Reading: Calculate conditional probability with Bayes's theorem

Recently, you learned that **Bayes's theorem** is a math formula for determining conditional probability. The theorem is named after Thomas Bayes, an 18th-century mathematician from London, England. Recall that **conditional probability** refers to the probability of an event occurring given that another event has already occurred. For example, when you draw an ace from a deck of playing cards, this changes the probability of drawing a second ace from the same deck.

In this reading, you'll learn more about the different parts of Bayes's theorem, and how you can use the theorem to calculate conditional probability.

Bayes's theorem

Bayes's theorem provides a way to update the probability of an event based on new information about the event.

Posterior and prior probability

In Bayesian statistics, **prior probability** refers to the probability of an event before new data is collected. **Posterior probability** is the updated probability of an event based on new data.

Bayes's theorem lets you calculate posterior probability by updating the prior probability based on your data.

For example, let's say a medical condition is related to age. You can use Bayes's theorem to more accurately determine the probability that a person has the condition based on age. The prior probability would be the probability of a person having the condition. The posterior, or updated, probability would be the probability of a person having the condition if they are in a certain age group.

The theorem

Let's examine the theorem itself.

Bayes's theorem states that for any two events A and B, the probability of A given B equals the probability of A multiplied by the probability of B given A divided by the probability of B.

Bayes's theorem

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

In the theorem, prior probability is the probability of event A. Posterior probability, or what you’re trying to calculate, is the probability of event A given event B.

- **P(A)**: Prior probability
- **P(A|B)**: Posterior probability

Sometimes, statisticians and data professionals use the term “likelihood” to refer to the probability of event B given event A, and the term “evidence” to refer to the probability of event B.

- **P(B|A)**: Likelihood
- **P(B)**: Evidence

Using these terms, you can restate Bayes’s theorem as:

- Posterior = Likelihood * Prior / Evidence

$$\overset{\text{posterior}}{\downarrow \atop P(A|B)} = \frac{\overset{\text{likelihood}}{\downarrow \atop P(B|A)} * \overset{\text{prior}}{\downarrow \atop P(A)}}{\underset{\text{evidence}}{P(B) \atop \uparrow}}$$

It can be helpful to think about the calculation from these different perspectives and help to map your problem onto the equation.

One way to think about Bayes’s theorem is that it lets you transform a prior belief, P(A), into a posterior probability, P(A|B), using new data. The new data are the likelihood, P(B|A), and the evidence, P(B).

Note*: This reading provides an introduction to the basic concepts and terms associated with Bayes’s theorem. A detailed examination of Bayesian statistics is beyond the scope of this course. As you progress in your career as a data professional, you’ll have the opportunity to further explore Bayes’s theorem and its various applications.*

For now, a key point to remember is that Bayes's theorem includes both the conditional probability of B given A and the conditional probability of A given B. If you know one of these probabilities, Bayes’s theorem can help you determine the other.

Let's explore an example to get a better understanding of how the theorem works.

Example: spam filter

A well-known application of Bayes's theorem in the digital world is spam filtering, or predicting whether an email is spam or not. In practice, a sophisticated spam filter deals with many different variables, including the content of the email, its title, whether it has an attachment, the domain type of the sender address (.edu or .org), and more. However, we can use a simplified version of a Bayesian spam filter for our example.

Let's say you want to determine the probability that an email is spam given a specific word appears in the email. For this example, let's use the word "money."

You discover the following information:

- The probability of an email being spam is 20%.
- The probability that the word "money" appears in an email is 15%.
- The probability that the word "money" appears in a spam email is 40%.

In this example, your prior probability is the probability of an email being spam. Your posterior probability, or what you ultimately want to find out, is the probability that an email is spam given that it contains the word "money." The new data you will use to update your prior probability is the probability that the word "money" appears in an email and the probability that the word "money" appears in a spam email.

When you work with Bayes's theorem, it's helpful to first figure out what event A is and what event B is—this makes it easier to understand the relationship between events and use the formula.

Let's call event A a spam email and event B the appearance of the word "money" in an email. Now, you can re-write Bayes's theorem using the word "spam" for event A and the word "money" for event B.

P(A|B) = P(B|A) * P(A) / P(B)

P (Spam | Money) = P(Money | Spam) * P(Spam) / P(Money)

You want to find out the following:

- **P(Spam | Money), or posterior probability**: the probability that an email is spam given that the word "money" appears in the email

Now, enter your data into the formula:

- **P(Spam), or prior probability**: the probability of an email being spam = 0.2, or 20%
- **P(Money), or evidence**: the probability that the word “money” appears in an email = 0.15, or 15%
- **P(Money | Spam), or likelihood**: the probability that the word “money” appears in an email given that the email is spam = 0.4, or 40%

P (Spam | Money) = P(Money | Spam) * P(Spam) / P(Money) = 0.4 * 0.2 / 0.15 = 0.53333, or about 53.3%.

So, the probability that an email is spam given that the email contains the word “money” is 53.3%.

Key takeaways

Bayes's theorem is the foundation for the field of Bayesian statistics, also known as Bayesian inference, which is a powerful method for analyzing and interpreting data in modern data analytics. Data professionals use Bayes’s theorem in a wide variety of fields, from artificial intelligence to medical testing.

Having a basic understanding of Bayes’s theorem will enable you to learn more about Bayesian statistics as you advance in your career as a data professional.

Resources for more information

To learn more about Bayes’s Theorem, refer to the following resource:

- Bayes' theorem explained by Pennsylvania State University

For an interesting discussion of the "prosecutor's fallacy," check out this page:

- Explanation of the prosecutor's fallacy by the American Journal of Epidemiology

---