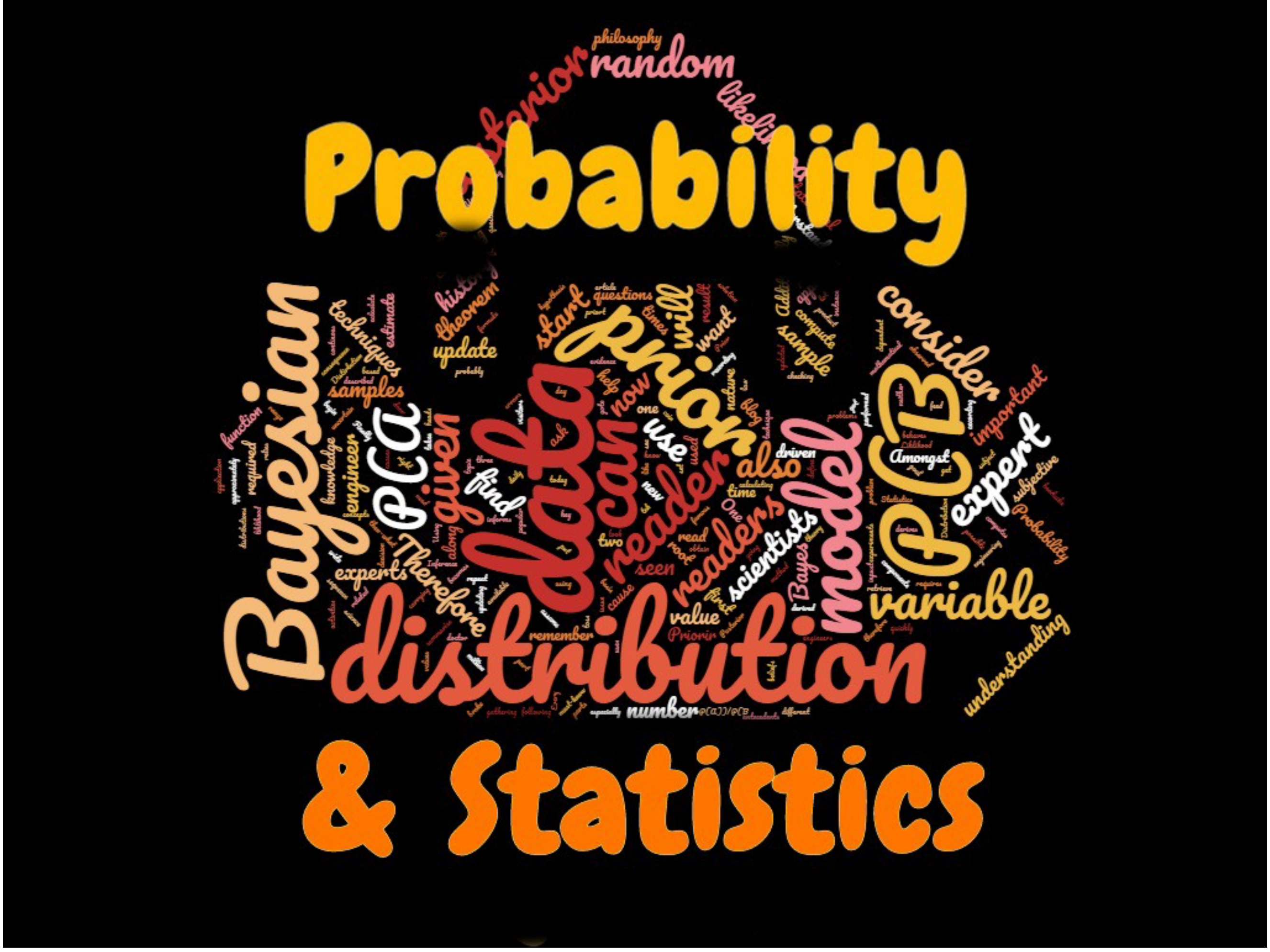
Understanding Statistics And Probability: Bayesian Inference

Understanding One Of The Key Fields Of Statistics That Is Gaining Interest In A Number Of Sectors

Bayesian inference is one of the most popular statistical techniques. It is a technique whereby the prior probabilities of an event are updated when the new data is gathered. Bayesian inference is a data-driven technique.

The Bayesian models are traditionally one of the first models to use. They are used as the baseline models as they are based on the simplistic view of the world and enable the scientists to explain the reasoning easier. Consequently, Bayesian inference is one of the most important techniques to learn in statistics.

> This article will introduce readers to Bayesian inference. It's one of the must-know topics.



An important concept of Probability And Statistics

Article Aim

This article will provide an overview of the following concepts:

1. What Is Bayesian Inference?
2. What Is Baye's Theorem?
3. Examples To Understand The Concepts
4. Naive Bayes Model

Bayesian inference is used in a large number of sectors including insurance, healthcare, e-commerce, sports, law amongst others. Bayesian inference is heavily used in classification algorithms whereby we attempt to classify/group text or numbers into their appropriate classes.

Furthermore, it is growing in interest in banking and in particular in the finance sector.

1. What is Bayesian Inference?

Before I explain what Bayesian Inference is, let's understand the key building blocks first.

I will start by illustrating an example.

Computer Engineers Example

Let's consider that I have a computer that stopped working. There are two computer engineers in my neighborhood who can fix the computer.

Both of the engineers claim to have different techniques to diagnose and fix the problem.

First Engineer — Frequentist Approach

The first engineer has a model, made up of a mathematical equation. This model is built based on the frequency of an event. The model requires a set of inputs to compute the diagnosis of why the computer stopped working.

The way the first engineer diagnoses the problem is by asking questions that the model requires as inputs.

As an instance, the engineer would ask about the computer specification such as the operating system, hard disk size and processor name. He would then feed the answers to the model and the model would then give the reasons of why the computer broke down.

The model will use the observed frequency of the events to diagnose why the computer stopped working.

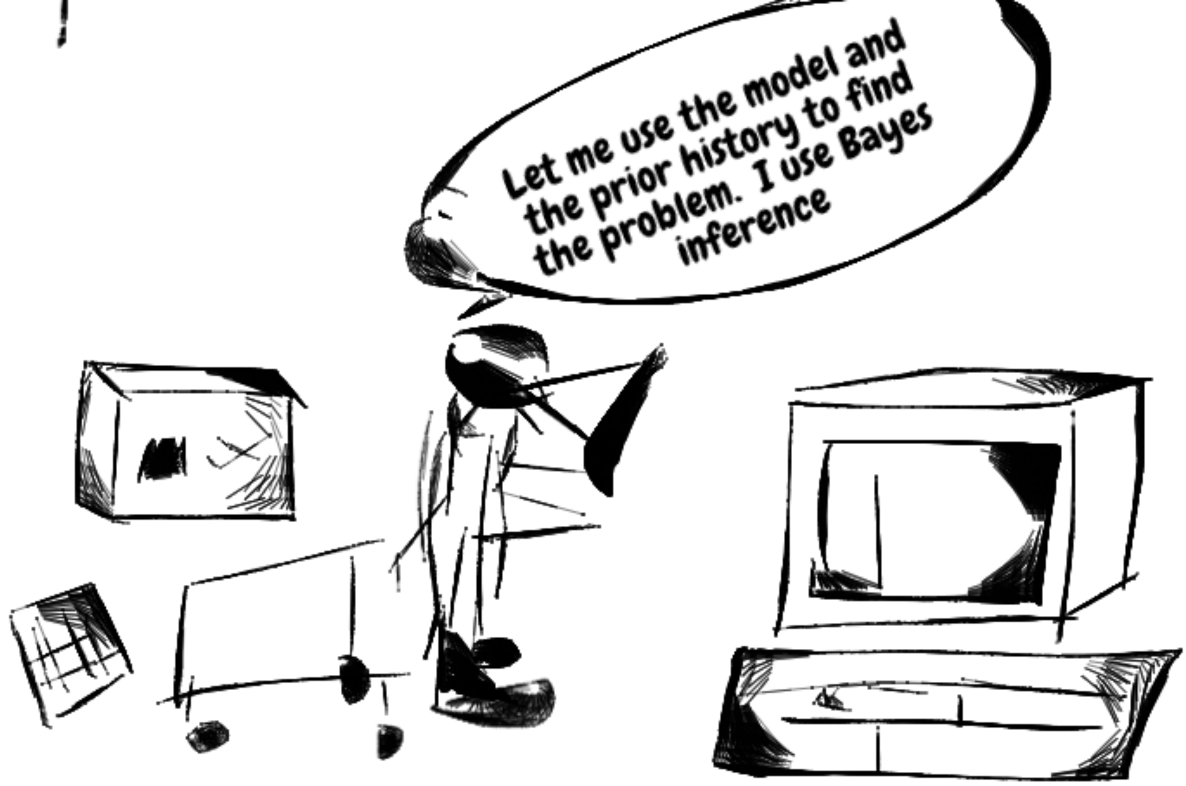
Frequentist Approach

Second Engineer — Bayesian Approach

The second engineer has a slightly different mechanism to diagnose the issue. He also uses the same model that the first engineer was using but along with the model, he also takes advantage of the prior history of the event to diagnose the problem. Therefore, the engineer would ask the same questions as the first engineer does but he would also inquire about the prior history of the problem.

As an instance, the second engineer would ask about the computer specification such as the operating system, hard disk size, and processor name. Additionally, he would ask questions about the historical history of the computer such as whether it has stopped working in the past and the reasons why it happened if known along with the number of times the event occured.

The engineer would gather the appropriate history of that individual computer to get a better understanding of the problem. This technique is known as Bayesian inference. In a nutshell, in Bayesian inference, we use the prior history along with the known model to compute posterior results.



Bayesian Approach

> The first engineer used the Frequentist approach and the second engineer used the Bayesian inference approach.

The frequentist approach is not accurate with a small sample size as it is based on the observed frequency of positive events occurring whereas the Bayesian approach relies on the prior belief regarding the probability of an event occurring. Having said that, the frequentist approach is easier to implement and understand than the Bayesian inference and is typically used for large sample sizes.

Bayesian inference is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available. Bayesian updating is particularly important in the dynamic analysis of a sequence of data.

The technique of Bayesian inference is based on Bayes' theorem. Bayes' theorem can help us update our knowledge of a random variable by using the prior and likelihood distributions to calculate the posterior distribution.

This brings us to the second part of the article.

2. Bayes' Theorem

In simplistic terms, the Bayes' theorem calculates the posterior probability of an event. It uses the prior probability along with the likelihood probability of the event.

Let's consider that we want to estimate how a target variable behaves. This target variable can be random in nature. Therefore, we begin by gathering data for the random sample. This data is our sample set with its own sampling distribution. As each sample has different data points, the distribution can help us quantify the errors in our sampling techniques.

Once the samples are gathered and the experiments are performed, we can now use the Bayes' theorem to obtain new information to update our prior understanding.

Bayes' theorem is a framework that enables us in calculating the probability of one event occurring given that the other event has already occurred.

Bayes' Theorem Formula

Bayes' theorem for two random variables A and B is:

Bayes' Theorem Formula

Let's understand the formula

The above formula states that there are two events: A and B

We are trying to find the conditional probability; the probability of event A occurring given that event B has already occurred. This is known as the posterior distribution.

1. It is computed by taking the joint probability of events A and B by calculating the product of the probability of event B occurring given that event A has already occurred and the unconditional probability of event A. This is shown in the numerator.

2. Finally, we divide the joint probability by the probability of event B occurring.

The three main components

Bayesian inference derives the posterior probability. It assumes that the posterior probability is a result of two main inputs (for simplicity): a prior probability and a likelihood function. The likelihood function is derived from a statistical model itself. Bayesian inference computes the posterior probability according to Bayes' theorem.

1. What we are computing is the posterior distribution. This is P(A given B).

2. What we already have is P(B). This is the probability that event B has already occurred.

3. Finally, P(B|A) is the likelihood probability. It is the probability that B is occurring given that A has already occurred.

Every time we want to predict a random variable, we already have a prior known distribution. As an instance, we have a prior distribution that if we toss a coin a million times then the number of times we'll see heads is approximately 50%

> The goal of Bayesian statistics is to compute a posterior distribution.

With the prior distribution, we use the Bayes theorem to obtain a posterior distribution. This is our updated understanding now that we have seen the data. Using the posterior distribution, we can summarise our understanding of the data.

> Bayesian inference is data driven because the prior distribution and the posterior distribution is driven by

> the data

3. Examples To Understand The Concept

I believe in applying the concept to solve practical problems because only then we can understand the concepts thoroughly.

1. Let's go over a simple example: Running And Training

Let's consider that you are training to run a 5km race within 25 minutes.

You want to know the probability of achieving the target time for the run if you train in the gym.

Let's use the Bayesian formula to find out the probability

At a high level, we need four numbers

1. P(R) is the probability of running a 5km run within 25 minutes. It is 50% based on the gathered data therefore 50% of the people run a 5km within 25 minutes.
2. P(T) is the probability of training in the gym. It is 60% based on the gathered data therefore 60% of the people train regardless of whether they have historically completed the run within 25 minutes or

not.

3. P(T|R) is the probability that an individual is training in the gym given that he /she has completed a 5km run within 25 minutes. It is 75%.

Question: What is the probability of achieving the target time of running a 5km run within 25 minutes given you train in the gym?

Therefore the question is to find P(R|T). We know it is calculated as:

The answer is 62.5%

2. Another Example — Bond Default In Financial Organisation

Let's consider that we work in a financial organization and want to find the probability of a bond X defaulting given that another bond Y has already defaulted.

We can create a probability matrix to visualise the problem:

The probability matrix of the bonds

It is really straightforward to solve using the probability matrix.

1. The joint probability that the bond Y is defaulting given that the bond X has already defaulted is 10% (lower-right box)
2. The probability of bond Y defaulting regardless of whether bond X has defaulted or not is 20% (sum of the second column)

Answer for the example

Therefore there is a 50% chance that the bond X will default if the bond Y has already defaulted.

3. Let’s Understand With Another Example

Let's assume we know that we have X number of readers reading this blog on a daily basis. I want to calculate the probability that a reader is a Python developer given that he/she is a data scientist.

Let's refer to A as a reader who is a data scientist.

- The probability that a reader is a data scientist is P(A)
- The probability that a reader is not a data scientist is P(A')

Let's refer to B as a reader who is a Python developer.

- The probability that a reader is a Python developer is P(B)
- The probability that a reader is not a Python developer is P(B')

What is the probability that a reader is a Python developer given that he/she is a data scientist?

The solution is:

The formula for the question

Therefore, we can start by collecting the required data for a day.

Let’s consider the following numbers:

There were 200 readers who read this blog today

Amongst the 200 readers:

- 20 readers were data scientists given they were Python developers
- 30 readers were not data scientists and they were Python developers
- 60 readers were data scientists and not Python developers.
- 90 readers were neither Python developers nor data scientists.

The question is then — what is the probability of a reader being a Python expert given that he also is a data scientist? P(B|A)

We can compute a tree to visualise it:

The probability decision tree

The formula for the question

The probability is therefore calculated as:

The equation to get the probability

This gives us 25%

Therefore, there is a 25% chance that the reader is a

Python developer given he/she is a data scientist.

4. Bayesian Models — Naive Bayes

Lastly, I wanted to introduce the readers to the Naive Bayes model. Naive Bayes model is used as the benchmark/baseline model in most classification data science projects, in particular in text mining projects.

The model is probabilistic in nature. It is based on the concepts that I have explained above. Hence it uses Baye's theorem to calculate the results.

The model is called Naive because its foundation is on an extremely simplified version of reality. It assumes that the features in the data are independent of each other given the class label. This assumption is not entirely true for most of the data set.

As an instance, the features might be linked together. To explain, words like National History Museum or Buckingham Palace or White House or Black Day amongst others, have a completely different meaning when they are written together than when they are written separately in different sentences. Hence it shows that there is some form of dependence on the features.

> The Naïve Bayes model assumes that all of the features are independent for simplicity. Sometimes a model that can be explained in more important than an

> accurate model that can't be explained

The Bayesian models are traditionally one of the first models to use. They are easy to explain as they are based on the simplistic view of the world. Furthermore, the parameters are easy to understand. Hence why Naive Bayes classifier is seen as the baseline model.

Summary

This article explained what Bayesian inference and Bayesian theorem are.

Thank you for reading

In particular, it provided an overview of the following topics:

1. What Is Bayesian Inference?
2. What Is Baye’s Theorem?
3. Examples To Understand The Concepts
4. Naive Bayes Model

Bayesian inference is used in a large number of sectors including insurance, healthcare, e-commerce, sports, law amongst others. Bayesian inference is heavily used in classification algorithms whereby we attempt to classify/group text or numbers into their appropriate classes.

Furthermore, it is growing in interest in banking and in particular in the finance sector.