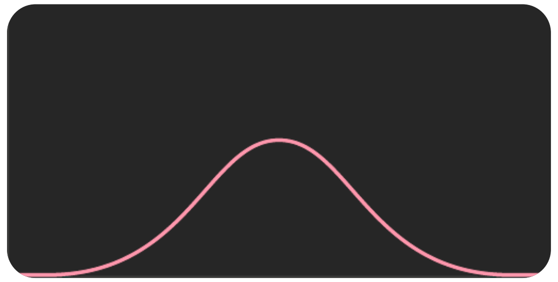
Kurtosis is a statistical measure that quantifies the extent to which the tails of a distribution deviate from those of a normal distribution.



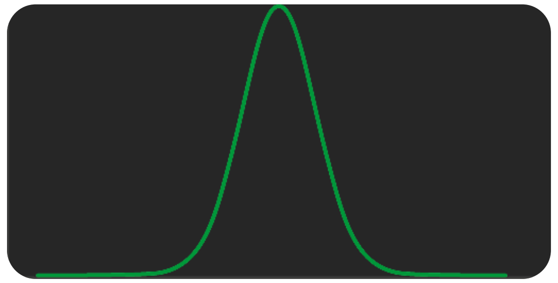
It indicates if the distribution is flatter or peaked compared to a normal distribution; it doesn’t affect the mean, median, or mode.

Measures of Asymmetry: Kurtosis

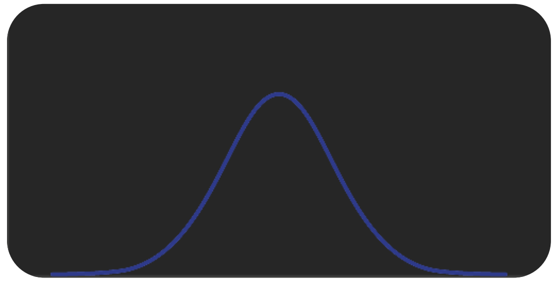
There are three types of kurtosis:



Platykurtic is negative kurtosis.



Leptokurtic is positive kurtosis.



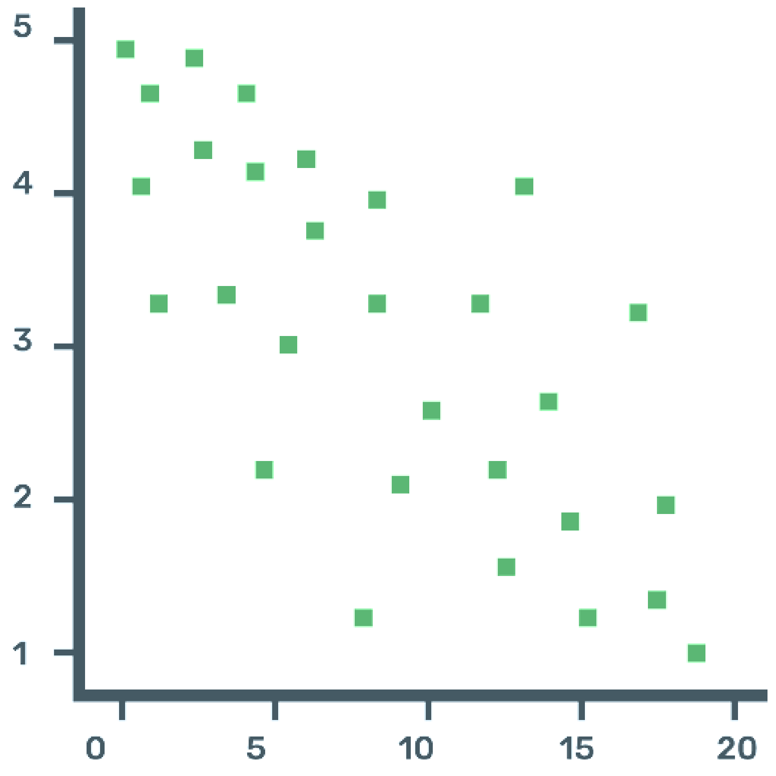
Mesokurtic distributions are normal distributions.

Source: https://indiafreenotes.com/kurtosis/

Measures of Variability

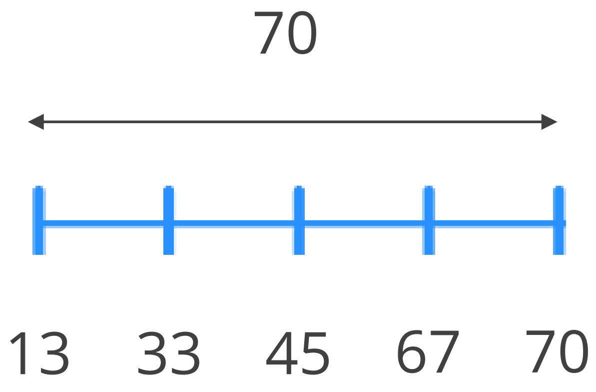
Measures of Variability: Dispersion

Measures of dispersion describe the spread of the data.



- The measure of central tendency provides a single value that represents the overall value; however, it cannot capture the complete perspective.
- The metric of dispersion allows us to examine the variability or inconsistency in the spread of data.
- Examples of dispersion measures include the range, interquartile range, standard deviation, and variance.

Measures of Variability: Range



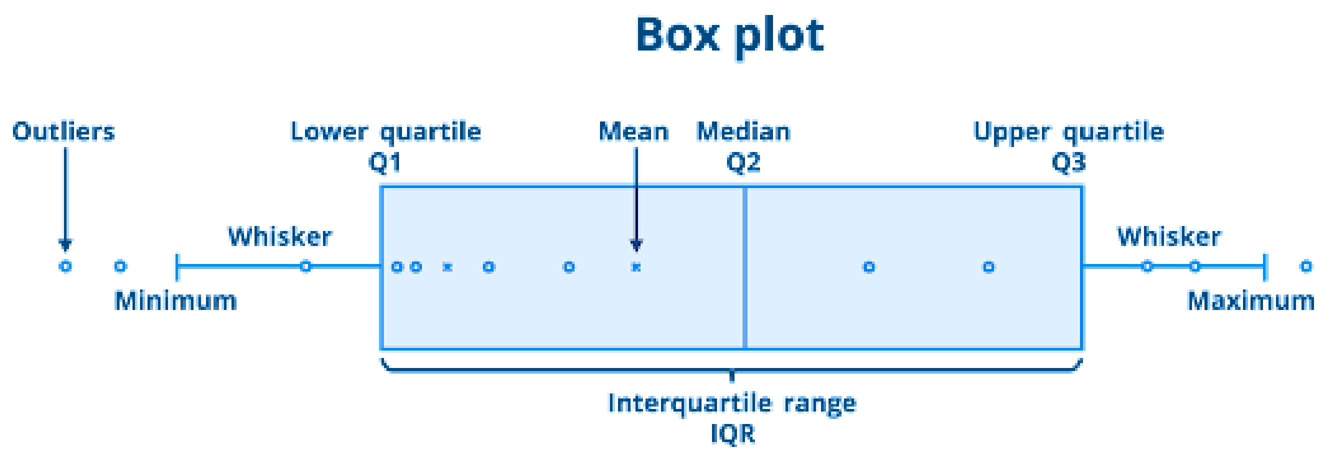
- The range of a distribution is determined by calculating the difference between the largest and smallest values in the dataset.
- The range alone does not provide a comprehensive view of the entire distribution.
- It primarily focuses on extreme values and may overlook other important aspects that are considered less significant.

Example

For {13,33,45,67,70} the range is 57 i.e., (70-13)

Measures of Variability: Interquartile Range

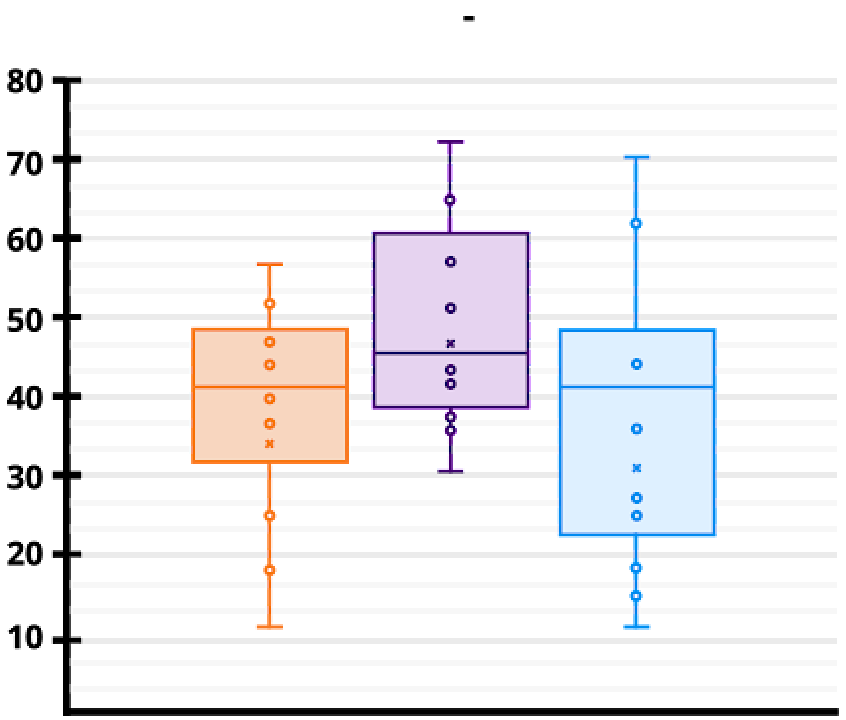
The interquartile range (IQR) is a statistical measure that provides insights about the spread or dispersion of a dataset.



- The IQR represents the range that contains the middle 50% of the data.
- It provides a measure of variability that is less influenced by extreme values or outliers compared to the range.

Interquartile Range: Example

Consider the following data:



Data: [10, 12, 15, 18, 20, 21, 22, 25, 30]
Lower half: [10, 12, 15, 18]
Upper half: [21, 22, 25, 30]
Q1 = 13.5, Q3 = 23.5
IQR = 23.5 – 13.5 = 10

Measures of Variability: Variance

Variance

Variance is the average of all squared deviations.

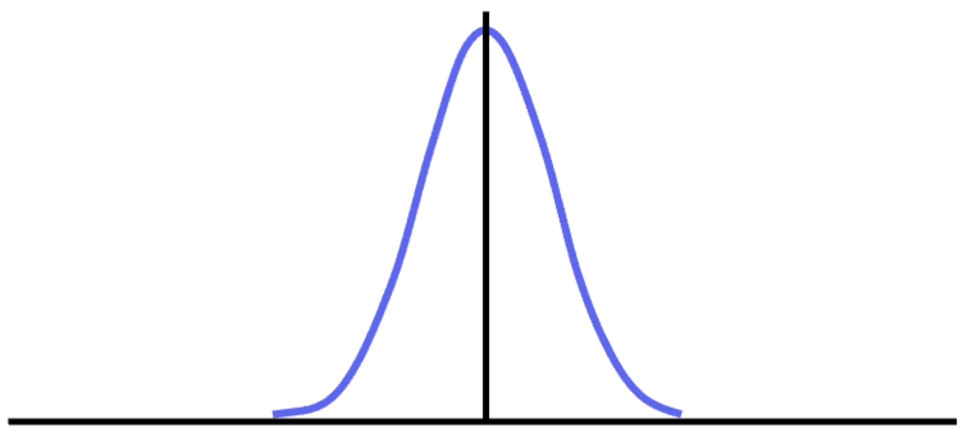
It is the sum of squared distances between each point and the mean, representing the dispersion around the mean.

The standard deviation is used instead of variance because it accounts for the unit difference that variance lacks.

Measures of Variability: Variance

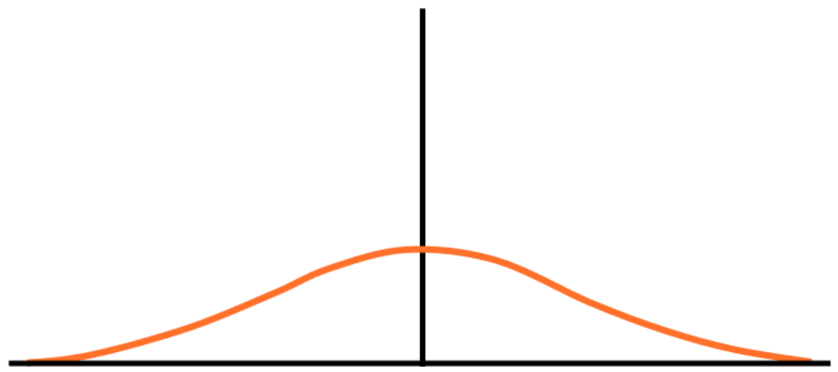
Types of variance:

Low variance



The data points are similar and not very far away from the mean.

High variance



Data values vary and are farther from the mean.

Measures of Variability: Variance

The formulas for calculating variance are as follows:

Population variance

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

Sample variance

$$s^2 = \frac{\sum (X - \overline{X})^2}{n - 1}$$

As the units of values and variance are not equal, another measure of variability is used.

Measures of Variability: Standard Deviation

Standard deviation

It quantifies the variability or dispersion around an average.

It is calculated as the square root of the variance.

It indicates the concentration of data around the mean of the dataset.

Measures of Variability: Standard Deviation

The formula for calculating variance is given below:

Population standard deviation:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Sample standard deviation:

$$S = \sqrt{\frac{\sum (X - \overline{X})^2}{n - 1}}$$

Measures of Variability: Example

Finding the mean, variance, and standard deviation for the following dataset:

Consider a dataset with the following values:
{3,5,6,9,10}

$$\text{Mean} = \frac{3+5+6+9+10}{5} = 6.6$$

$$\text{Variance} = \frac{(3-6.6)^2+(5-6.6)^2+(6-6.6)^2+(9-6.6)^2+(10-6.6)^2}{5} = 6.64$$

$$\text{Standard deviation} = \sqrt{\text{variance}} = \sqrt{6.64} = 2.576$$

Measures of Relationship

Measures of Relationship: Covariance

Relationship measures are employed to compare two variables.

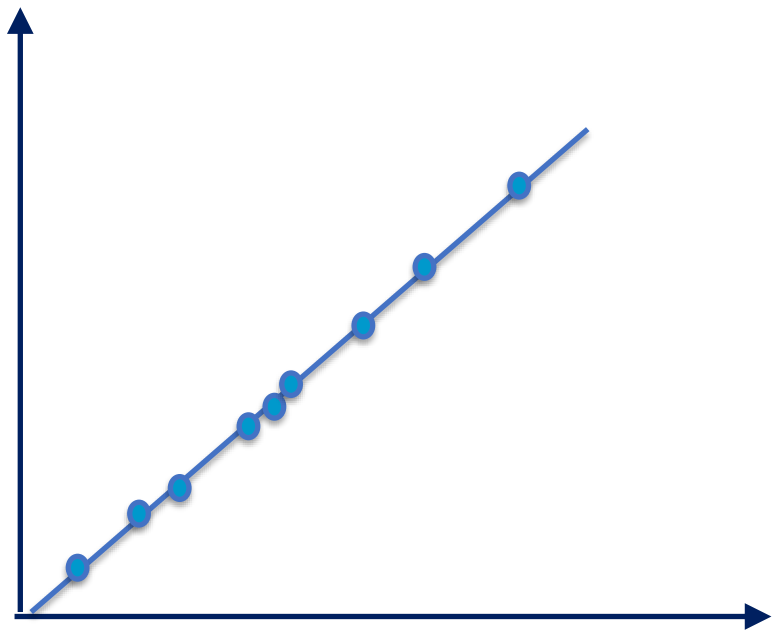
The covariance of two variables is a measure of their relationship.

It calculates the degree of change in the variables.

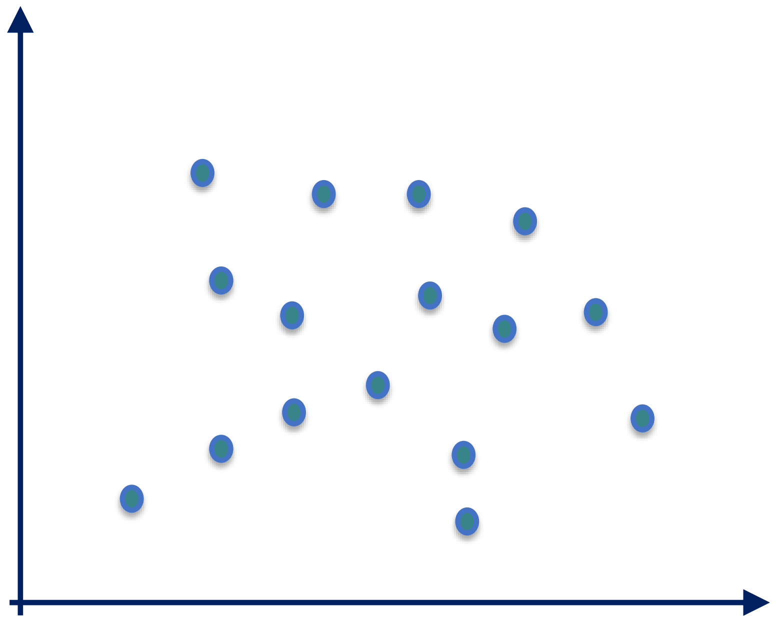
It determines if there is a similar association between two variables.

Types of Correlation

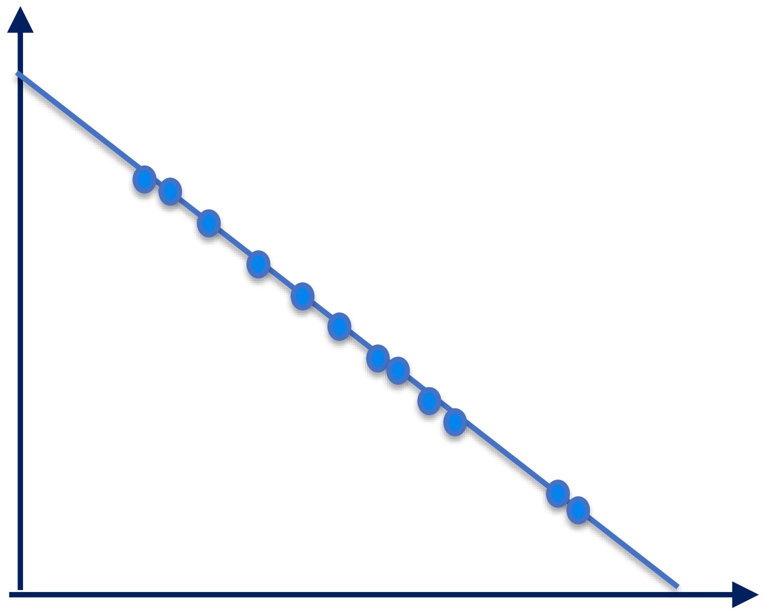
There are three types of correlations:



Perfect positive correlation: As one variable increases, the other variable tends to increase



Zero correlation: There is no linear relationship between the variables



Perfect negative correlation: As one variable increases, the other variable tends to decrease, and vice versa

Measures of Relationship: Correlation

Correlation offers a more comprehensive understanding of covariance.

It is a normalized version of covariance.

It measures the degree of change between variables.

It is commonly referred to as the Pearson Correlation Coefficient.

Measures of Relationship: Correlation

The formula for calculating correlation is as follows:

$$Correlation = \rho = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

The value of correlation ranges from -1 to 1.

Measures of Relationship: Correlation

Correlation = 1

- A correlation of 1 implies a positive relationship.
- Herein, an increase in one independent variable corresponds to an increase in the dependent variable.

Correlation = -1

- A correlation of -1 indicates a negative relationship.
- Herein, an increase in one independent variable corresponds to a decrease in the dependent variable.

Correlation = 0

- A correlation coefficient of 0 indicates independence between the variables.

Measures of Relationship: Example

Consider the following example for calculating correlation:

Height	Weight	$x - \overline{x}$	$y - \overline{y}$	$(x - \overline{x})(y - \overline{y})$	$(x - \overline{x})^2$	$(y - \overline{y})^2$
5	45	-0.14	-5	0.7	0.019	25
5.5	53	-0.36	3	-1.08	0.129	9
6	70	0.86	20	17.2	0.739	400
4.7	42	-0.44	-8	3.52	0.193	64
4.5	40	-0.64	-10	6.4	0.409	100

Measures of Relationship: Example

The formula for calculating correlation is given below:

Sum (height) = 25.7; Mean (height) = 5.14

Sum (weight) = 250; Mean (weight) = 50

$$\sum(x-\bar{x})(y-\bar{y}) = 26.74$$

$$\sum(x-\bar{x})^2 = 1.489$$

$$\sum(y-\bar{y})^2 = 598$$

$$\text{Correlation} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}$$

$$= \frac{26.74}{\sqrt{1.489}\sqrt{598}}$$

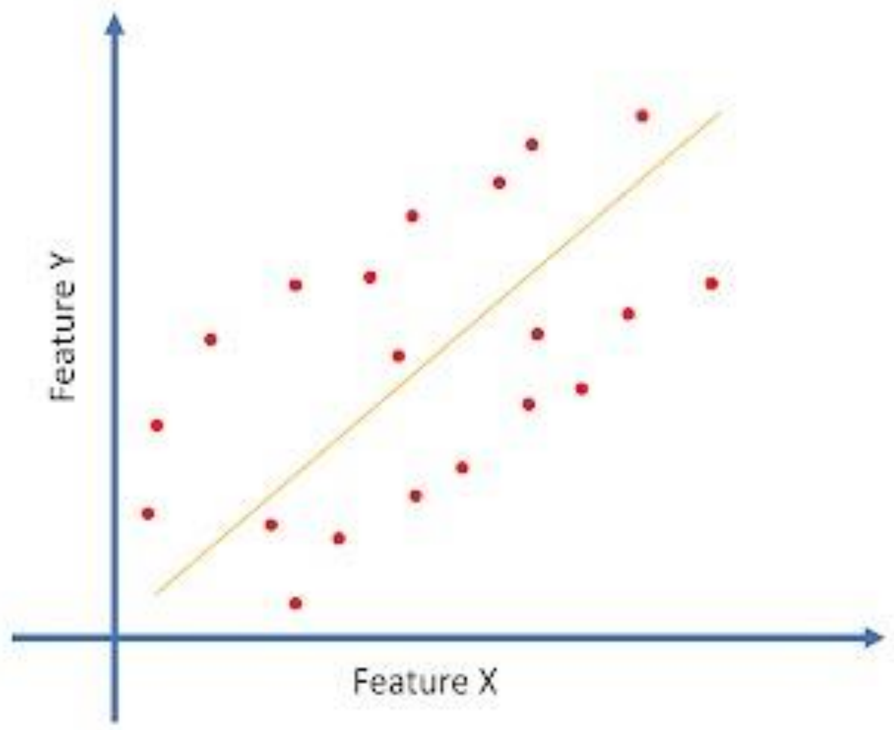
$$= \frac{26.74}{1.22 * 24.454}$$

$$= 0.896$$

A correlation coefficient of 0.896 indicates a strong positive relationship between height and weight. This indicates that as a person's height increases, their weight also tends to increase

Measures of Relationship

Covariance and correlation are statistical measures used to quantify the relationship between two variables.

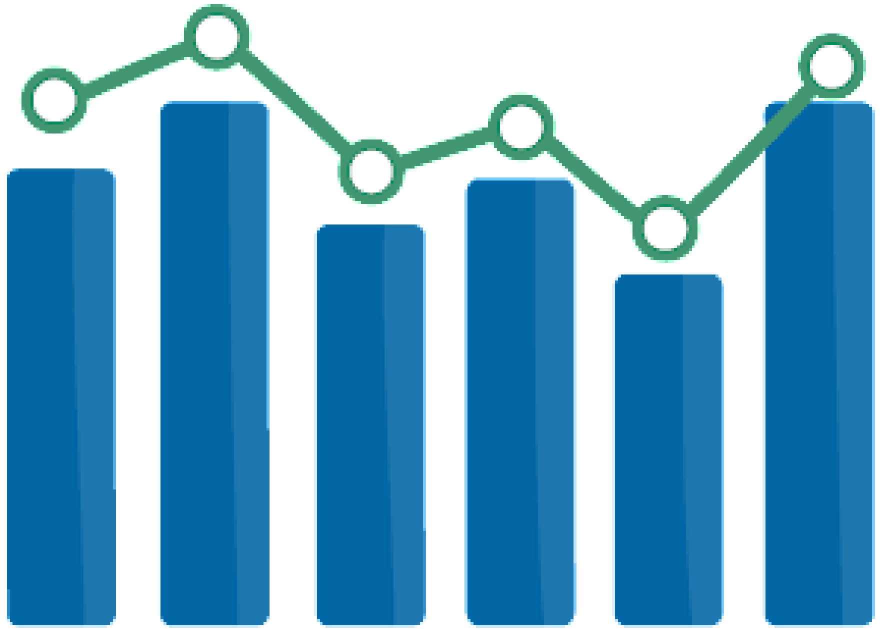


Although they are typically used with numerical values, they can also be applied to other types of data, including ordinal or interval data.

Expectation

Expectation

The expected value of a random variable X is a weighted average of the possible values that X can take.



Expectation: Example

If a coin is tossed 10 times, the outcome is likely to be heads 5 times and tails 5 times.

Introduction to Probability

Probability Theory

Probability is a quantitative measure that represents the likelihood of an event.

Example

- For example, in a coin toss scenario, the probability of getting heads is ½ or 50%.
- The probability of each given event is between 0 and 1 (inclusive).
- The probability of any given event falls within the range of 0 to 1 (inclusive).
- Therefore, the probability of an event, denoted as p(x), satisfies the condition $0 \leq p(x) \leq 1$.
- This means that $\int p(x)dx = 1$ (integral of p for a distribution over x).

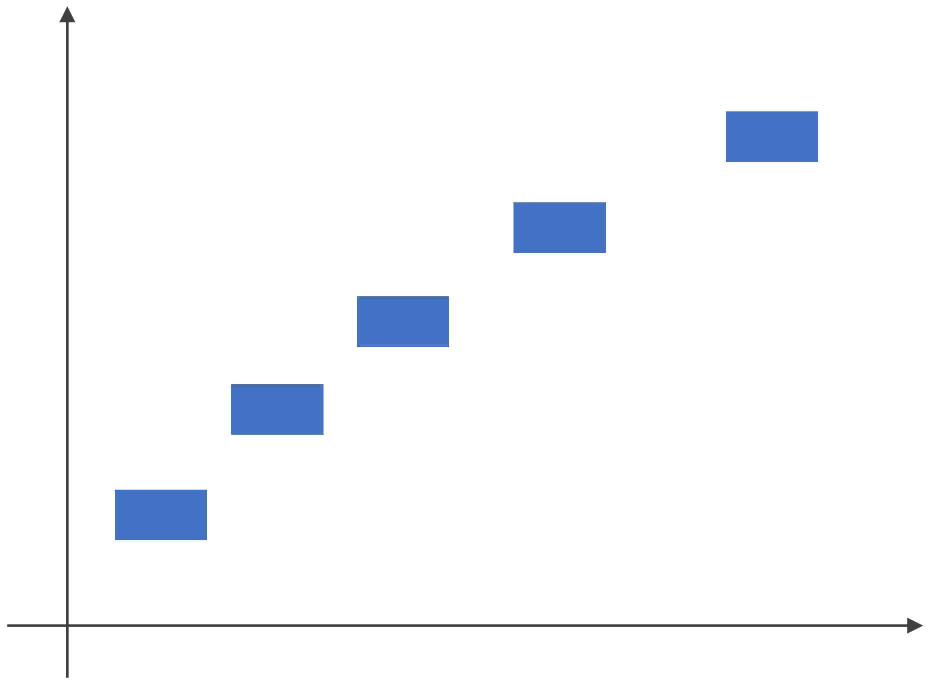
Types of Probability

In statistics, three main types of probability are commonly used:

1. Marginal probability
2. Conditional probability
3. Joint probability

Marginal Probability

Marginal probability refers to the probability of an event occurring without considering the occurrence or non-occurrence of any other event.



It focuses on the probabilities of individual variables in a multi-dimensional dataset, disregarding the influence of other variables.

Marginal Probability: Example

Consider a marginal probability of rolling an even number on a single die, regardless of the outcome of a second roll.

- The possible outcomes of rolling a single die are {1, 2, 3, 4, 5, 6}, each with an equal probability of 1/6.
- To calculate the marginal probability of rolling an even number, we add the probabilities of the individual even outcomes (2, 4, and 6).
- P(even) = P(2) + P(4) + P(6) = 1/6 + 1/6 + 1/6 = 3/6 = 1/2
- Therefore, the marginal probability of rolling an even number is 1/2, which is equivalent to 50%.

Conditional Probability

Conditional probability refers to the probability of an event occurring given that another event has already occurred.

It quantifies the likelihood of an event based on additional information or a specific condition.

P(A|B) denotes the probability of event A given event B has occurred.

The formula for conditional probability is, $P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)}$

Conditional Probability: Example

Consider a bag of colored marbles with 10 red, 8 blue, and 12 green marbles.

Define the events as follows:

A: Selecting a red marble
B: Selecting either a blue or green marble

To compute P(A|B), we must calculate the probabilities P(A and B) and P(B).

Conditional Probability: Example

When a marble is picked randomly from a bag, and the objective is to find the probability of choosing a red marble while it might be blue or green, conditional probability can be used effectively.

P(A and B) = P (selecting a red marble and selecting either a blue or green marble = P(selecting a red marble)

= 10 / (10 + 8 + 12) = 10 / 30 = 1/3

P(B) = P(selecting either a blue or green marble) = (8 + 12) / (10 + 8 + 12) = 20 / 30 = 2/3

Hence, conditional probability, $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$ = (1/3) / (2/3) = ½ or 50%

Bayesian Conditional Probability

Bayesian probability is a specific approach within the framework of conditional probability that incorporates prior knowledge and updates probabilities based on new evidence.

Thomas Bayes

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

Source: https://en.wikipedia.org/wiki/Thomas_Bayes

Bayesian Conditional Probability

The Bayes model specifies the probability of event A occurring if event B has already occurred.

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

Where, P(A) = Probability of event A
P(B) = Probability of event B
P(A|B) = Probability of A given B is true
P(B|A) = Probability of B given A is true
P(A ∩ B) = Probability of both events happening

Bayesian Conditional Probability: Example

Consider a two-coin flip experiment, where:

- P(Getting a head with the first coin) = P(coin 1-H) = 2/4
- P(Getting a head with the second coin) = P(coin 2-H) = 2/4
- P (coin2-H) = 2/4
- P (coin 1-H ⋂ coin2-H) = 1/4

Two coin flip	
Coin 1	Coin 2
H	T
T	H
H	H
T	T

The probability of coin 1-H, given coin 2-H, can be calculated as follows:

P(coin1-H | coin2-H) = (1/4)/(2/4) = ½ = 50%

Simplifying the Bayes Equation

Events A and B are statistically independent if

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

Where:

$P(A|B) = P(A)$, assuming $P(B)$ is not zero

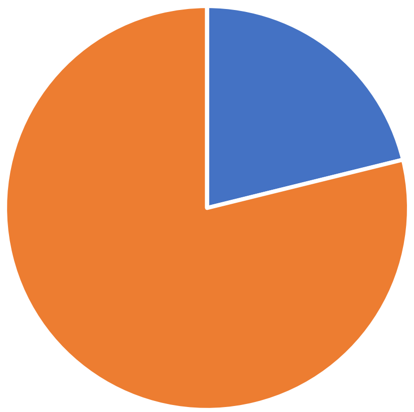
$P(B|A) = P(B)$, assuming $P(A)$ is not zero

Data Analytics with Bayes Model: Example

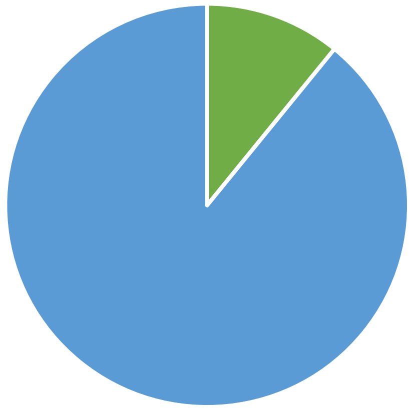
Consider an example to calculate the probability of developing diabetes based on the frequency of fast food:

Observed Data

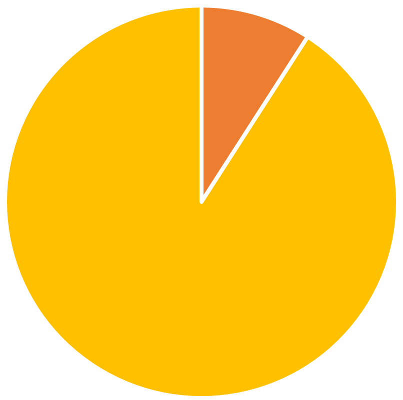
Fast food consumers = 20%



Diabetes prevalence = 10%



Diabetics that consume fast food= 5%



Chances of diabetes, given fast food consumption: (conditional probability)

⇒ (D and F)/F = 5%/20% = ¼ = 25%

Analysis: Fast food consumption increases the chances of diabetes by 25%.

Joint Probability

Joint probability denotes the likelihood of two or more events occurring simultaneously.

- It gauges the probability of an intersection or overlap between multiple events.
- For two events, A and B, we denote the joint probability as P(A and B) or P(A ∩ B).
- This notation represents the likelihood of both events A and B occurring together.

Joint Probability

The joint probability is calculated by multiplying the individual probabilities of each event.

If the events are independent, meaning that the occurrence of one event does not affect the probability of the other, the joint probability is simply the product of their probabilities.

→ P(A and B) = P(A) * P(B)

Joint Probability: Example

Consider a deck of playing cards. The probability of a red card (event A) and a face card (event B) being drawn from the deck is to be determined.

Joint Probability: Example

The joint probability can be calculated as follows:

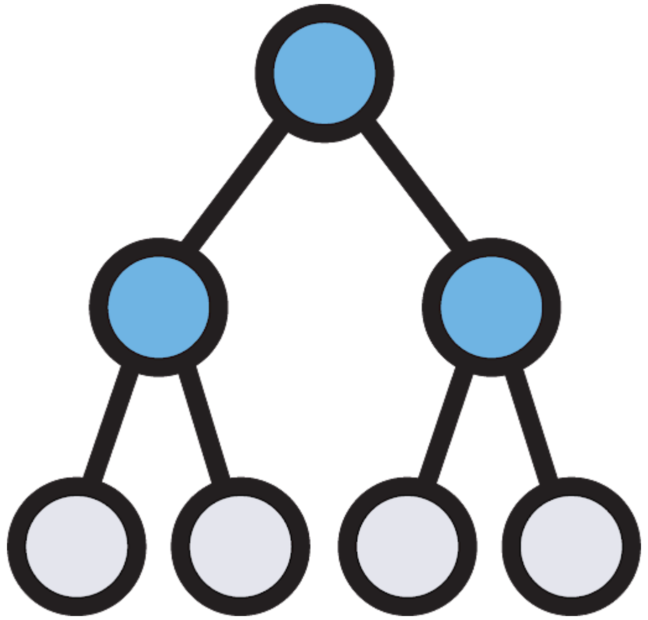
P(A) represents the probability of a red card being drawn. Since there are 26 red cards out of a total of 52 cards, P(A) = 26/52 = 1/2.

P(B) represents the probability of a face card being drawn. Since there are 12 face cards out of a total of 52 cards, P(B) = 12/52 = 3/13.

P(A and B) can be calculated by multiplying P(A) and P(B) = (1/2) * (3/13) = 3/26.

Chain Rule of Probability

The chain rule of probability, also recognized as the multiplication rule, is a fundamental concept in probability theory.



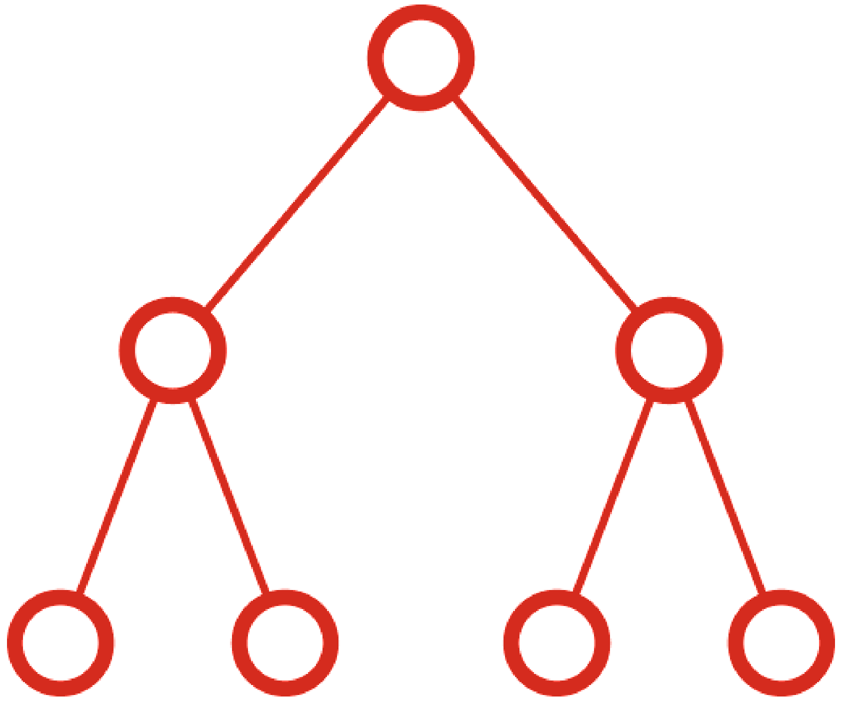
- It enables us to compute the probability of several events occurring simultaneously.
- Mathematically, the chain rule of probability can be expressed as follows:

P(A and B and C and ...) = P(A) * P(B|A) * P(C|A and B) * ...

In principle, the joint probability of events A, B, C, etc., is equal to the product of the initial and subsequent conditional probabilities.

Chain Rule of Probability

The chain rule of probability pertains to both conditional probability and joint probability.



This rule helps in calculating the likelihood of multiple events happening concurrently, whether they are joint events or events dependent on the occurrence of preceding events.

Chain Rule of Probability: Example

Assume three events A, B, and C. The objective is to determine the likelihood of all three events happening together, given that A and B have already occurred. In mathematical terms, P(C | A and B) is desired.

Using the chain rule of probability, this can be expressed as:

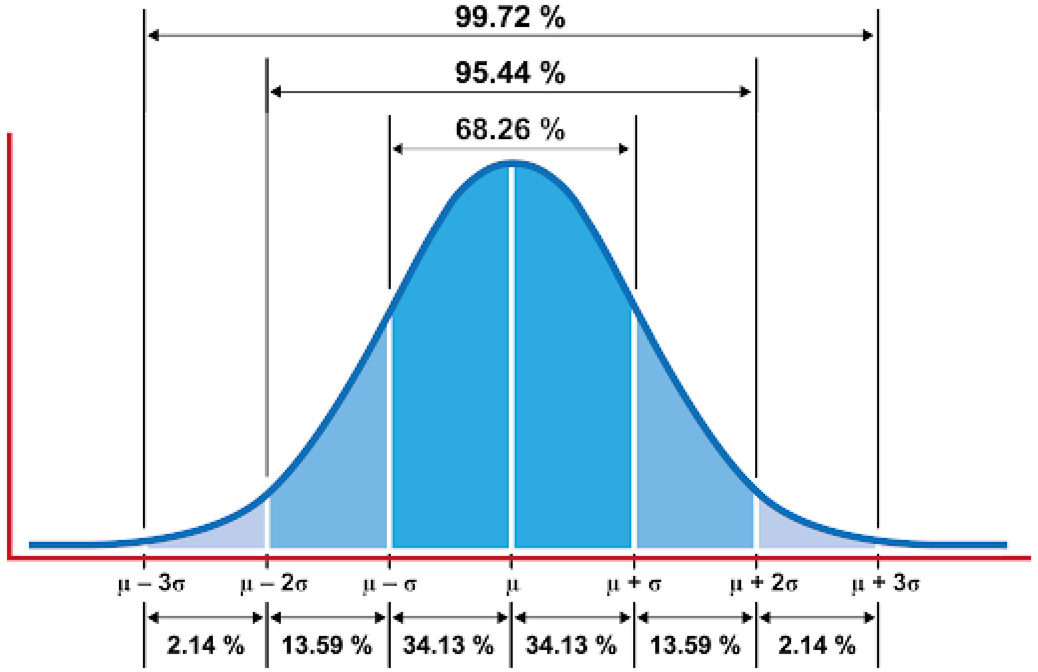
P(C|A and B) = P(C|A and B) * P(B|A) * P(A)

Here, P(C|A and B) denotes the likelihood of C given A and B. P(B|A) represents the likelihood of B given A. P(A) is the probability of event A.

Probability Distribution

Probability Distribution

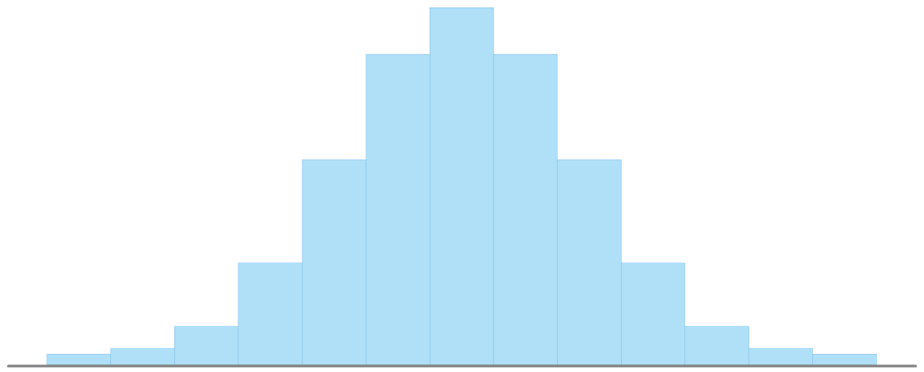
A probability distribution refers to a mathematical function or table.



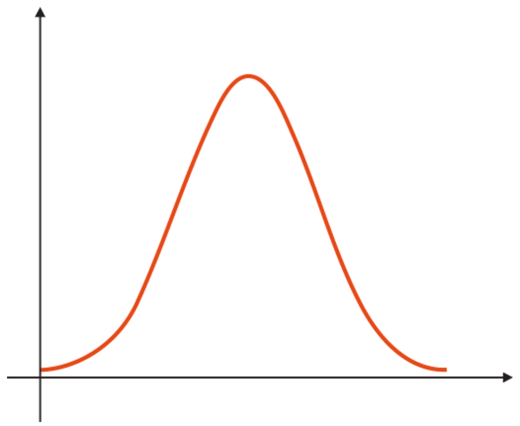
- It describes the likelihood of different outcomes or events in a random experiment or process.
- It provides a systematic method to assign probabilities to various possible outcomes.

Probability Distribution

The following are the two main types of probability distributions:



Discrete probability distribution

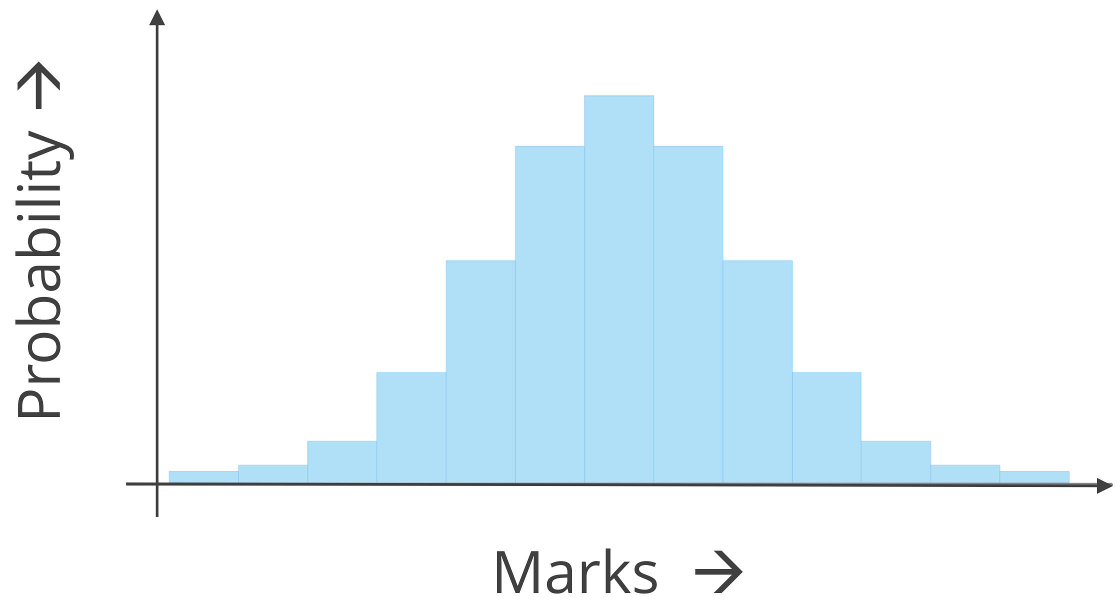


Continuous probability distribution

Discrete Probability Distribution

Discrete Probability Distribution

A discrete probability distribution describes the probabilities of a discrete random variable.

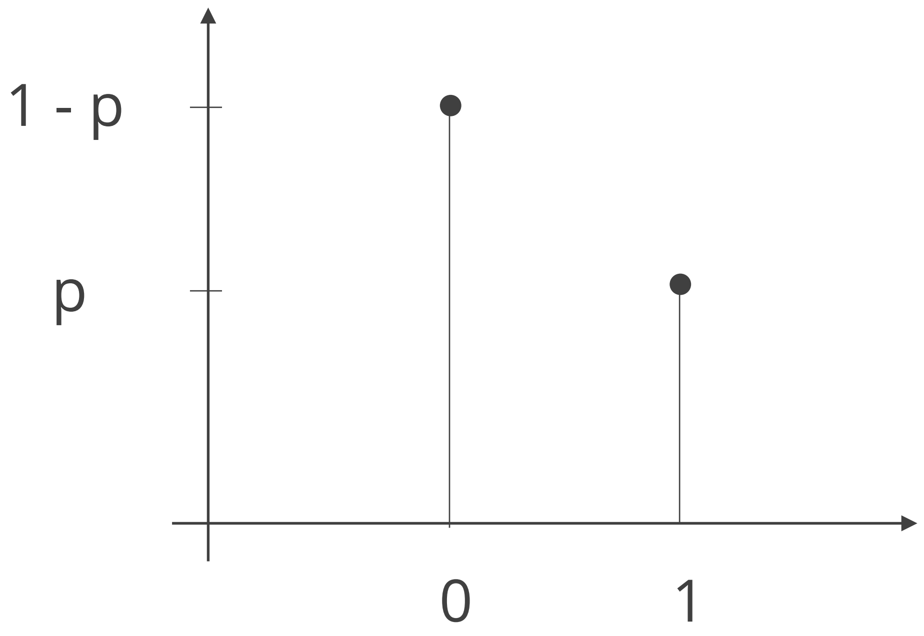


A discrete random variable can only take on a countable number of distinct values.

Examples of discrete probability distributions include the Bernoulli distribution, binomial distribution, Poisson distribution, and geometric distribution.

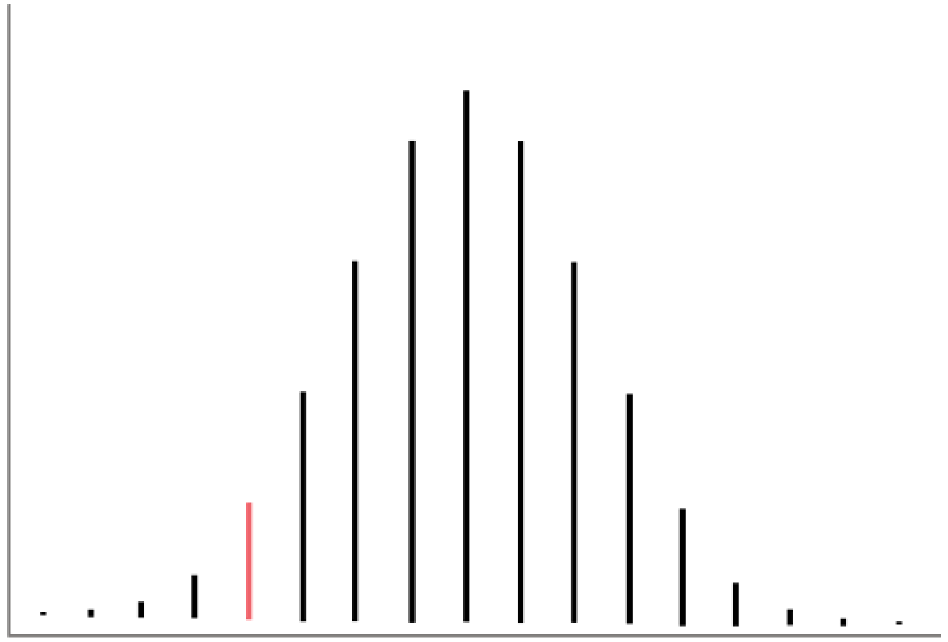
Discrete Probability Distributions

Bernoulli distribution



It describes the probability of a binary outcome, which includes success or failure, with a fixed probability of success.

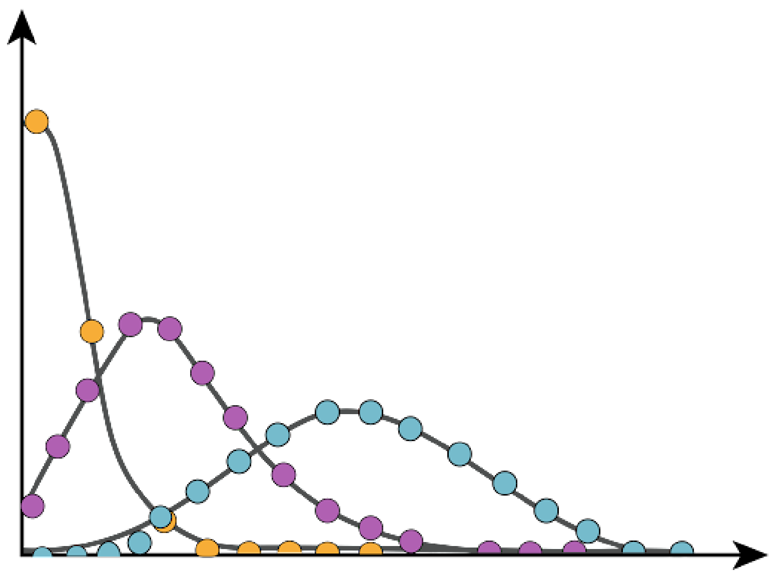
Binomial distribution



It describes the probability of achieving a specific number of successes in a fixed number of independent Bernoulli trials.

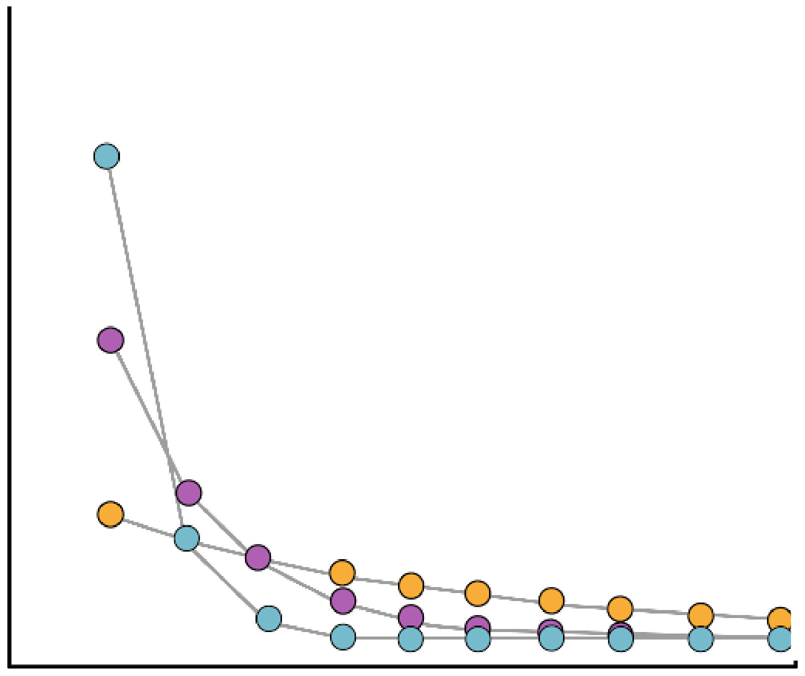
Discrete Probability Distributions

Poisson distribution



It describes the probability of a specific number of events occurring within a given timeframe or area, assuming there is a constant average rate.

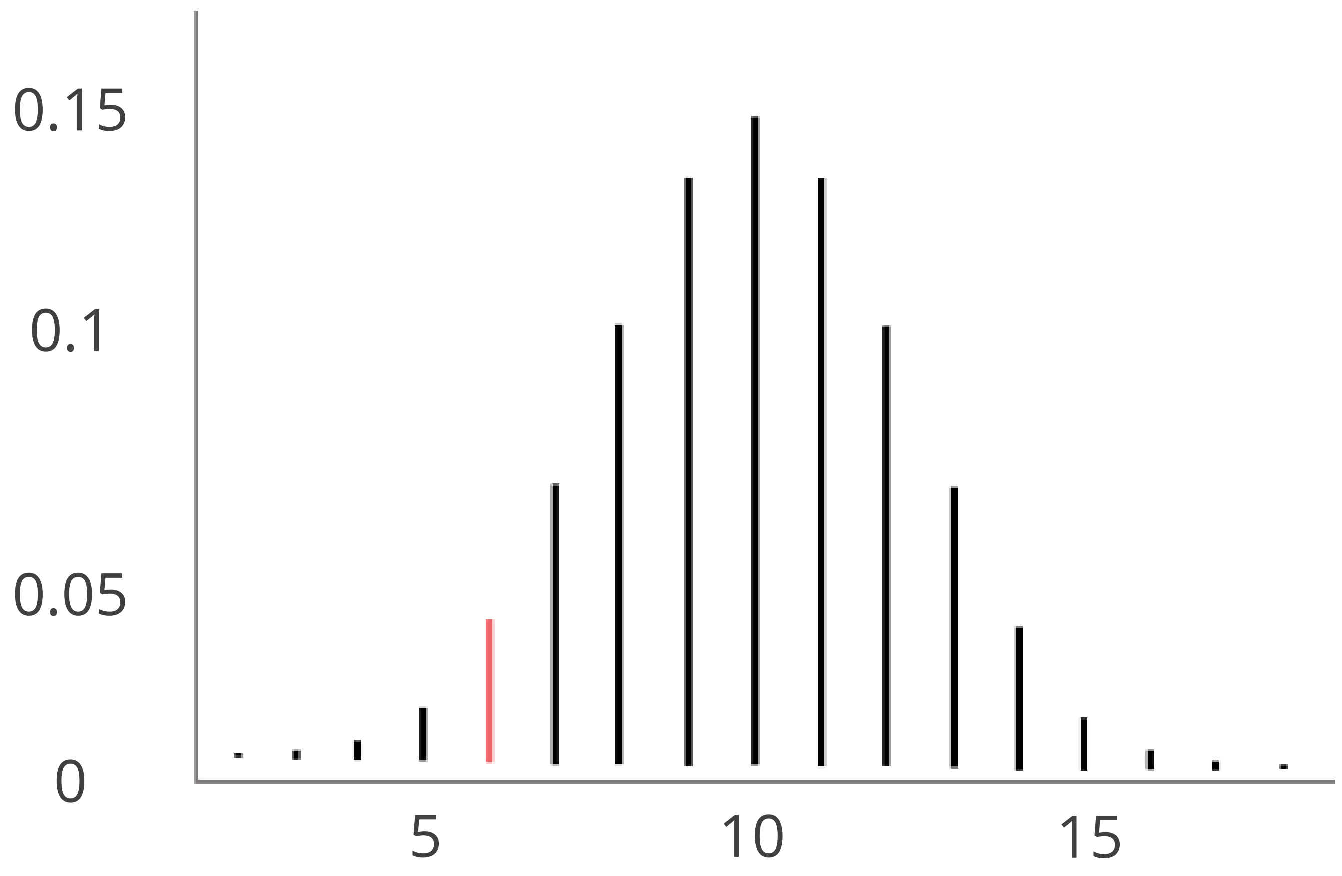
Geometric distribution



It describes the probability of the number of attempts required to achieve the first successful outcome in a series of independent Bernoulli trials.

Binomial Distribution

The binomial distribution is a discrete probability distribution that models the number of successful outcomes in a predetermined number of independent Bernoulli trials.



It is an extension of the Bernoulli distribution, which models a single binary outcome.

Binomial Distribution

The binomial distribution is characterized by two parameters: the number of trials (n) and the probability of success in each trial (p).

The probability mass function (PMF) of the binomial distribution is given by:

$$P_x = {}^nC_x \, p^x \, q^{(n-x)}$$

Where:

P: Binomial probability

x: Number of times for a specific outcome within n trials

${}^nC_x$: Number of combinations

P = Probability of success on a single trial

q = Probability of failure on a single trial

n = Number of trials

Properties of Binomial Distribution

Support: The binomial distribution is defined for non-negative integer values of k, ranging from 0 to n.

Mean: The mean (or expected value) of the binomial distribution is equal to, **E(X) = n * p.**

Variance: The variance of the binomial distribution is given by **Var = n * p * (1-p).**

Skewness: The skewness of the binomial distribution is determined by the values of n and p. Depending on the relationship between n and p, the distribution can be positively skewed, negatively skewed, or symmetrical.

Kurtosis: The kurtosis of the binomial distribution is affected by the values of n and p. It can be classified as leptokurtic (with a taller peak and heavier tails), mesokurtic (resembling a normal distribution), or platykurtic (with a flatter peak and lighter tails).

Applications of Binomial Distribution

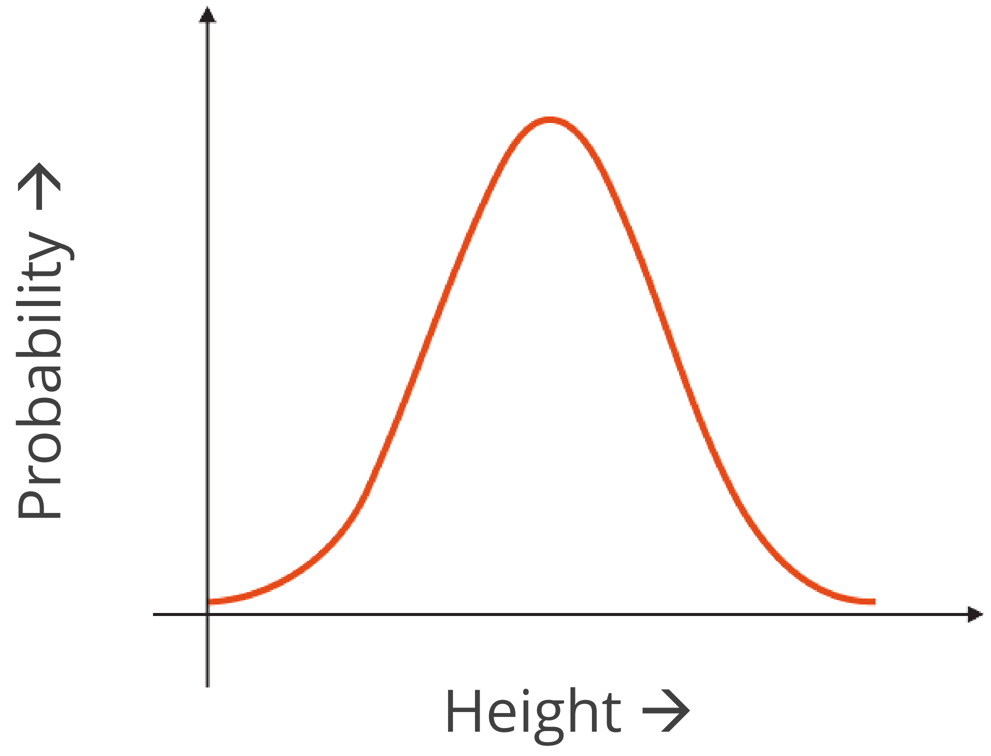
The binomial distribution finds diverse applications in statistics and practical situations, including:

- Modeling the number of successes or failures in a predetermined number of trials
- Estimating the probability of specific outcomes in games of chance
- Analyzing survey results that involve categorizing responses into two distinct categories
- Evaluating the performance of binary classification models
- Hypothesis testing and constructing confidence intervals for proportions

Continuous Probability Distribution

Continuous Probability Distribution

A continuous probability distribution describes the probabilities associated with a continuous random variable.



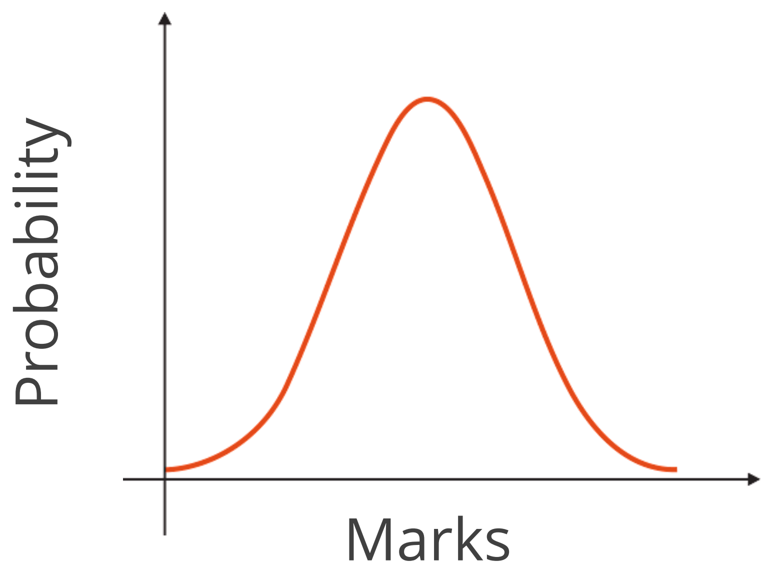
This type of random variable can take on any value within a specified range.

Examples of continuous probability distributions include the normal distribution, uniform distribution, exponential distribution, and gamma distribution.

Continuous Probability Distribution

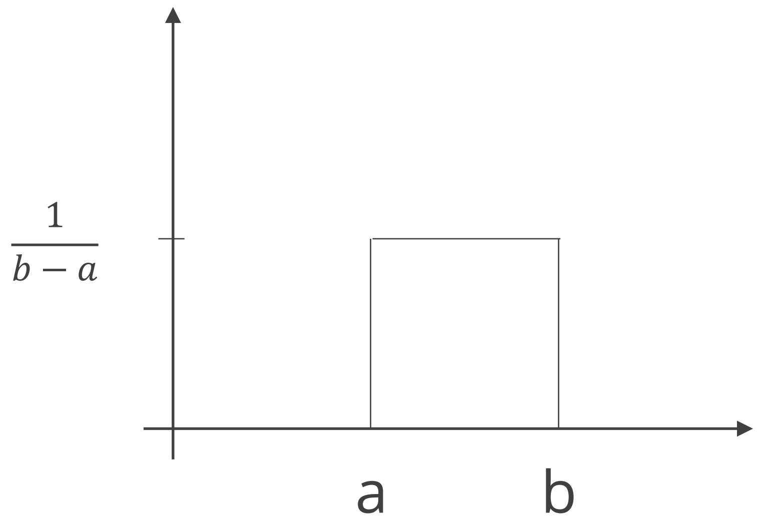
Types of continuous probability distributions:

Normal distribution



Describes a bell-shaped distribution that is symmetric and commonly observed in natural phenomena

Uniform distribution

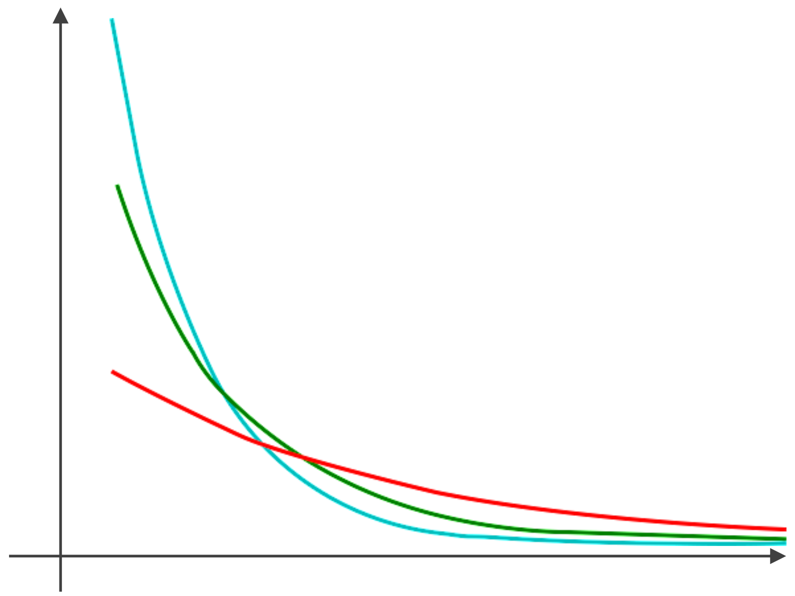


Describes a uniform distribution where all values within a range have equal likelihood

Continuous Probability Distribution

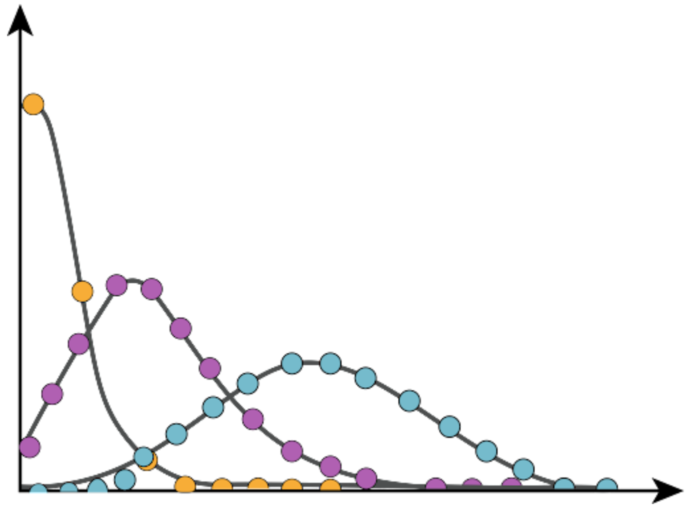
Types of continuous probability distributions:

Exponential distribution



Describes the probability of the time between events occurring in a Poisson process

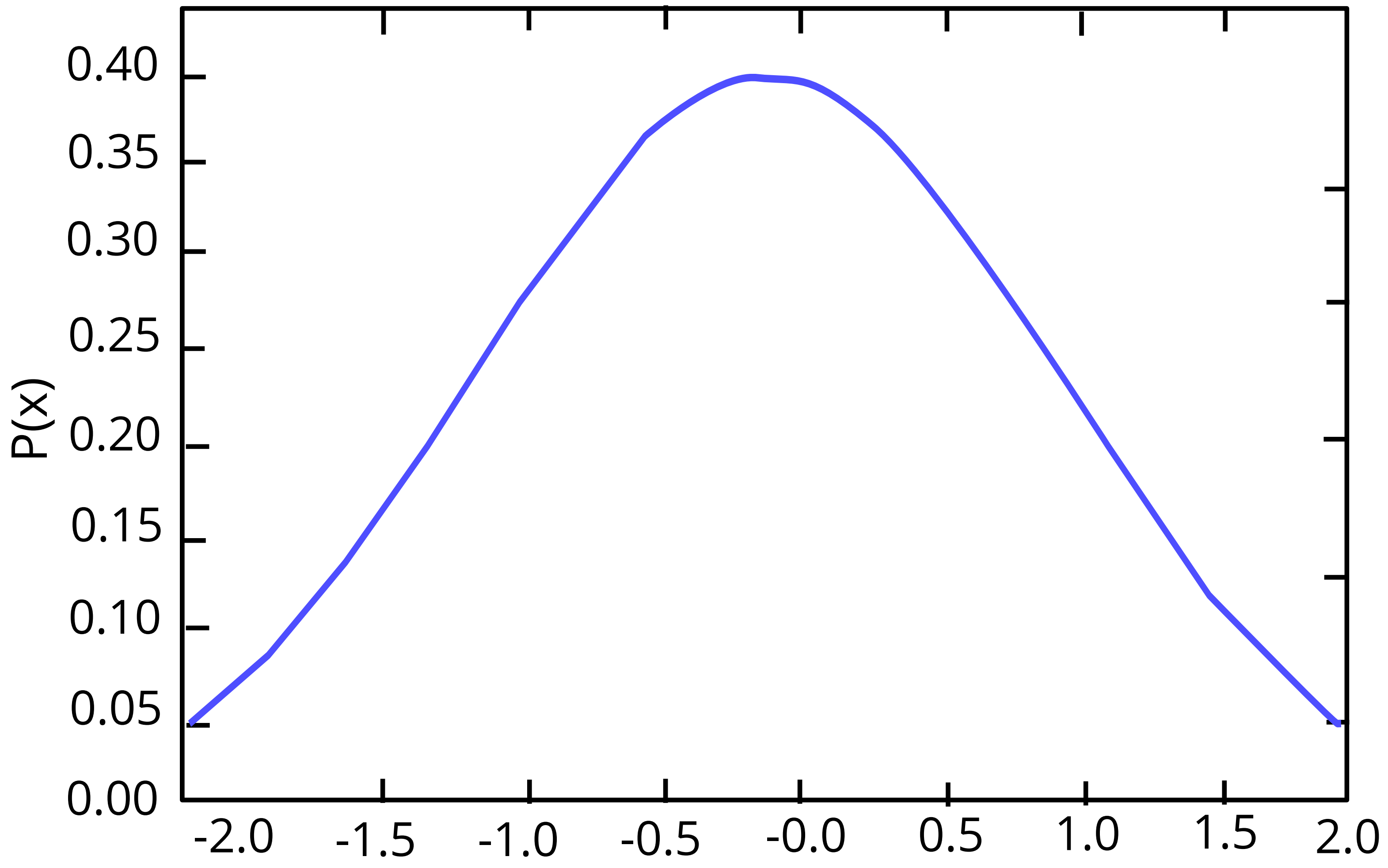
Gamma distribution



Describes the probability of the time it takes until a specified number of events occur in a Poisson process

Normal Distribution

It is a type of distribution where data tends to cluster around a central value without any significant bias to the left or right.



It is also known as the Gaussian distribution.

In machine learning, when there is a lack of prior information, the normal distribution is considered a reasonable assumption.

Properties of Normal Distribution

The normal distribution, also known as the Gaussian distribution, exhibits the following properties:

01 **Symmetry:** The normal distribution is symmetric with respect to its mean. This implies that the left and right tails of the distribution mirror each other.

02 **Bell-shaped curve**: The shape of the normal distribution closely resembles a bell curve. It is characterized by a peak at the mean and gradually decreasing values on both sides.

03 **Unimodal:** The normal distribution is unimodal, indicating a single peak.

Properties of Normal Distribution

The normal distribution, also known as the Gaussian distribution, exhibits the following properties:

04 **Mean and median equality**: In a normal distribution, the mean, median, and mode are all equal and located at the center of the distribution.

05 **Standard deviation and variance**: The spread of a normal distribution is determined by the standard deviation (σ) and variance (σ^2). A larger standard deviation indicates a wider spread of data points.

06 **Empirical rule**: The empirical rule, also known as the 68-95-99.7 rule, applies to normal distribution. It states that approximately 68% of the data falls within one standard deviation of the mean, approximately 95% falls within two standard deviations, and approximately 99.7% falls within three standard deviations.

Properties of Normal Distribution

The normal distribution, also known as the Gaussian distribution, exhibits the following properties:

07 **Transformations**: Normal distributions maintain their normality under linear transformations, such as adding or multiplying by constants.

08 **Characterized by two parameters**: The normal distribution is characterized by its mean (μ) and standard deviation (σ). These parameters determine the distribution's location and spread.

09 **Related to the standard normal distribution**: The standard normal distribution is a specific instance of the normal distribution with a mean of 0 and a standard deviation of 1. Other normal distributions can be standardized to the standard normal distribution using a process called standardization.

Normal Distribution: Equation

Formulas for calculating normal distribution are given below:

$$N(x;\mu,\sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2}}\,\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

$$N(x;\mu,\beta^{-1}) = \sqrt{\frac{\beta}{2\pi}}\,\exp\left(-\frac{1}{2}\beta(x-\mu)^2\right)$$

Here:

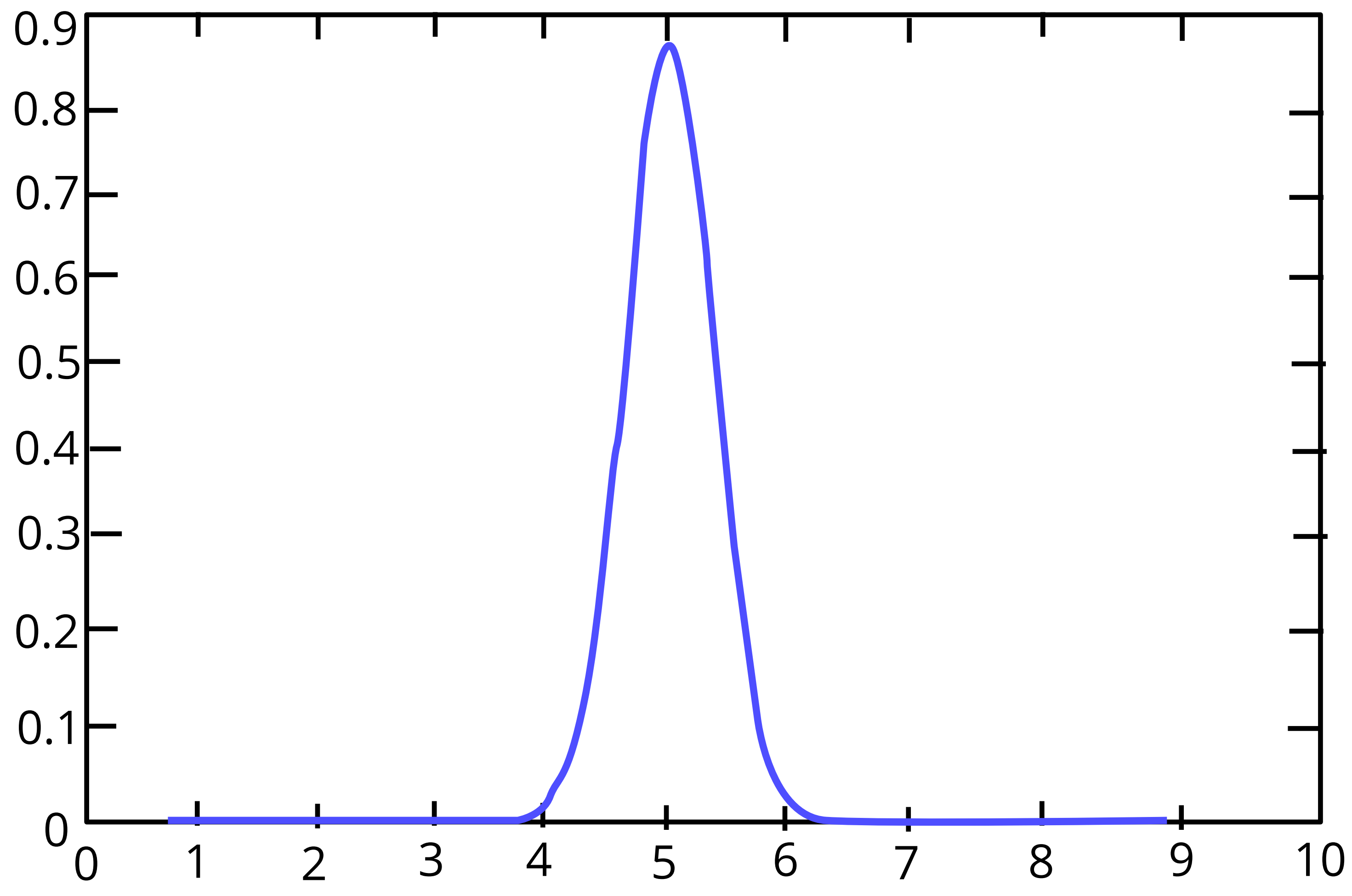
- μ = mean or peak value, which also means E[x] = μ
- σ = standard deviation, and σ^2 = variance

Note:

- A standard normal distribution has μ = 0 and σ = 1
- For efficient handling, invert σ and use precision β (inverse variance) instead

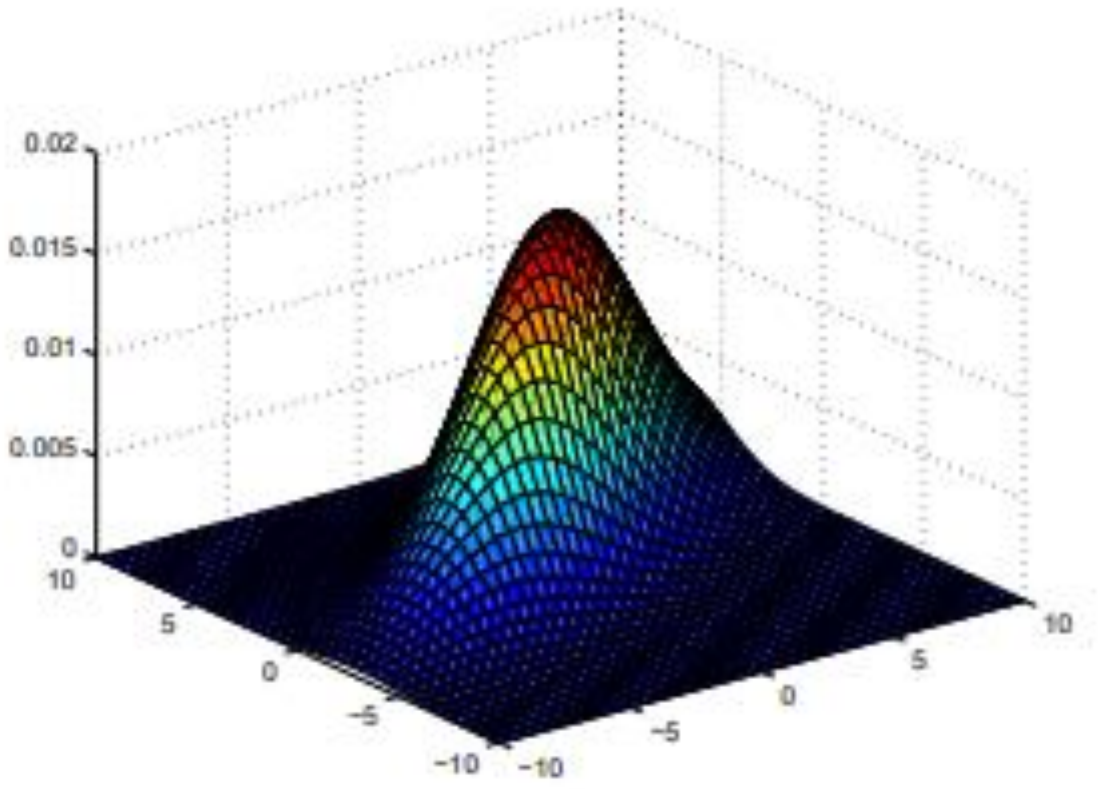
Types of Normal Distribution: Univariate

The distribution of a single variable is known as a univariate normal distribution.



Types of Normal Distribution: Multivariate

The multivariate normal distribution is an extension of the univariate normal distribution to several variables.



The multivariate normal distribution is a distribution over two variables, x1 and x2.

Source: https://cs229.stanford.edu/section/gaussians.pdf

Applications of Normal Distribution

The following are the applications of normal distribution:

Central limit theorem	It states that the normal distribution applies to the sum or average of many independent random variables.
Data modeling and analysis	It involves using the normal distribution to model continuous variables such as heights, weights, and errors in measurements.
Estimation and inference	It uses the normal distribution to estimate population parameters and construct confidence intervals.
Hypothesis testing	It relies on the normal distribution to evaluate the significance of observed data.

Applications of Normal Distribution

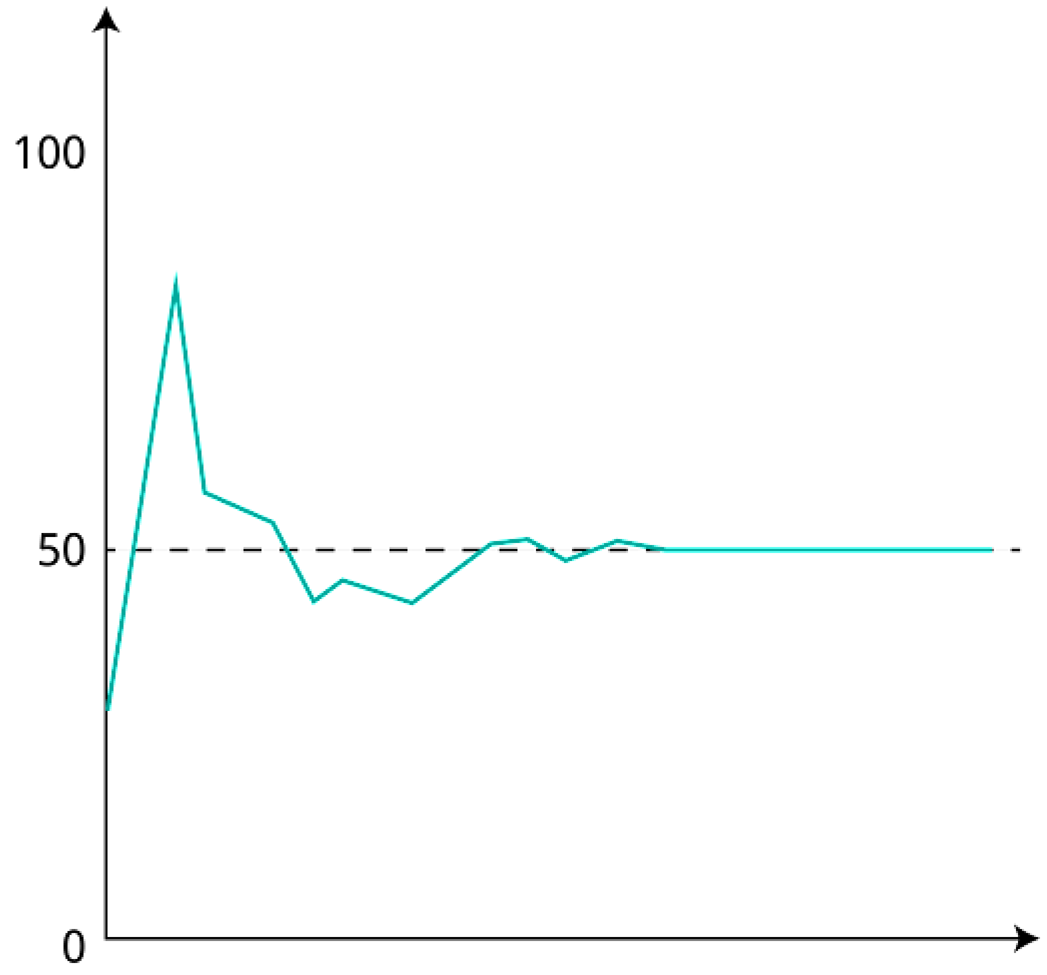
The following are the applications of normal distribution:

Process control	It uses the normal distribution for monitoring and controlling industrial processes.
Risk management and finance	It involves using the normal distribution to model asset returns and implement risk management strategies.
Quality control	It employs the normal distribution to assess the variability of product characteristics and establish tolerance limits.
Simulation and **Monte Carlo** methods	It uses the normal distribution in simulations and Monte Carlo methods to evaluate probabilities.

Law of Large Numbers

Law of Large Numbers

The law of large numbers is a theorem that describes the result of performing the same experiment numerous times.



Source: *https://bluebox.creighton.edu/demo/modules/en-boundless-old/www.boundless.com/definition/law-of-large-numbers/index.html*

Law of Large Numbers

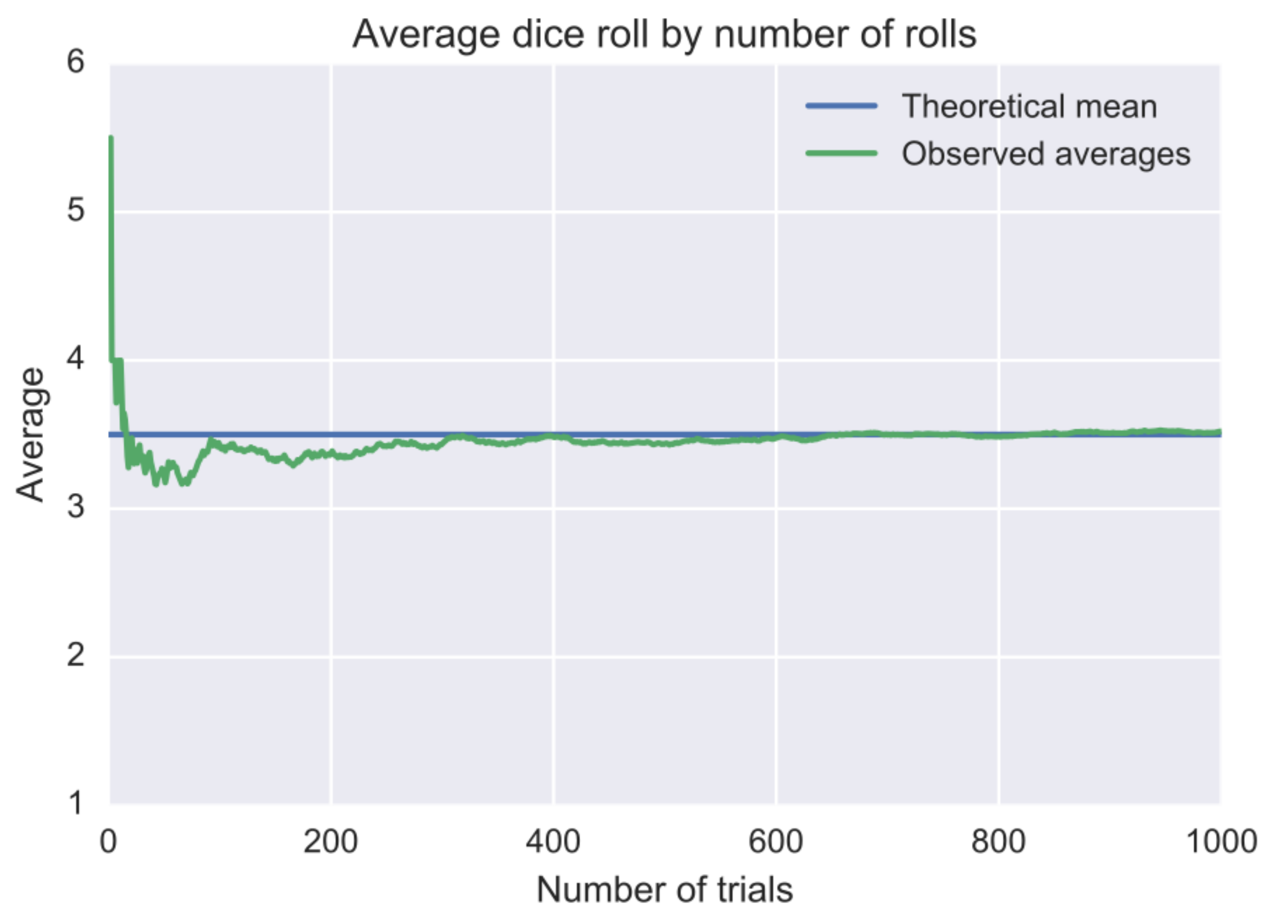
Tossing a coin numerous times gives the following outcomes:

- The results split evenly between heads and tails
- Expected average value is ½ (50%)

Note: Tossing the coin 1000 times may result in an even split between heads and tails, but this may not be the case if it is only tossed 10 times.

Law of Large Numbers

The graph indicates that as the number of rolls increases, the average value of the results approaches 3.5.



Source: Wikipedia

Key Takeaways

- Probability and statistics form the foundation of data analysis.
- Data helps predict future outcomes and make assessments based on past information patterns.
- Central tendency refers to a single value that describes the data by identifying its central position. The mean, median, and mode are measures of central tendency.
- Gaussian distribution is a type of distribution in which the data tends to cluster around a central value with a minimal bias towards the left or right.

Knowledge Check

Knowledge Check 1

If x_1, x_2, x_3,....., x_n are the observations of a given data, then the mean of the observations will be:

A. Sum of observations/Total number of observations

B. Total number of observations/Sum of observations

C. Sum of observations + Total number of observations

D. Sum of observations + Total number of observations/2

Knowledge Check 1

If x_1, x_2, x_3,….., x_n are the observations of a given data, then the mean of the observations will be:

A. Sum of observations/Total number of observations

B. Total number of observations/Sum of observations

C. Sum of observations + Total number of observations

D. Sum of observations + Total number of observations/2

The correct answer is **A**

Mean = Sum of observations/Total number of observations. In the example, mean = $(x_1 + x_2 + x_3 + \ldots + x_n)/n$

Knowledge Check 2

Which of the following can be the probability of an event?

A. - 0.4

B. 1.004

C. 18/23

D. 10/7

Knowledge Check

2

Which of the following can be the probability of an event?

A. - 0.4

B. 1.004

C. 18/23

D. 10/7

The correct answer is **C**

The probability of an event is always between 0 and 1. In the given options, only 18/23 falls between 0 and 1.

Knowledge Check

3

Which of the following is true about the normal distribution?

A. It is skewed to the right

B. It is a discrete probability distribution

C. Its mean, median, and mode are equal

D. It has a uniform shape

Knowledge Check

3

Which of the following is true about the normal distribution?

A. It is skewed to the right

B. It is a discrete probability distribution

C. Its mean, median, and mode are equal

D. It has a uniform shape

The correct answer is **C**

The mean, median, and mode are all equal for a normal distribution.

Knowledge Check

4

A binomial distribution is characterized by:

A. Continuous outcomes

B. Number of trials and probability of success

C. Multiple peaks in the distribution

D. Mean equal to the probability of success

Knowledge Check

4

A binomial distribution is characterized by:

A. Continuous outcomes

B. Number of trials and probability of success

C. Multiple peaks in the distribution

D. Mean equal to the probability of success

The correct answer is **B**

A binomial distribution is characterized by two parameters, the number of trials and the probability of success.

Knowledge Check

5

Conditional probability is defined as:

A. The probability of two independent events occurring together

B. The probability of an event occurring given that another event has already occurred

C. The probability of an event occurring in a single trial

D. The probability of an event occurring in a series of trials

Knowledge Check

5

Conditional probability is defined as:

A. The probability of two independent events occurring together

B. The probability of an event occurring given that another event has already occurred

C. The probability of an event occurring in a single trial

D. The probability of an event occurring in a series of trials

The correct answer is **B**

Conditional probability is defined as the probability of an event occurring, given that another event has already occurred.