Bayes' Theorem: The Holy Grail of Data Science

Artem Oppermann   December 22, 2018

Intuitive derivation of the Bayes' Theorem

Artem Oppermann
Dec 22

Bayes' Theorem is perhaps the most important theorem in the field of mathematical statistics and probability theory. For this reason, the theorem finds its use very often in the field of data science.
In this article, I am going to use a practical problem to intuitively derive the Bayes' Theorem.

0. Introduction

Bayes' theorem, named after 18th-century British mathematician Thomas Bayes, is a mathematical formula for determining conditional probabilities. This theorem has enormous importance in the field of data science. For example one of many applications of Bayes' theorem is the Bayesian inference, a particular approach to statistical inference.

Bayesian inference is a method in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available. Bayesian inference has found application in a wide range of activities, including **science**, **engineering**, **philosophy**, **medicine**, **sport**, **and law**.

In finance, for example, Bayes' theorem can be used to rate the risk of lending money to potential borrowers. **In medicine**, the theorem can be used to determine the accuracy of medical test results by taking into consideration how likely any given person is to have a disease and the general accuracy of the test.

Let us get practical now…

2. Problem Statement

Consider two bowls **X** and **Y**, filled with oranges and blueberries. In this case, **you know exactly** how many oranges and blueberries are in each of the two bowls.

If I ask you how likely it is to pick an orange from bowl **X**, you can tell exactly the probability. Since there are 11 items in the bowl **X** and 3 of those are oranges, the probability to pick an orange would be *p(orange)*=3/11.
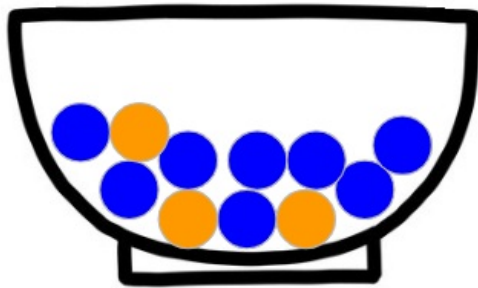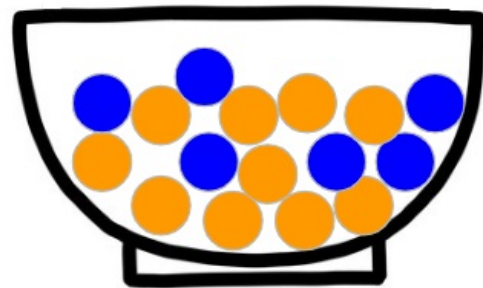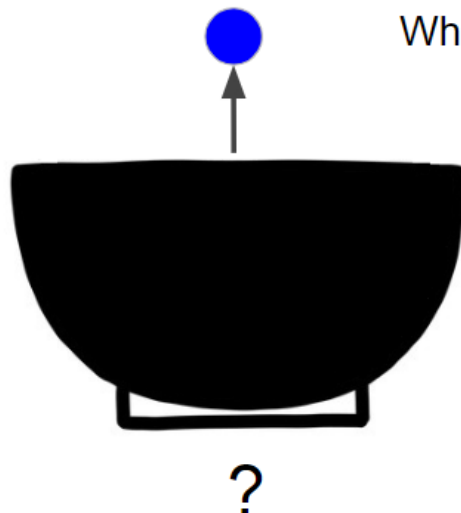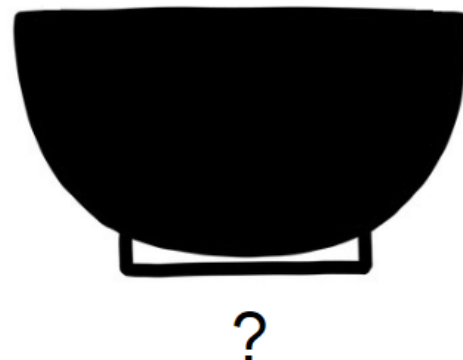

Bowl **X** and Bowl **Y** filled with oranges and blueberries.

Reversed Case

I pick blindly an item and get a blueberry.

> But what happens if I pick an item from a random bowl. And let's say I picked a blueberry. Can you tell the probability which bowl the blueberry was picked from?

> This is the question that can be answered using the Bayes' Theorem.

3. Bayes Theorem Derivation

In order to derive the Bayes' Theorem, we are going to simulate an experiment. In this experiment, we roll a dice. Every time the dice shows the number 3 or less we will pick an item from bowl **X**, for the number 4 or higher we will pick an item from bowl **Y**. And we are

going to do this *N=300* times. And to simplify matters, we introduce the following abbreviations:

Blueberry := B, Orange := O, Bowl X := X, Bowl Y := Y

After we roll the dice *N=300* times we will obtain some statistical results regarding the number of items which were picked from the two bowls. A hypothetical result of the experiment is shown in Fig. 1. Here **s** represents the bowl or the "source" where an item was picked from. **y** is the observable variable (blueberry or orange).
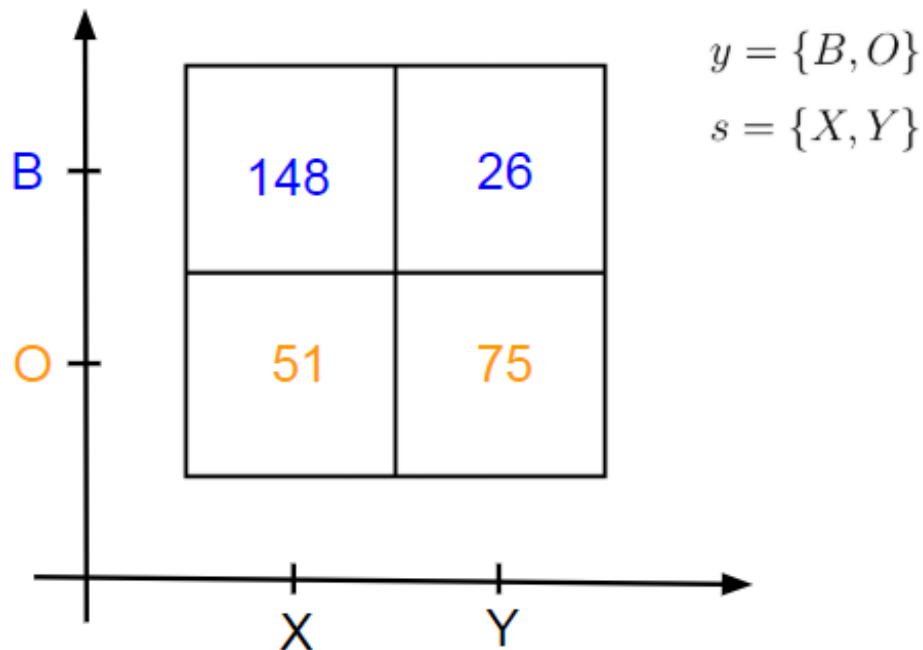

Fig. 1 Statistical results

The figure tells us that we have picked…

- … 148 times a blueberry from the bowl X: **n(s=X, y=B)=148**
- … 26 times a blueberry from the bowl Y: **n(s=Y, y=B)=26**
- … 51 times an orange from the bowl X: **n(s=X, y=O)=51**
- … 75 times an orange from the bowl Y: **n(s=Y, y=O)=75**

> Given these statistical numbers, we can now ask some interesting questions…

What is a probability to pick a random item from bowl X?

The obtain this probability that we denote as **p(s=X)** we must divide the number of items picked only from bowl **X** divided by the number *N=300* of total picks. Here is **n(s=X, y=B)=148** the number of blueberries picked from **X** and **n(s=X, y=O)=51** the number of orangespicked from **X.** Thus, the probability to pick any item from **X** looks as follows:

$$p(s = X) = \frac{n(s = X, y = B) + n(s = X, y = O)}{N}$$

$$= \frac{148 + 51}{300} \approx \frac{2}{3}$$

Eq. 1 Probability to pick an item from bowl X.

> **Note**: This kind of probability is called the "Prior Probability". In Bayesian statistical inference, the prior probability is the probability of an event before new data is collected. In this case **p(s=X)** tells the probability for picking an item from **X**, without knowing which item it is exactly.

Accordingly, the probability **p(s=Y)** to pick an item from **Y** is:

Eq. 2 Probability to pick an item from bowl Y.

$$p(s = Y) \approx \frac{1}{3}$$

What is a probability to pick an orange/blueberry?

This time we want to find out how likely it is to pick an orange or blueberry without considering a specific bowl. We denote these probabilities as **p(y=O)** and **p(y=B)**. The calculation is done analogously to the previous case. We are dividing the number of picks of a specific item by the number of total picks. The resulting probabilities are given by Eq. 3 and Eq. 4:

$$p(y = O) = \frac{n(s = X, y = O) + n(s = Y, y = O)}{N}$$

$$= \frac{51 + 75}{300} \approx \frac{4}{10}$$

Eq. 3 Probability to pick an orange.

Eq. 4 Probability to pick a blueberry.

$$p(y = B) \approx \frac{6}{10}$$

What is the probability to pick a blueberry from X?

Now we are going to calculate the joint probability **p(s=X, y=B)** which tell us the likelihood of picking a blueberry from **X**.

> **Note**: Joint probability is the probability of event Nr.1 occurring at the same time event Nr. 2 occurs. In this case one event is picking from the bowl that happens to be **X**. The other event is the fact that we have picked a blueberry.

In order to calculate the joint probability, we need to divide the number of times we picked a blueberry from *X* by the total number of picks:

$$p(s = X, y = B) = \frac{n(s = X, y = B)}{N}$$

$$= \frac{148}{300} \approx \frac{1}{2}$$

Eq. 5 Probably to pick a blueberry from **X**.

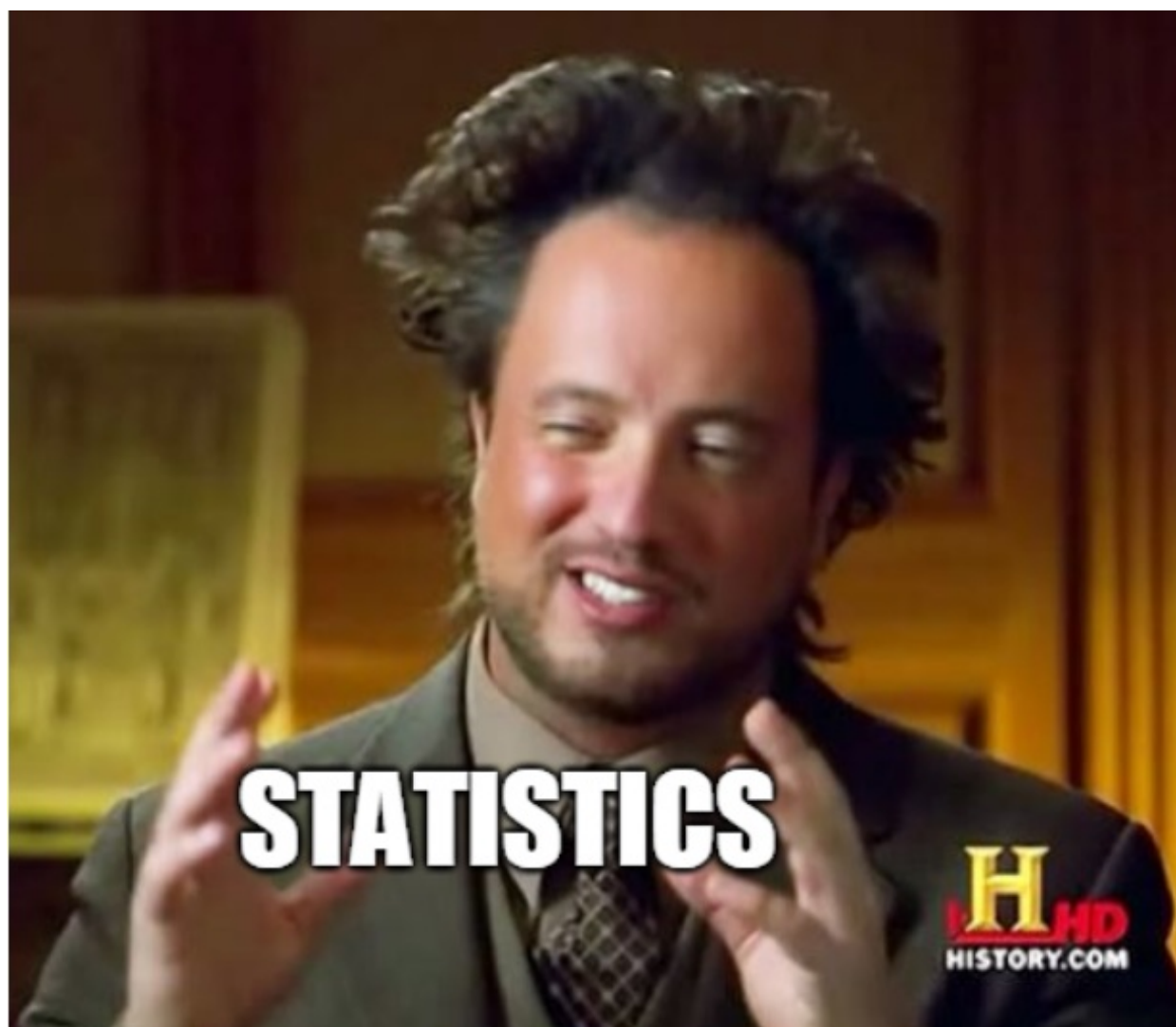Accordingly, the probability for picking a blueberry from **Y** is:

$$p(s = Y, y = B) = \frac{n(s = Y, y = B)}{N}$$

$$= \frac{26}{300} \approx \frac{1}{12}$$

Eq. 6 Probably to pick a blueberry from **Y**.

And the probability to pick an orange from **X** is:

$$p(s = X, y = O) = \frac{n(s = X, y = O)}{N}$$

$$= \frac{51}{300} \approx \frac{1}{6}$$

Eq. 7 Probably to pick an orange from **X**.

Given that we have picked from X, what is the probability that it is a blueberry?

Now it gets interesting. We calculate the first conditional probability. In this case, we know for sure which bowl we pick from. In this case, let's say, that we pick from **X**. Given this knowledge, we can calculate the probability that tells us the likelihood to pick a blueberry.

This conditional probability is denoted as **p(y=B| s=X)**, **s=X** being the condition that we pick the item from **X**. To calculate **p(y=B| s=X)** we need to divide the number of times we have picked blueberries from *X* by the total number of items picked from *X*:

$$p(y = B | s = X) = \frac{n(s = X, y = B)}{n(s = X, y = B) + n(s = X, y = O)}$$

$$= \frac{148}{200} \approx \frac{3}{4}$$

Eq. 8 Probably to pick a blueberry, given that we pick from **X**.

Product Rule

It is time for the first important statistical rule. Here we take the previously derived

probability to pick blueberries from **X p(s=X, y=B)** and extend this equation by multiplying it with (**n(s=X, y=B)+n(s=X, y=O))** in the denominator and numerator. We can do this because the value of the probability **p(s=X, y=B)** will not be changed by this extension.

Now if you take a closer look on the equation, you will notice that the new expression for **p(s=X, y=B)** consists of the product between two other probabilities **p(y=B|s=X)** and **p(s=X)** that we have derived earlier.

$$p(s=X, y=B) = \frac{n(s=X, y=B)}{N} \cdot \frac{n(s=X, y=B) + n(s=X, y=O)}{n(s=X, y=B) + n(s=X, y=O)}$$

$$= p(y=B|s=X) \cdot p(s=X)$$

$$\approx \frac{3}{4} \cdot \frac{2}{3} \approx \frac{1}{2}$$

Eq. 9 Product rule

This relation between probabilities is called the product rule. The product rule allows us to calculate the joint probability **p(s=X, y=B)** by using the conditional probability **p(y=B| s=X)** and the prior probability **p(s=X).**

Sum Rule

Now let's revisit the prior probability **p(s=X)** which gives us the likelihood to pick any item from **X**. If you divide the equation into two summands as can be seen in the second line in Eq. 10 you can observe that these two summands are nothing else than two joint probabilities that we have derived earlier.

$$p(s=X) = \frac{n(s=X, y=B) + n(s=X, y=O)}{N}$$

$$= \frac{n(s=X, y=B)}{N} + \frac{n(s=X, y=O)}{N}$$

$$= p(s=X, y=B) + p(s=X, y=O)$$

$$= \frac{1}{2} + \frac{1}{6} \approx \frac{2}{3}$$

Eq. 10 Sum rule

This relation is called the sum rule. The sum rule allows calculating the prior **p(X)** probability by doing the summation of joint probabilities that contain the random variable **s=X** from the prior and any other random variable **y**.

The Bayes Rule

For the product rule, the order of the random variables in the joint does not matter. Hence **p(s,y)** and **p(y,s)** have the same value.

If we equate the values **p(s, y)** and **p(y, s)** and do some reorganizing we get a new mathematical expression for **p(s|y)**. This new expression of **p(s|y)** is the Bayes Rule.

Eq. 12 Bayes'Theorem/Rule.

$$p(s, y) = p(y|s)p(s)$$

$$p(y, s) = p(s|y)p(y)$$

$$\rightarrow p(s, y) = p(y, s)$$

$$p(s, y) = p(y, s)$$

$$p(y|s)p(s) = p(s|y)p(y)$$

$$\rightarrow p(s|y) = \frac{p(y|s)p(s)}{p(y)}$$

Finally: Which bowl was the blueberry taken from?

The Bayes' Theorem provides us with the formula for calculation of the conditional probability **p(s|y),** which is the answer to our initial question.

The fact that we have picked a blueberry can be represented by the condition **y=B**. To answer the question which bowl was the blueberry picked from we must calculate **p(s|y=B)** for **s=X** and **s=Y**. Both values of **p(s|y)** tell us the likelihood that the blueberry was picked either from bowl **X** or bowl **Y**.

Let's do the calculation for **s=X**. Fortunately, all the probabilities that we need, we have already calculated in the previous sections. If we insert these probabilities into **p(s=X|y=B)** in Eq. 13 we come to the following conclusion: **Given that we have picked a blueberry, the probability that this blueberry was picked from bowl X is approximately 85 %.** The calculation can be done analogously for any other case.

$$p(s = X|y = B) = \frac{p(s = X, y = B)p(s = X)}{p(y = B)}$$

$$= \frac{p(s = X, y = B)p(s = X)}{p(s = X, y = B) + p(y = B, s = Y)}$$

$$= \frac{\frac{3}{4} \cdot \frac{2}{3}}{\frac{1}{12} \cdot \frac{1}{2}} \approx 85\%$$

Eq. 13 Bayes'Theorem

Without the Bayes'Theorem the calculation of **p(s|y)** would very difficult. The theorem, however, allows us to calculate this probability using probabilities that can be calculated with much less effort. This is the magic of the Bayes' Theorem: **A hard-to-compute probability distribution is represented by probabilities that are very easy to calculate.**