Learning Outcome

After completing this module, a student will be able to learn:

* Probability Theories
* Bayes’ Theorem, Maximum Likelihood
* Hypothesis Testing
* Central limit theorem
* Chi-square test

## 

## 

## 

# Probability Theories

## What Is Probability Theory?

Probability theory is a branch of mathematics focusing on the analysis of random phenomena. It is an important skill for data scientists using data affected by chance.

With randomness existing everywhere, the use of probability theory allows for the analysis of chance events. The aim is to determine the likelihood of an event occurring, often using a numerical scale of between 0 and 1, with the number “0” indicating impossibility and “1” indicating certainty.

A classic example of this is a coin toss, where there can be two possible options: heads or tails. Here the possibility of flipping a head or a tail on a single toss is 50%. When conducting your own experiment, you may find that the outcomes can vary. But if you continue flipping the coin, the outcome grows closer to 50/50.

Probability plays a vital role in many areas of scientific research. Researchers can integrate uncertainty into their research models as a way of describing their findings. This allows for a predictive distribution of findings tied to what may have been observed in the past.

Randomness and uncertainty are popular themes tied to probability. In Nassim Taleb’s bestselling books The Black Swan and Fooled By Randomness, the claim is made that rare events typically hold more importance than common ones because their effect size is not as restricted. Also, because of their rarity, results are unlikely to be determined.

Taleb popularized what he calls a “black swan” event, one that is rare, has a catastrophic impact when it does occur and can be explained in hindsight in a way that leads many to believe that it was actually predictable.

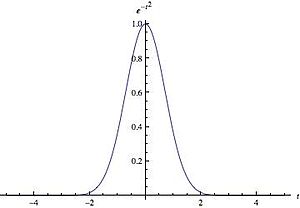


Image 1: Probability Theory

Reference: <https://upload.wikimedia.org/wikipedia/commons/thumb/d/d2/Gaussian_distribution_2.jpg/300px-Gaussian_distribution_2.jpg>

## Practical Uses for Probability Theory

Probability is commonly used by data scientists to model situations where experiments, conducted during similar circumstances, yield different results (as in the case of throwing dice or a coin).

It also has many practical uses in the business world. Take for example the insurance industry, where actuarial records chart life expectancy of individuals of a certain age. Instead of predicting what will happen to any one individual, the aim is to capture a collective result encompassing a large number of people.

Similar approaches have been taken in genetic science, where assessing the likelihood of a genetic disease is tied to frequency of occurrence as opposed to predictions about a specific individual.

Another common application of probability is also commonly applied in clinical trials where new disease treatments, drugs or surgical treatments are being sought. In assessing whether a treatment can be deemed a success or failure, the clinical trial aims to determine whether the new treatment is more successful than a prevailing treatment standard.

An example here is testing the efficacy of a new vaccine, such as the poliomyelitis testing done for the Salk vaccine in 1954 involving almost two million children. Organized by the U.S. Public Health Service, the vaccine nearly eliminated polio as a health problem in the industrialized world.

## What Are the 3 Types of Probability?

There are three types of probability commonly used to gather statistical inference data. These are:

1. Classical
2. Relative Frequency
3. Subjective Probability

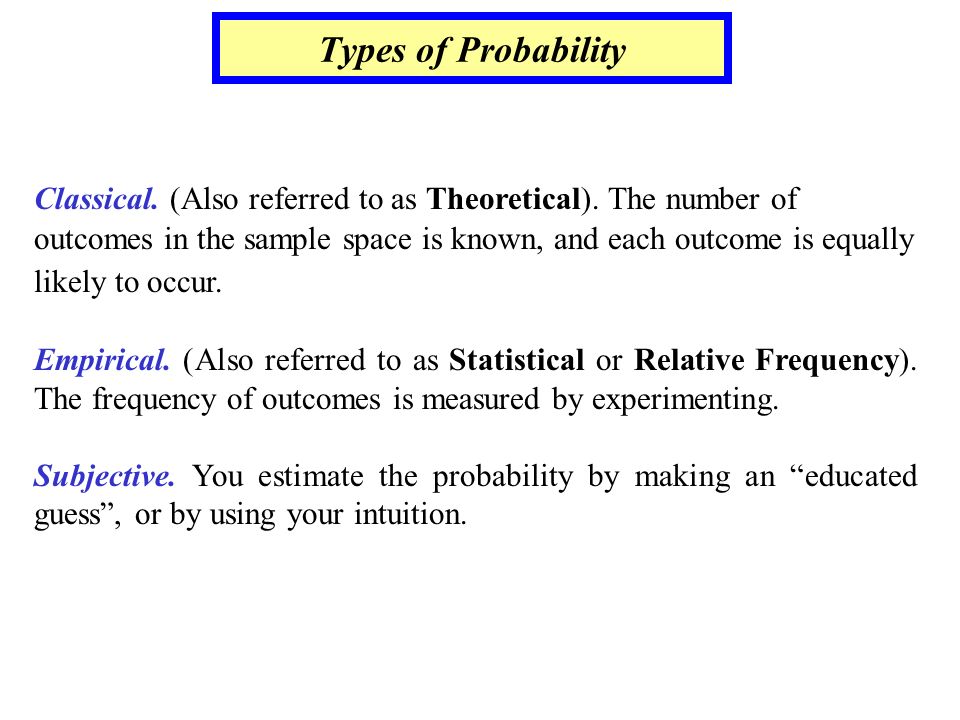


Image : Types Of Probability

Reference: <https://slideplayer.com/slide/6379961/22/images/3/Types+of+Probability.jpg>

1. **Classical**

Also known as the axiomatic method, this type of probability involves a set of axioms (rules) attached to it. For example, you could have a rule that the probability must be greater than 0.5% in order for it to be valid.

1. **Relative Frequency**

This involves looking at the occurrence ratio of a singular event in comparison to the total number of outcomes. This type of probability is often used after data from an experiment has been gathered to compare a subset of data to the total amount of collected data.

1. **Subjective Probability**

When using the subjective approach, probability is the likelihood of something happening based on one’s experiences or personal judgment. Here there are no formal calculations for subjective probability for it is based on one’s beliefs, judgment and personal reasoning.

By way of example, during a sporting event, fans of one team share who they are rooting for. This is based on facts or opinions they personally hold regarding the game, the two teams playing and the odds of the team winning.

## Probability Theory Examples

Probability theory is a tool employed by researchers, businesses, investment analysts and countless others for risk management and scenario analysis.

**Epidemiology**

Take epidemiology, which is the science of disease distribution. Researchers in this field study disease frequency, assessing how the probability differs across groups of people. A present-day example of this is the use of probability by epidemiologists to assess the cause-effect relationship between exposure and illness to the coronavirus.

Probability theory is often used to unlock key factors denoting the relationship between exposures and health risks. The aim here is to quantify uncertainty. This knowledge can fuel a course of action based on best outcomes for those affected by various diseases.

**Insurance**

The actuaries who are often employed in the insurance industry make primary use of probability, statistics and other data science tools to calculate the probability of uncertain future events occurring over a period of time. They then apply other data concepts to determine the amount of money that needs to be set aside to pay for future losses.

**Small Business**

Then there’s the small-business world where owners cannot always turn to their hunches and instincts to run a successful company. In today’s competitive business environment, probability analysis can provide entrepreneurs with key metrics pointing the way to the most profitable and productive paths. This analysis offers a controlled way to anticipate potential results.

For example, if a business enterprise expects to receive between $500,000 and $750,000 in revenue each month, the graph will begin with $500,000 at the low end and $750,000 at the high end. For a typical probability distribution, the graph will resemble a bell curve, where the least likely outcomes fall nearer the extreme ends of the range and the most likely nearer to the midpoint of the extremes.

**Meteorology**

A weather forecast serves as another example of probability theory. The probability for precipitation or severe weather is tied to a specific geographic location. As a result, forecasting can be viewed as the combination of the chance of a weather occurrence and the coverage of that event. According to an information statement of the American Meteorological Society:

“A probability forecast includes a numerical expression of uncertainty about the quantity or event being forecast. Ideally, all elements (temperature, wind, precipitation, etc.) of a weather forecast would include information that accurately quantifies the inherent uncertainty. Surveys have consistently indicated that users desire information about uncertainty or confidence of weather forecasts. The widespread dissemination and effective communication of forecast uncertainty information is likely to yield substantial economic and social benefits, because users can make decisions that explicitly account for this uncertainty.”

## Advantages and Disadvantages of Probability Theory

For data scientists, there are a number of advantages and disadvantages with probability that need to be considered.

1. **Classical**

The classical method of probability is used when all probable outcomes have an equal likelihood of happening and every outcome is known in advance. The coin toss example above uses the classical approach to probability. The classical approach offers a simple approach to real-world examples that is easy to digest for those not possessing a math or science background.

With respect to limitations, the classical approach is unable to handle projects where an infinite number of possible outcomes exist. It’s also ineffective in scenarios where each outcome is not equally likely, as in the case of tossing a weighted die. These limitations affect the ability of this approach to handling more complicated tasks.

1. **Relative Frequency**

Unlike the classical approach, relative frequency offers the advantage of being able to handle scenarios where outcomes have different theoretical probability (or likelihood) of occurring. This approach can also manage a probability situation where possible outcomes are unknown.

Although you can use relative frequency probability in more diverse situations and settings than classical probability, it has several limitations. The first limitation to relative frequency involves the problem of “infinite repetitions.” This is where experiments possessing an infinite number of times cannot be analyzed with this theory. So while a large number of trials can be conducted, that number can’t be infinite.

1. **Subjective**

Problems that benefit from subjective probability are those that require some level of belief to make possible. For example, a candidate who may be down in the polls may use subjective probability to make a case for staying in the race.

Subjective probability also benefits from what is known as the reference class problem. In a reference class problem, assigning a probability to a certain event might require that event to be classified. That classification can be subjective, and thus changing the classification can change the probability of the event.

For example, if you want to determine the probability of a person contracting an infectious disease like COVID-19, we need to begin with assessing which classes of people are relevant to the problem. It’s here where various reference classes can be established. A broad class such as “all U.S. residents” could be used. Or it could be narrowed down to, say, “all residents of the states of X, Y and Z, where 80% of the deaths are occurring.” In other words, depending on the reference class chosen, different probabilities will emerge.

## What is the probability formula?

The formula for probability states that the possibility of an event happening, or P(E), equals the ratio of the number of favorable outcomes to the number of total outcomes. Mathematically, it looks like this:

P(E) = favorable outcomes/total outcomes

## How Data Scientists Use Probability Theory

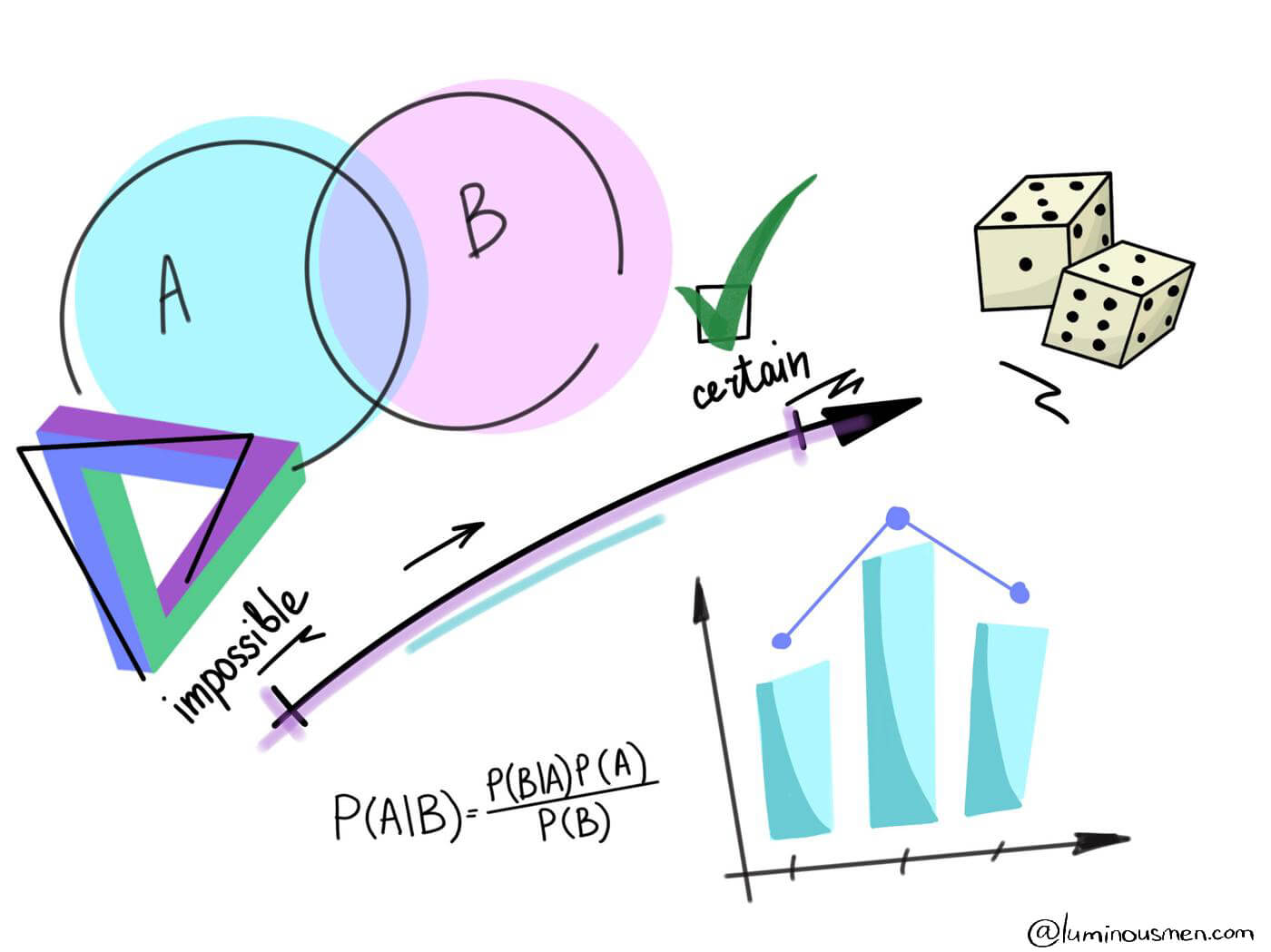


Image: Data Science. Probability

References: <https://luminousmen.com/media/data-science-probability_2.jpg>

Probability allows data scientists to assess the certainty of outcomes of a particular study or experiment. An experiment is a planned study that is executed under controlled conditions. When a result is not already predetermined, the experiment is referred to as a chance experiment. Conducting a coin toss twice is an example of a chance experiment.

Today’s data scientists need to have an understanding of the foundational concepts of probability theory including key concepts involving probability distribution, statistical significance, hypothesis testing and regression. Learn more statistics concepts that data scientists use regularly; probability distribution is only one of them.

# Bayes’ Theorem, Maximum Likelihood

## Bayes’ Theorem

### **What is the Bayes’ Theorem?**

In statistics and probability theory, the Bayes’ theorem (also known as the Bayes’ rule) is a mathematical formula used to determine the conditional probability of events. Essentially, the Bayes’ theorem describes the probability of an event based on prior knowledge of the conditions that might be relevant to the event.

The theorem is named after English statistician, Thomas Bayes, who discovered the formula in 1763. It is considered the foundation of the special statistical inference approach called the Bayes’ inference.

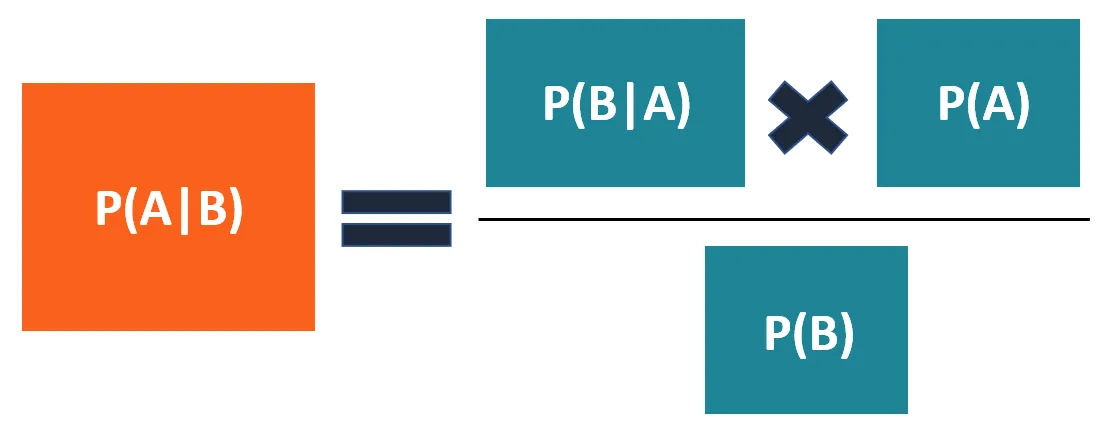


Image: Bayes' Theorem

Reference: <https://cdn.corporatefinanceinstitute.com/assets/bayes-theorem.png>

Besides statistics, the Bayes’ theorem is also used in various disciplines, with medicine and pharmacology as the most notable examples. In addition, the theorem is commonly employed in different fields of finance. Some of the applications include but are not limited to, modeling the risk of lending money to borrowers or forecasting the probability of the success of an investment.

### **Formula for Bayes’ Theorem**

The Bayes’ theorem is expressed in the following formula:

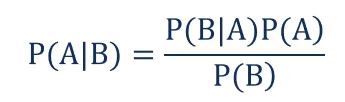


Image: Bayes’ Theorem – Formula

Reference: <https://cdn.corporatefinanceinstitute.com/assets/bayes-theorem1.png>

Where:

* P(A|B) – the probability of event A occurring, given event B has occurred
* P(B|A) – the probability of event B occurring, given event A has occurred
* P(A) – the probability of event A
* P(B) – the probability of event B

Note that events A and B are independent events (i.e., the probability of the outcome of event A does not depend on the probability of the outcome of event B).

A special case of the Bayes’ theorem is when event A is a binary variable. In such a case, the theorem is expressed in the following way:

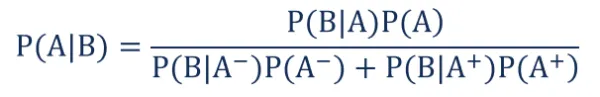


Image: Special Case

Reference: <https://cdn.corporatefinanceinstitute.com/assets/bayes-theorem2-600x97.png>

Where:

* P(B|A–) – the probability of event B occurring given that event A– has occurred
* P(B|A+) – the probability of event B occurring given that event A+ has occurred

In the special case above, events A– and A+ are mutually exclusive outcomes of event A.

Example of Bayes’ Theorem

Imagine you are a financial analyst at an investment bank. According to your research of publicly-traded companies, 60% of the companies that increased their share price by more than 5% in the last three years replaced their CEOs during the period.

At the same time, only 35% of the companies that did not increase their share price by more than 5% in the same period replaced their CEOs. Knowing that the probability that the stock prices grow by more than 5% is 4%, find the probability that the shares of a company that fires its CEO will increase by more than 5%.

Before finding the probabilities, you must first define the notation of the probabilities.

* P(A) – the probability that the stock price increases by 5%
* P(B) – the probability that the CEO is replaced
* P(A|B) – the probability of the stock price increases by 5% given that the CEO has been replaced
* P(B|A) – the probability of the CEO replacement given the stock price has increased by 5%.

Using the Bayes’ theorem, we can find the required probability:

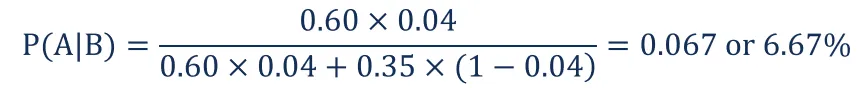


Image: Sample Calculation

Reference: <https://cdn.corporatefinanceinstitute.com/assets/bayes-theorem3.png>

Thus, the probability that the shares of a company that replaces its CEO will grow by more than 5% is 6.67%.

## Maximum likelihood

### **What is Maximum likelihood**

Maximum likelihood is a widely used technique for estimation with applications in many areas including time series modeling, panel data, discrete data, and even machine learning.

Here we cover the fundamentals of maximum likelihood estimation.

In particular, we discuss:

1. The basic theory of maximum likelihood.
2. The advantages and disadvantages of maximum likelihood estimation.
3. The log-likelihood function.
4. Modeling applications.

In addition, we consider a simple application of maximum likelihood estimation to a linear regression model.

### **What is Maximum Likelihood Estimation?**

Maximum likelihood estimation is a statistical method for estimating the parameters of a model. In maximum likelihood estimation, the parameters are chosen to maximize the likelihood that the assumed model results in the observed data.

This implies that in order to implement maximum likelihood estimation we must:

1. Assume a model, also known as a data generating process, for our data.
2. Be able to derive the likelihood function for our data, given our assumed model

Once the likelihood function is derived, maximum likelihood estimation is nothing more than a simple optimization problem.

### **What are the Advantages and Disadvantages of Maximum Likelihood Estimation?**

At this point, you may be wondering why you should pick maximum likelihood estimation over other methods such as least squares regression or the generalized method of moments. The reality is that we shouldn't always choose maximum likelihood estimation. Like any estimation technique, maximum likelihood estimation has advantages and disadvantages.

#### **Advantages of Maximum Likelihood Estimation**

There are many advantages of maximum likelihood estimation:

* If the model is correctly assumed, the maximum likelihood estimator is the most efficient estimator.
* It provides a consistent but flexible approach which makes it suitable for a wide variety of applications, including cases where assumptions of other models are violated.
* It results in unbiased estimates in larger samples.

Efficiency is one measure of the quality of an estimator. An efficient estimator is one that has a small variance or mean squared error.

#### **Disadvantages of Maximum Likelihood Estimation**

* It relies on the assumption of a model and the derivation of the likelihood function which is not always easy.
* Like other optimization problems, maximum likelihood estimation can be sensitive to the choice of starting values.
* Depending on the complexity of the likelihood function, the numerical estimation can be computationally expensive.
* Estimates can be biased in small samples.

### **What is the Likelihood Function?**

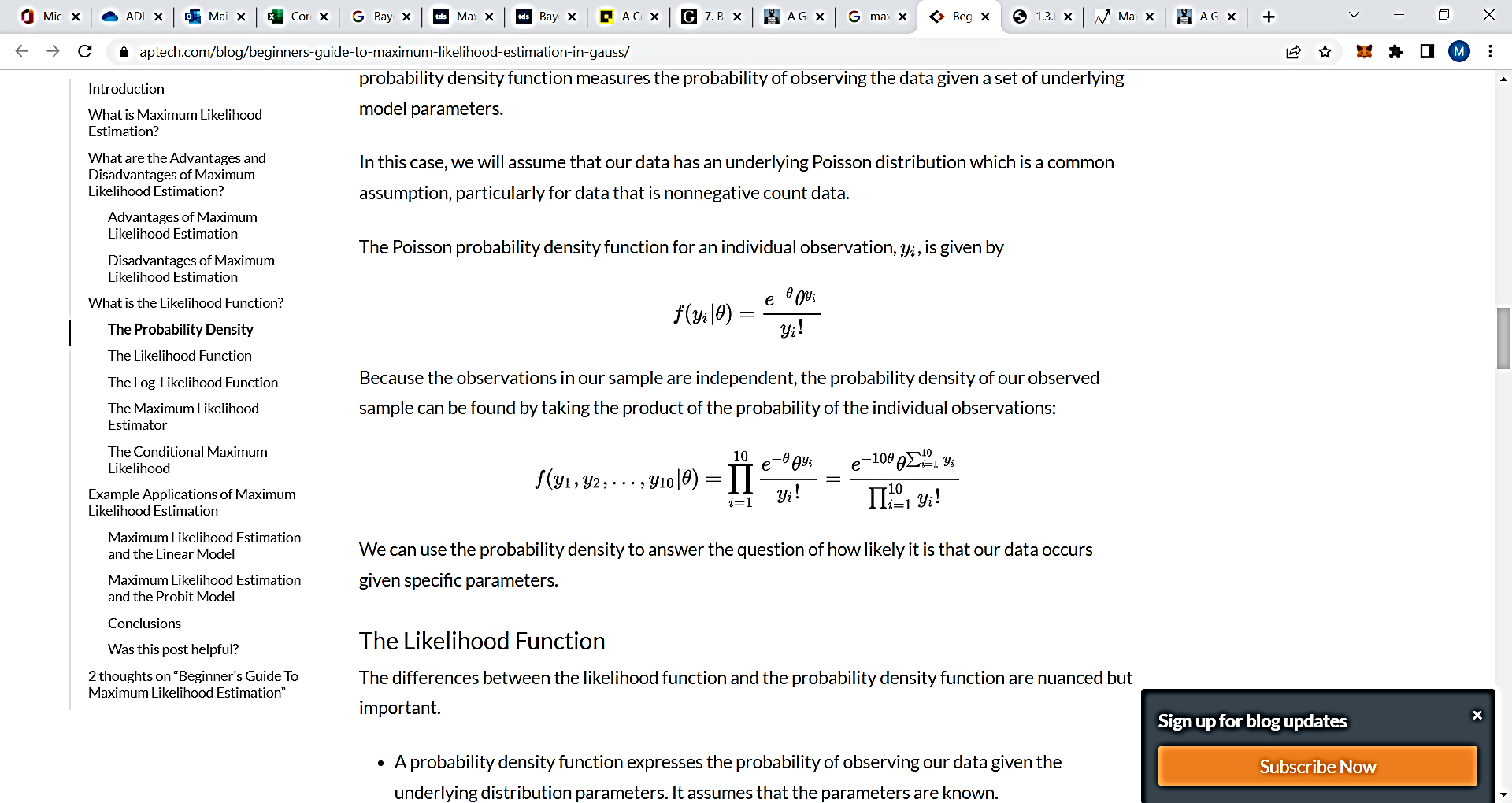
Maximum likelihood estimation hinges on the derivation of the likelihood function. For this reason, it is important to have a good understanding of what the likelihood function is and where it comes from.

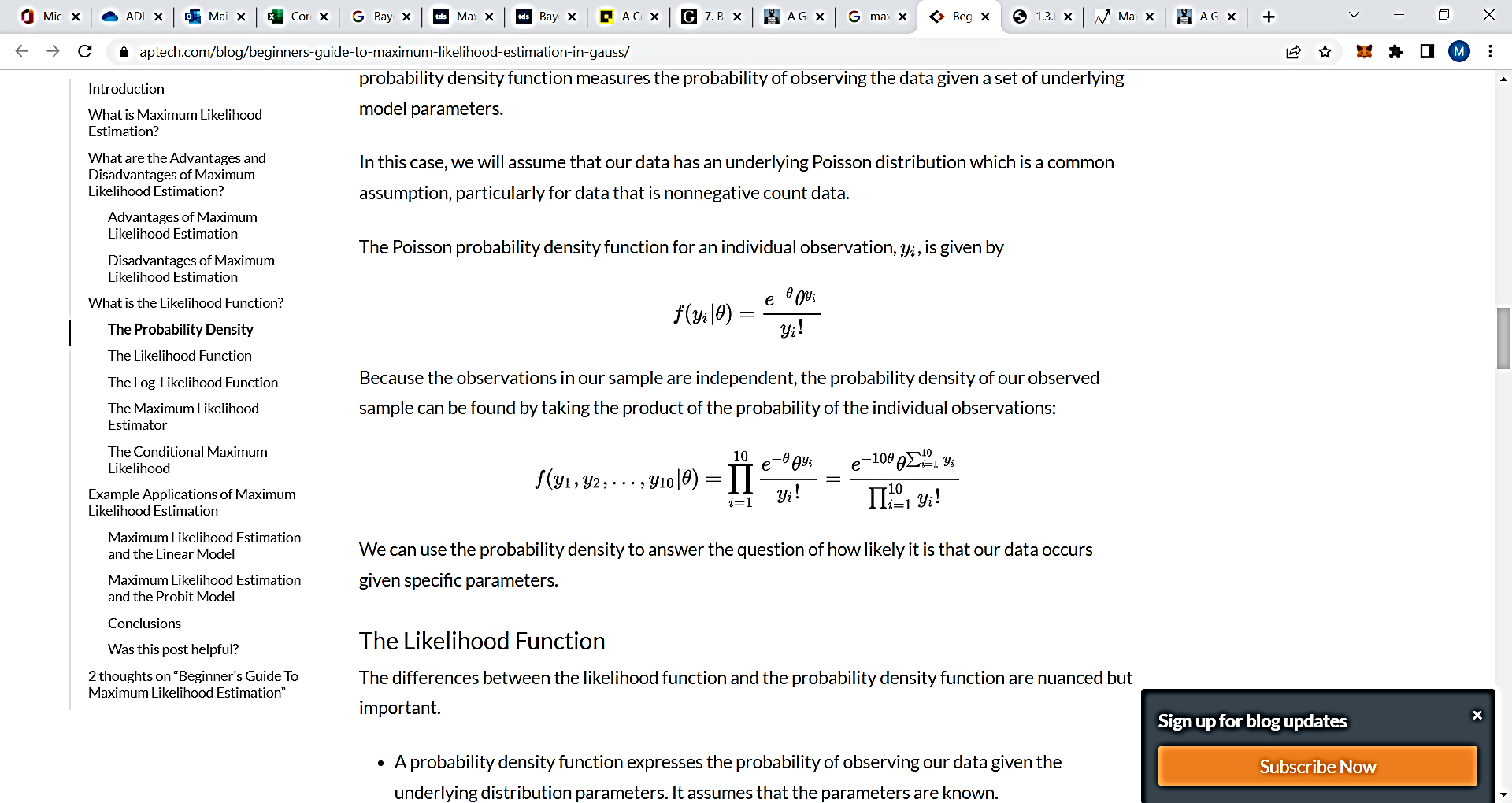
Let's start with the very simple case where we have one series with 10 independent observations: 5, 0, 1, 1, 0, 3, 2, 3, 4, 1.

#### **The Probability Density**

The first step in maximum likelihood estimation is to assume a probability distribution for the data. A probability density function measures the probability of observing the data given a set of underlying model parameters.

In this case, we will assume that our data has an underlying Poisson distribution which is a common assumption, particularly for data that is nonnegative count data.

The Poisson probability density function for an individual observation, is given by

Because the observations in our sample are independent, the probability density of our observed sample can be found by taking the product of the probability of the individual observations:

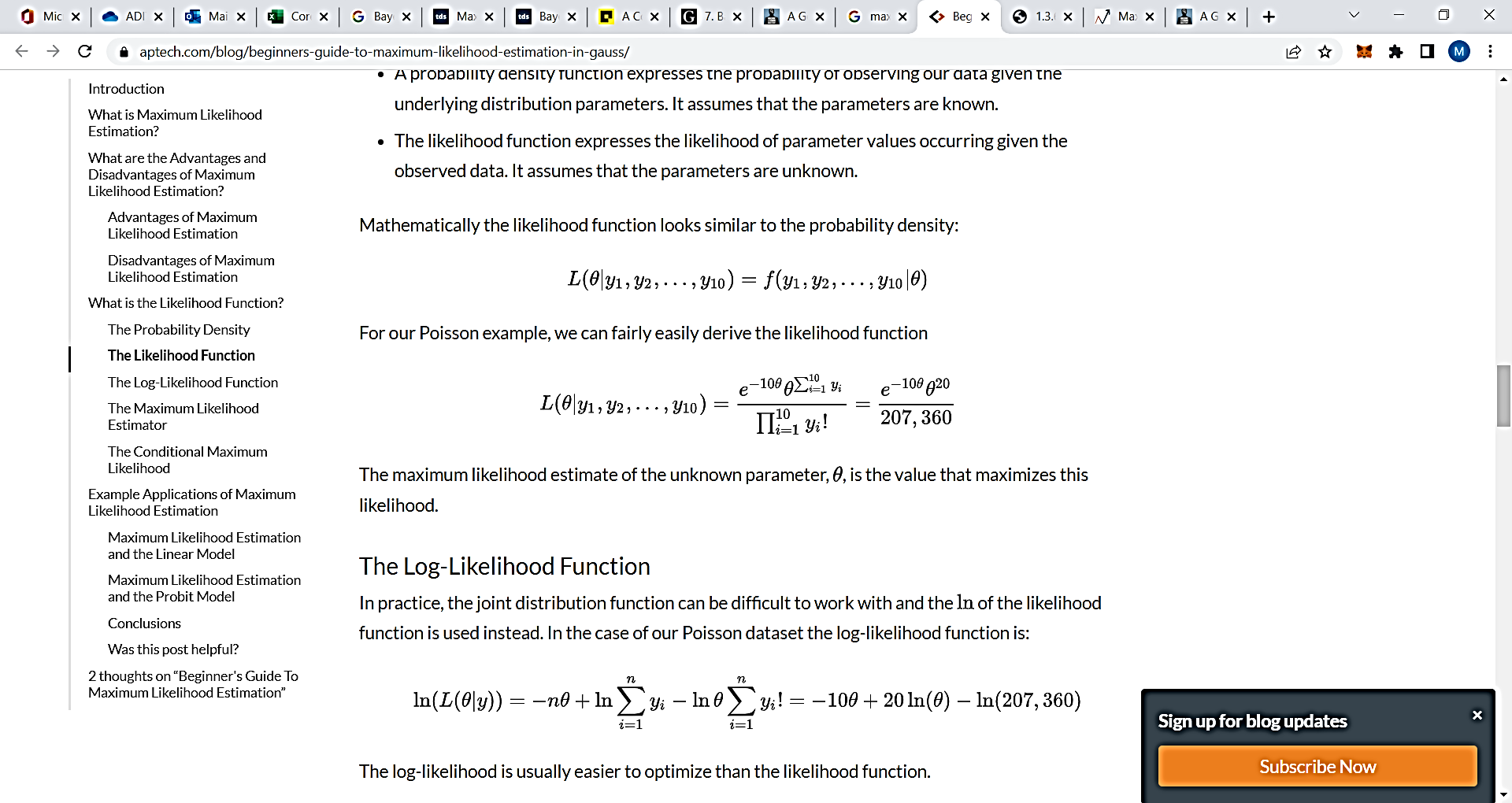
We can use the probability density to answer the question of how likely it is that our data occurs given specific parameters.

#### **The Likelihood Function**

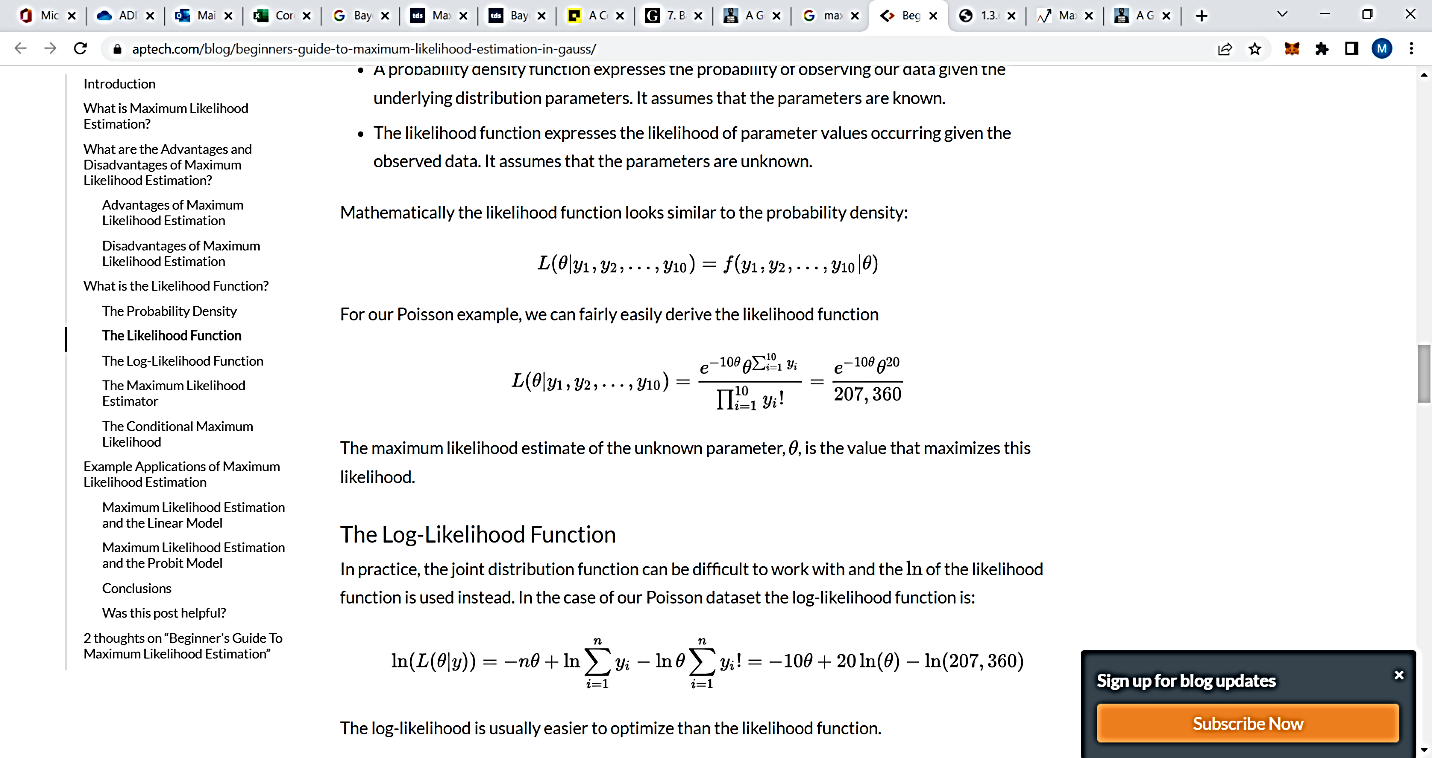
The differences between the likelihood function and the probability density function are nuanced but important.

* A probability density function expresses the probability of observing our data given the underlying distribution parameters. It assumes that the parameters are known.
* The likelihood function expresses the likelihood of parameter values occurring given the observed data. It assumes that the parameters are unknown.

Mathematically the likelihood function looks similar to the probability density:

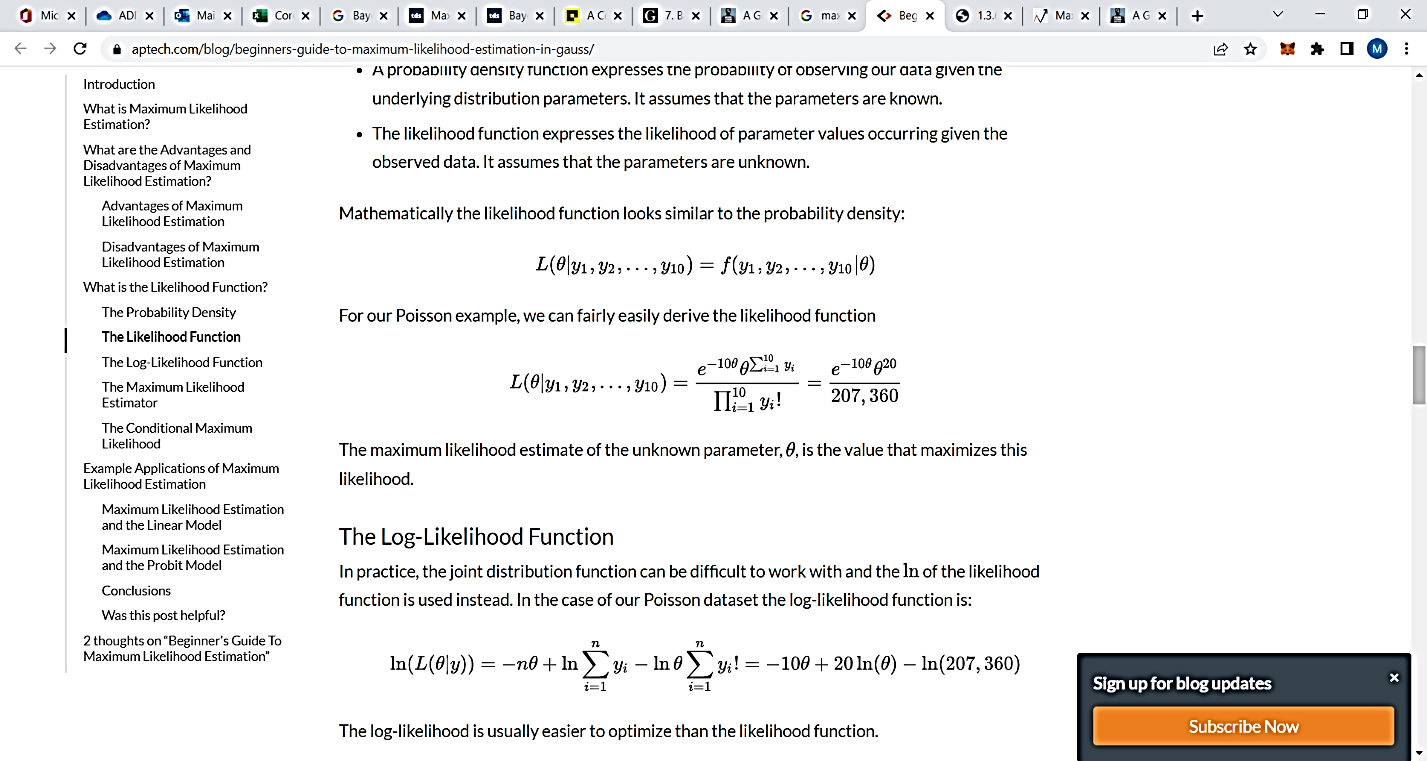


For our Poisson example, we can fairly easily derive the likelihood function



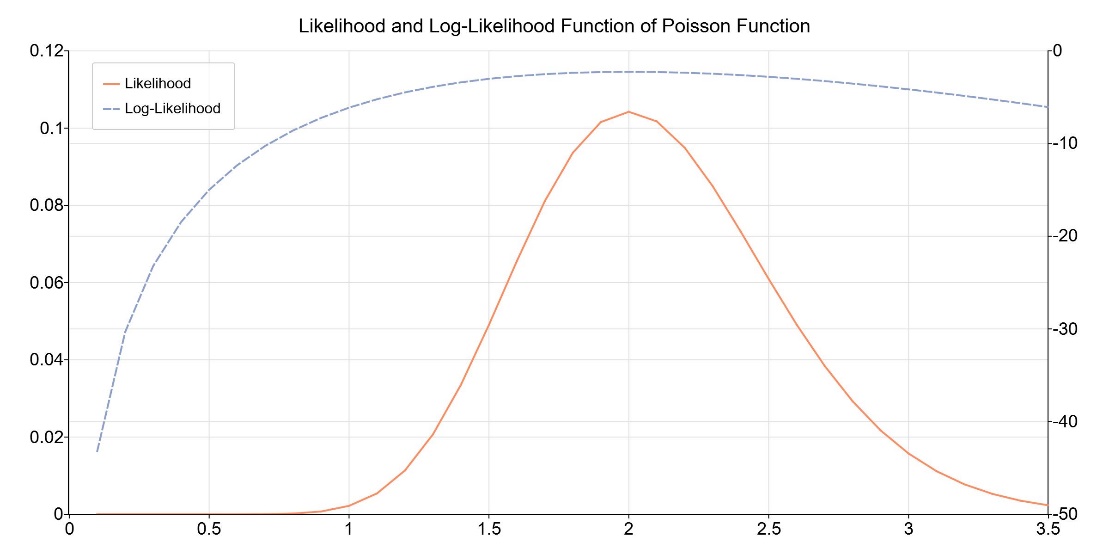
The maximum likelihood estimate of the unknown parameter, , is the value that maximizes this likelihood.

#### **The Log-Likelihood Function**

In practice, the joint distribution function can be difficult to work with and the of the likelihood function is used instead. In the case of our Poisson dataset the log-likelihood function is:

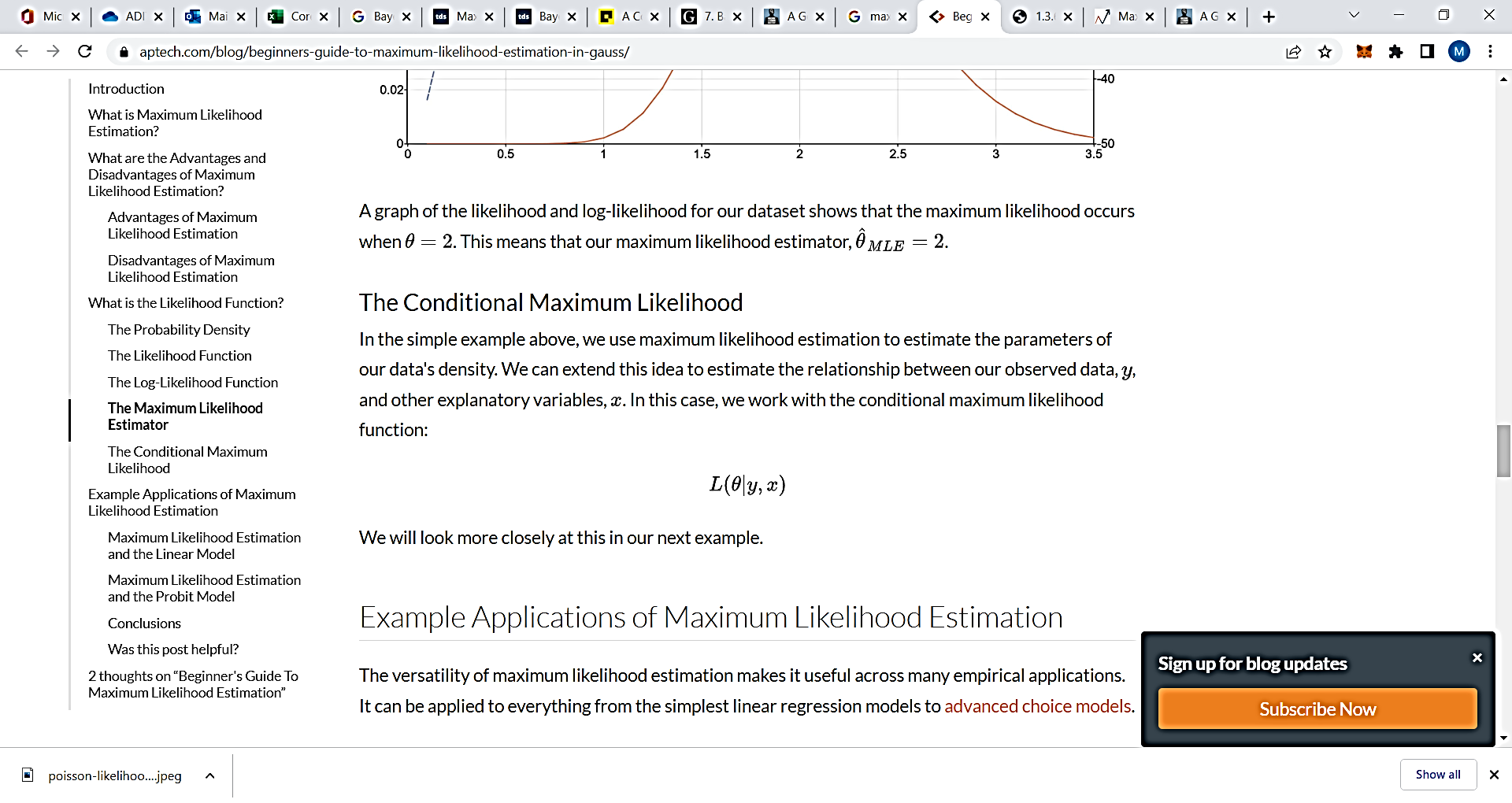
The log-likelihood is usually easier to optimize than the likelihood function.

#### **The Maximum Likelihood Estimator**



#### Image: The Maximum Likelihood Estimator

Reference: https://www.aptech.com/wp-content/uploads/2020/09/poisson-likelihood-function.jpeg

A graph of the likelihood and log-likelihood for our dataset shows that the maximum likelihood occurs when θ= 2. This means that our maximum likelihood estimator,

## Maximum Likelihood Estimation (MLE) vs Bayesian Estimation.

|  |  |  |
| --- | --- | --- |
|  | MLE | Bayesian Estimation |
| Predictions | We make predictions utilizing the latent variables in the density function to compute a probability. | We make predictions using the posterior distribution and the parameters which are considered as the random variables. |
| Situations to working with | Data with minimal values and the knowledge of prior is low. We can use MLE. | Data with sparse value and knowledge about the reliability of priors is high. We can use Bayesian estimation. |
| Complexity | MLE is less complex because we require to compute only the likelihood function | Bayesian estimation is more complex because the computation requires the likelihood function, evidence, and prior. |

# Hypothesis Testing

## What Is Hypothesis Testing?

Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis.

Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data. Such data may come from a larger population, or from a data-generating process. The word "population" will be used for both of these cases in the following descriptions.

* Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data.
* The test provides evidence concerning the plausibility of the hypothesis, given the data.
* Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed.

## How Hypothesis Testing Works

In hypothesis testing, an analyst tests a statistical sample, with the goal of providing evidence on the plausibility of the null hypothesis.

Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed. All analysts use a random population sample to test two different hypotheses: the null hypothesis and the alternative hypothesis.

The null hypothesis is usually a hypothesis of equality between population parameters; e.g., a null hypothesis may state that the population mean return is equal to zero. The alternative hypothesis is effectively the opposite of a null hypothesis (e.g., the population mean return is not equal to zero). Thus, they are mutually exclusive, and only one can be true. However, one of the two hypotheses will always be true.

## Types of Hypothesis Testing

There are two types of Hypothesis Testing

1. Null hypothesis
2. Alternative hypothesis
3. **Null Hypothesis:** It is denoted by H0. A null hypothesis is the one in which sample observations result purely from chance. This means that the observations are not influenced by some non-random cause.
4. **Alternative Hypothesis:** It is denoted by Ha or H1. An alternative hypothesis is the one in which sample observations are influenced by some non-random cause. A hypothesis test concludes whether to reject the null hypothesis and accept the alternative hypothesis or to fail to reject the null hypothesis. The decision is based on the value of X and R.

## Steps in Hypothesis Testing

Econometricians follow a formal process to test a hypothesis and determine whether it is to be rejected. The steps include:

1. **Stating the Hypotheses**

The first step involves positioning the null and alternative hypotheses. Remember, that these are mutually exclusive. If one hypothesis states a fact, the other must reject it.

1. **Making Statistical Assumptions**

Consider statistical assumptions – such as independence of observations from each other, normality of observations, random errors and probability distribution of random errors, randomization during sampling, etc.

1. **Formulating an Analysis Plan**

This includes deciding the test which is to be carried out to test the hypothesis. At the same time, we need to decide how sample data will be used to test the null hypothesis.

1. **Investigating Sample Data**

At this stage, sample data is examined. It’s when we find scores – mean values, normal distribution, t distribution, z score, etc.

1. **Interpreting Results**

This stage involves making decision to either reject the null hypothesis in favor of alternative hypothesis or not to reject the null hypothesis.

## Accepting or Rejecting Null Hypothesis

This is an extension of the last step - interpreting results in the process of hypothesis testing. A null hypothesis is accepted or rejected basis P value and the region of acceptance.

P value – it is a function of the observed sample results. A threshold value is chosen before the test is conducted and is called the significance level, which is represented as α. If the calculated value of P ≤ α, it suggests the inconsistency between the observed data and the assumption that the null hypothesis is true. This suggests that the null hypothesis must be rejected. However, this doesn’t mean that alternative hypothesis can be accepted as true. This is when Type I error occurs.

## Real-World Example of Hypothesis Testing

If, for example, a person wants to test that a penny has exactly a 50% chance of landing on heads, the null hypothesis would be that 50% is correct, and the alternative hypothesis would be that 50% is not correct.

Mathematically, the null hypothesis would be represented as Ho: P = 0.5. The alternative hypothesis would be denoted as "Ha" and be identical to the null hypothesis, except with the equal sign struck-through, meaning that it does not equal 50%.

A random sample of 100 coin flips is taken, and the null hypothesis is then tested. If it is found that the 100 coin flips were distributed as 40 heads and 60 tails, the analyst would assume that a penny does not have a 50% chance of landing on heads and would reject the null hypothesis and accept the alternative hypothesis.

If, on the other hand, there were 48 heads and 52 tails, then it is plausible that the coin could be fair and still produce such a result. In cases such as this where the null hypothesis is "accepted," the analyst states that the difference between the expected results (50 heads and 50 tails) and the observed results (48 heads and 52 tails) is "explainable by chance alone."

# Central limit theorem

## What is the Central Limit Theorem (CLT)?

The Central Limit Theorem (CLT) is a statistical concept that states that the sample mean distribution of a random variable will assume a near-normal or normal distribution if the sample size is large enough. In simple terms, the theorem states that the sampling distribution of the mean approaches a normal distribution as the size of the sample increases, regardless of the shape of the original population distribution.

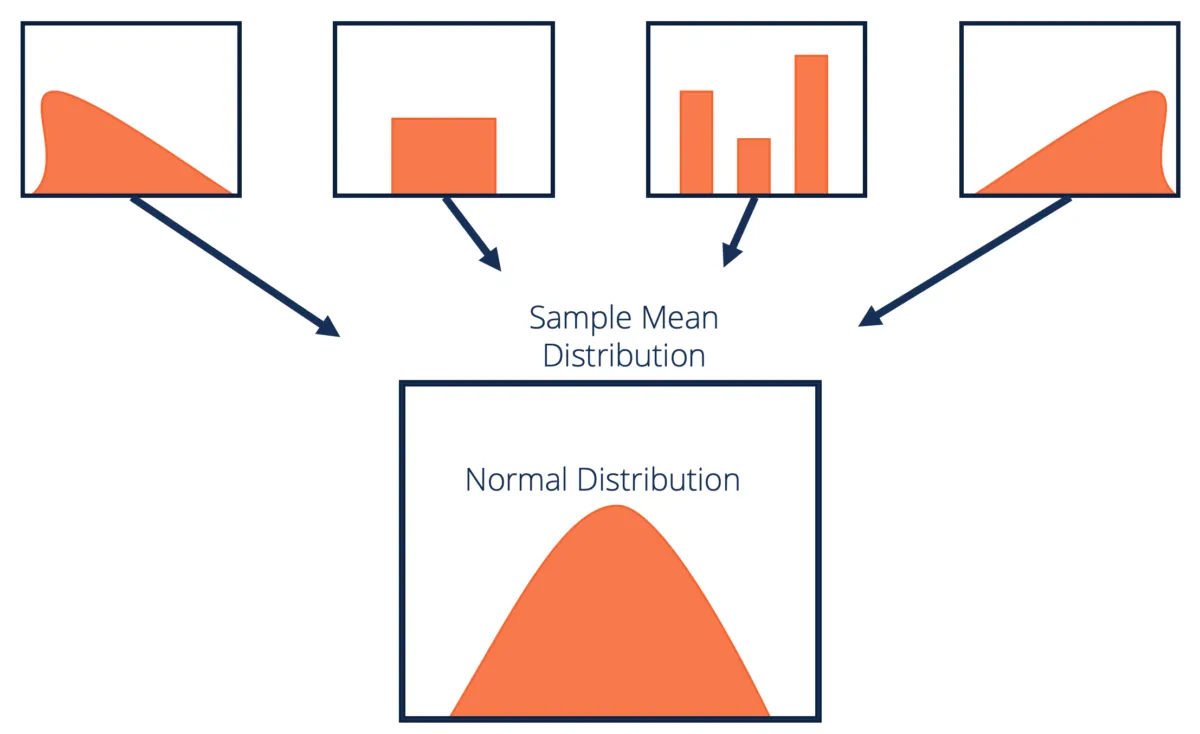


Image: Central Limit Theorem (CLT) Diagram showing Convergence to Normal Distribution

Reference: <https://cdn.corporatefinanceinstitute.com/assets/Central-Limit-Theorem-CLT-Diagram-1200x734.png>

As the user increases the number of samples to 30, 40, 50, etc., the graph of the sample means will move towards a normal distribution. The sample size must be 30 or higher for the central limit theorem to hold.

One of the most important components of the theorem is that the mean of the sample will be the mean of the entire population. If you calculate the mean of multiple samples of the population, add them up, and find their average, the result will be the estimate of the population mean.

The same applies when using standard deviation. If you calculate the standard deviation of all the samples in the population, add them up, and find the average, the result will be the standard deviation of the entire population.

## How Does the Central Limit Theorem Work?

The central limit theorem forms the basis of the probability distribution. It makes it easy to understand how population estimates behave when subjected to repeated sampling. When plotted on a graph, the theorem shows the shape of the distribution formed by means of repeated population samples.

As the sample sizes get bigger, the distribution of the means from the repeated samples tends to normalize and resemble a normal distribution. The result remains the same regardless of what the original shape of the distribution was. It can be illustrated in the figure below:

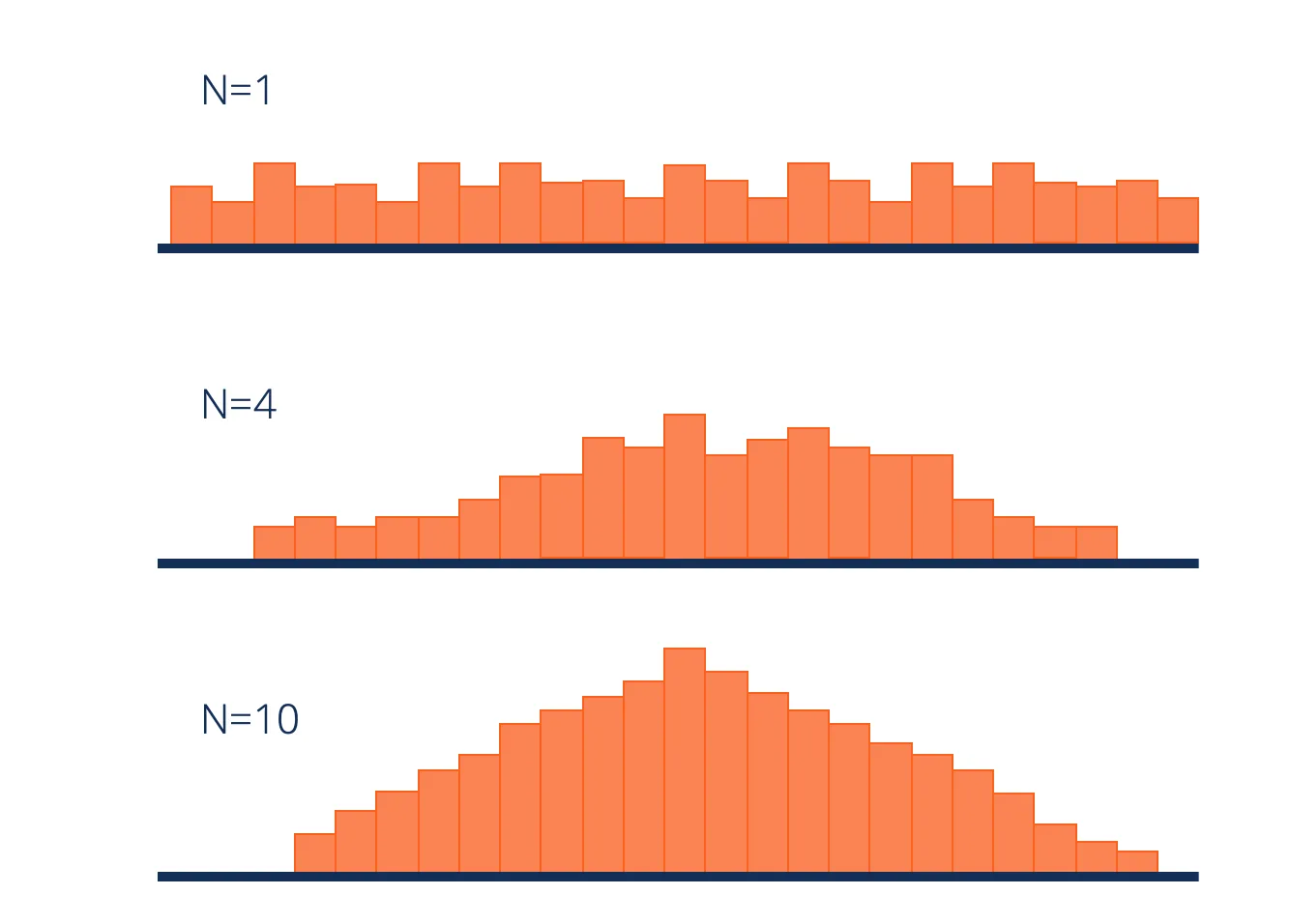


Image: Central Limit Theorem (CLT) - How it arises

Reference: <https://cdn.corporatefinanceinstitute.com/assets/Central-Limit-Theorem-CLT-How-it-works-and-how-it-arises.png>

From the figure above, we can deduce that despite the fact that the original shape of the distribution was uniform, it tends towards a normal distribution as the value of n (sample size) increases.

Apart from showing the shape that the sample means will take, the central limit theorem also gives an overview of the mean and variance of the distribution. The sample mean of the distribution is the actual population mean from which the samples were taken.

The variance of the sample distribution, on the other hand, is the variance of the population divided by n. Therefore, the larger the sample size of the distribution, the smaller the variance of the sample mean.

## Example of Central Limit Theorem

An investor is interested in estimating the return of ABC stock market index that is comprised of 100,000 stocks. Due to the large size of the index, the investor is unable to analyze each stock independently and instead chooses to use random sampling to get an estimate of the overall return of the index.

The investor picks random samples of the stocks, with each sample comprising at least 30 stocks. The samples must be random, and any previously selected samples must be replaced in subsequent samples to avoid bias.

If the first sample produces an average return of 7.5%, the next sample may produce an average return of 7.8%. With the nature of randomized sampling, each sample will produce a different result. As you increase the size of the sample size with each sample you pick, the sample means will start forming their own distributions.

The distribution of the sample means will move toward normal as the value of n increases. The average return of the stocks in the sample index estimates the return of the whole index of 100,000 stocks, and the average return is normally distributed.

## History of the Central Limit Theorem

The initial version of the central limit theorem was coined by Abraham De Moivre, a French-born mathematician. In an article published in 1733, De Moivre used the normal distribution to find the number of heads resulting from multiple tosses of a coin. The concept was unpopular at the time, and it was forgotten quickly.

However, in 1812, the concept was reintroduced by Pierre-Simon Laplace, another famous French mathematician. Laplace re-introduced the normal distribution concept in his work titled “Théorie Analytique des Probabilités,” where he attempted to approximate binomial distribution with the normal distribution.

The mathematician found that the average of independent random variables, when increased in number, tends to follow a normal distribution. At that time, Laplace’s findings on the central limit theorem attracted attention from other theorists and academicians.

Later in 1901, the central limit theorem was expanded by Aleksandr Lyapunov, a Russian mathematician. Lyapunov went a step ahead to define the concept in general terms and prove how the concept worked mathematically. The characteristic functions that he used to provide the theorem were adopted in modern probability theory.

# Chi-square test

## What Is a Chi-Square Statistic?

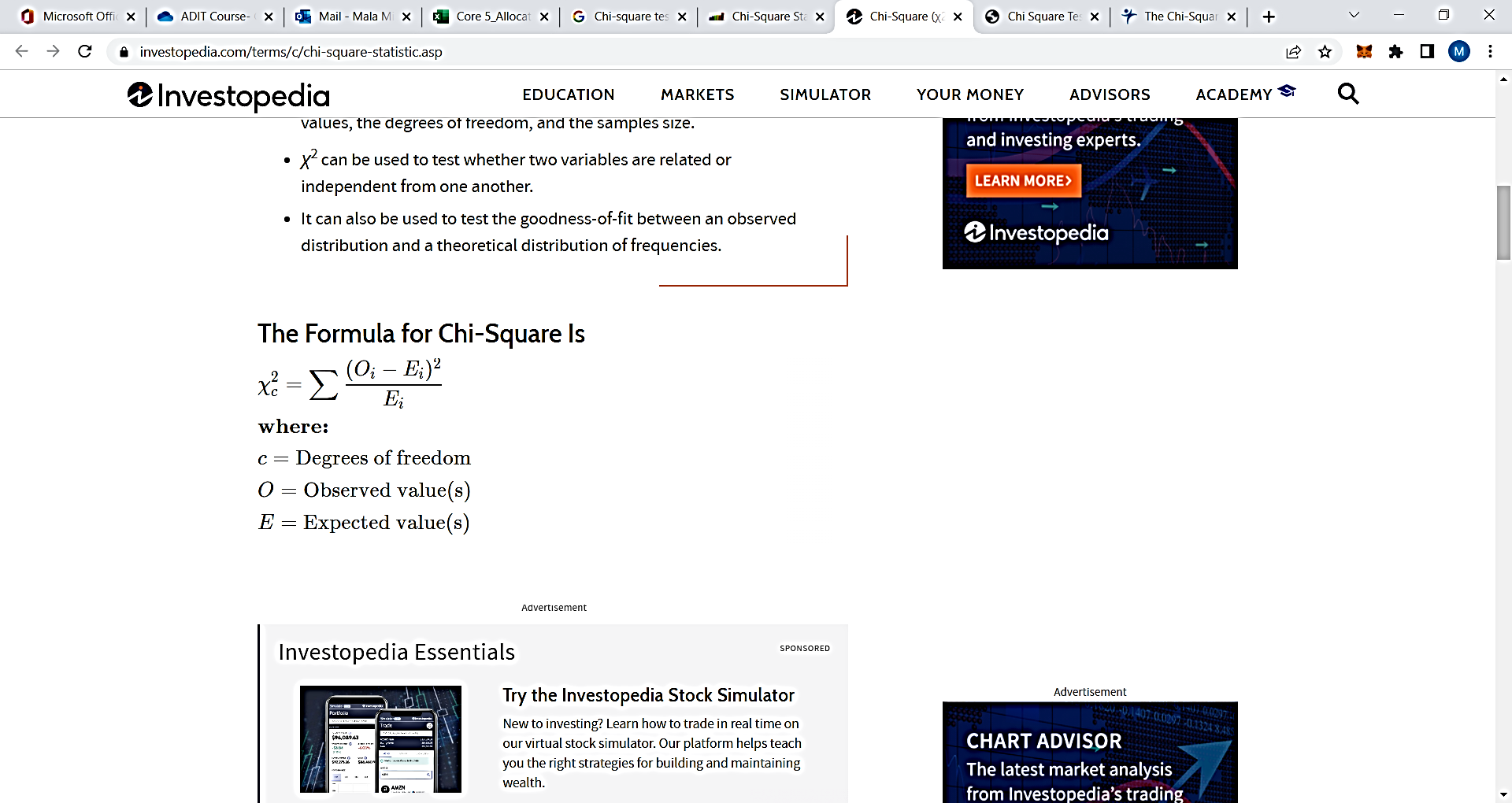
A chi-square (χ2) statistic is a test that measures how a model compares to actual observed data. The data used in calculating a chi-square statistic must be random, raw, mutually exclusive, drawn from independent variables, and drawn from a large enough sample. For example, the results of tossing a fair coin meet these criteria.

Chi-square tests are often used in hypothesis testing. The chi-square statistic compares the size of any discrepancies between the expected results and the actual results, given the size of the sample and the number of variables in the relationship.

For these tests, degrees of freedom are utilized to determine if a certain null hypothesis can be rejected based on the total number of variables and samples within the experiment. As with any statistic, the larger the sample size, the more reliable the result.

* A chi-square (χ2) statistic is a measure of the difference between the observed and expected frequencies of the outcomes of a set of events or variables.
* Chi-square is useful for analyzing such differences in categorical variables, especially those nominal in nature.
* χ2 depends on the size of the difference between actual and observed values, the degrees of freedom, and the samples size.
* χ2 can be used to test whether two variables are related or independent from one another.
* It can also be used to test the goodness-of-fit between an observed distribution and a theoretical distribution of frequencies.

## The Formula for Chi-Square Is



## What Does a Chi-Square Statistic Tell You?

There are two main kinds of chi-square tests: the test of independence, which asks a question of relationship, such as, "Is there a relationship between student sex and course choice?"; and the goodness-of-fit test, which asks something like "How well does the coin in my hand match a theoretically fair coin?"

Chi-square analysis is applied to categorical variables and is especially useful when those variables are nominal (where order doesn't matter, like marital status or gender).

**Independence**

When considering student sex and course choice, a χ2 test for independence could be used. To do this test, the researcher would collect data on the two chosen variables (sex and courses picked) and then compare the frequencies at which male and female students select among the offered classes using the formula given above and a χ2 statistical table.

If there is no relationship between sex and course selection (that is, if they are independent), then the actual frequencies at which male and female students select each offered course should be expected to be approximately equal, or conversely, the proportion of male and female students in any selected course should be approximately equal to the proportion of male and female students in the sample.

A χ2 test for independence can tell us how likely it is that random chance can explain any observed difference between the actual frequencies in the data and these theoretical expectations.

**Goodness-of-Fit**

χ2 provides a way to test how well a sample of data matches the (known or assumed) characteristics of the larger population that the sample is intended to represent. This is known as goodness of fit. If the sample data do not fit the expected properties of the population that we are interested in, then we would not want to use this sample to draw conclusions about the larger population.

**Example**

For example, consider an imaginary coin with exactly a 50/50 chance of landing heads or tails and a real coin that you toss 100 times. If this coin is fair, then it will also have an equal probability of landing on either side, and the expected result of tossing the coin 100 times is that heads will come up 50 times and tails will come up 50 times.\

In this case, χ2 can tell us how well the actual results of 100 coin flips compare to the theoretical model that a fair coin will give 50/50 results. The actual toss could come up 50/50, or 60/40, or even 90/10. The farther away the actual results of the 100 tosses is from 50/50, the less good the fit of this set of tosses is to the theoretical expectation of 50/50, and the more likely we might conclude that this coin is not actually a fair coin.

## When to Use a Chi-Square Test

A chi-square test is used to help determine if observed results are in line with expected results, and to rule out that observations are due to chance. A chi-square test is appropriate for this when the data being analyzed is from a random sample, and when the variable in question is a categorical variable. A categorical variable is one that consists of selections such as type of car, race, educational attainment, male vs. female, how much somebody likes a political candidate (from very much to very little), etc.

These types of data are often collected via survey responses or questionnaires. Therefore, chi-square analysis is often most useful in analyzing this type of data.

## What is a chi-square test used for?

Chi-square is a statistical test used to examine the differences between categorical variables from a random sample in order to judge goodness of fit between expected and observed results.

## Who uses chi-square analysis?

Since chi-square applies to categorical variables, it is most used by researchers who are studying survey response data. This type of research can range from demography to consumer and marketing research to political science and economics.

Is chi-square analysis used when the independent variable is nominal or ordinal?

A nominal variable is a categorical variable that differs by quality, but whose numerical order could be irrelevant. For instance, asking somebody their favorite color would produce a nominal variable. Asking somebody's age, on the other hand, would produce an ordinal set of data. Chi-square can be best applied to nominal data. ​

﻿

References

1. <https://www.mastersindatascience.org/learning/statistics-data-science/probability-theory/#:~:text=Probability%20theory%20is%20a%20branch,the%20analysis%20of%20chance%20events>.
2. <https://corporatefinanceinstitute.com/resources/knowledge/other/bayes-theorem/>
3. <https://www.aptech.com/blog/beginners-guide-to-maximum-likelihood-estimation-in-gauss/>
4. <https://www.investopedia.com/articles/active-trading/092214/hypothesis-testing-finance-concept-examples.asp>
5. <https://www.managementstudyguide.com/hypothesis-testing.htm>
6. <https://www.investopedia.com/terms/h/hypothesistesting.asp>
7. <https://corporatefinanceinstitute.com/resources/knowledge/other/central-limit-theorem/>
8. <https://www.investopedia.com/terms/c/chi-square-statistic.asp>