Bayes' theorem

Chrysafis Vogiatzis

Lecture 4

Learning objectives

After this lecture, we will be able to:

- Recall and explain the law of total probability.
- Use the law of total probability to calculate probabilities.
- Formulate Bayes' theorem.
- Describe Bayes' theorem.
- Explain what Bayes' theorem implies for probabilities.
- Apply Bayes' theorem in calculating probabilities.

Motivation: The Mantoux test

The Mantoux (sometimes called the Mendel–Mantoux) test is a diagnostic tool for tuberculosis (TB). In the test, a dosage of tuberculin units is injected: some time later, the reaction on the skin is measured and a positive or negative reaction is given. It is assumed that about 0.05% of the children in the world have TB. The test is pretty accurate, with 99% success rate – that is a person with TB receives a positive result 99% of the time, and a person without TB receives a negative result 99% of the time.

The Mantoux test is mandatory in most European countries schools. A random kid did the test, which came up positive. Are you 99% certain the kid has TB?

Motivation: Pilot season

Studios typically make decisions on shows based on a single episode made early on, called a "pilot". This pilot episode is viewed by a carefully selected audience who then reports either favorable or unfavorable reviews. A show is considered highly successful, moderately successful, or unsuccessful depending on its performance while on air.

Historically, 95% of highly successful shows received favorable reviews, 50% of moderately successful shows received favorable reviews, and 10% of unsuccessful shows received favorable reviews.

You are one of the producers of a new TV show, and are showing the pilot episode to a major studio. The audience loved it and gave generally favorable reviews. Will your show definitely be a huge success?

The law of total probability

Motivation

The Spring 2020 semester saw a rapid change of plans for most university courses due to the pandemic. Students were left needing to make a decision about selecting credit or no credit for their classes. Let's focus on one particular case.

Credit/no credit or graded?

A UIUC course requires students to end up with an average of 70 or above to qualify for credit, whereas an average of 60 is enough to qualify for a passing grade. A student believes they will end up with a score between 65 and 80 – so they are definitely passing the class – but they are thinking of opting for the credit/no credit option. Unfortunately, the class is missing two important grades: the final project and a final (non-cumulative) exam.

We are commonly facing problems like this in every day life. Decision-making under uncertainty revolves around us making decisions where the outcomes are not guaranteed. In such cases, the decision-maker *weighs the different futures* and aims to quantify the probability of a favorable outcome. Let's revisit the student from the example.

Credit/no credit or graded?

The student has been quite enjoying the material of the final exam and they are optimistic that they will score very highly. They believe that they will end up with a score of 90/100 with probability 50% or a score of 80/100 with probability 50%. They are not as confident for the final project, where they believe they received either a 50/100, a 60/100, or a 70/100 (with probability 30%, 60%, and 10%, respectively). So, say, they go ahead and put all these eventualities in a table.

A table can be used to keep track of all of the events whose outcomes are uncertain. In the case of the student, they would have to be enumerate a total of 6 cases (why? [1]), which are presented next.

[1] Remember counting!

Credit/no credit or graded?

	Final project	Final exam	Final grade
Scenario 1	50	80	65
Scenario 2	50	90	68
Scenario 3	60	80	69
Scenario 4	60	90	72
Scenario 5	70	80	73
Scenario 6	70	90	76

If the student picks credit/no credit, what is the probability they do not receive credit?

Derivation

Consider two events A and B, marked below as the blue area of the rectangle and the circle in the middle, respectively. We also mark the complement of A in the figure.

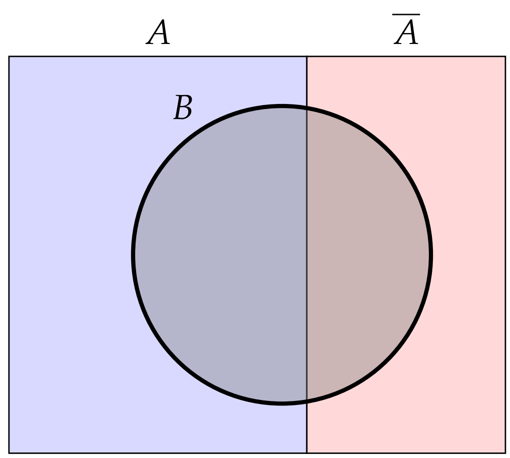

Figure 1: Two events A and B. A is represented by the blue area, whereas B contains all outcomes in the circle.

Say, we are interested in the probability of B happening. We can present this as a function of A as follows:

From the second figure, we observe that B can be written as a union of two mutually exclusive events as in

$$B = (B \cap A) \cup \left(B \cup \overline{A}\right) \implies$$
$$P(B) = P\left(B \cap A\right) + P\left(B \cap \overline{A}\right). \tag{1}$$

Finally, we recall here that $P(A \cap B) = P(A) \cdot P(B|A)$ [2], which we can replace in (1) to get:

$$P(B) = P(A) \cdot P(B|A) + P(\overline{A}) \cdot P(B|\overline{A}). \tag{2}$$

[2] The multiplication rule we saw during our previous lecture.

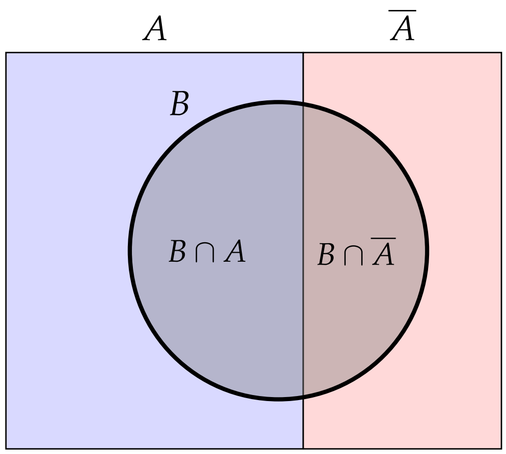

Figure 2: Marking the two (mutually exclusive) parts of B. It is true that $B = (B \cap A) \cup (B \cap \overline{A})$.

This is the law of total probability for two events.

> Two urns contain red and blue balls. The first urn contains 3 red and 3 blue balls, while the second urn contains 5 red and 2 blue balls. We pick one ball at random from the first urn and (without seeing its color) place it in the second urn. What is the probability we pick a blue ball from the second urn?

Use and interpretation

The law of totally probability can be generalized to more than 2 states. Assume we have m mutually exclusive and *collectively exhaustive* events. With the term *collectively exhaustive* we mean events whose union is the whole sample space. Formally:

Definition 1 (Collectively exhaustive events) *Let S be the sample space, and let $A_i, i = 1, \ldots, m$ be some events. Then, events A_i are collectively exhaustive if $\cup_{i=1}^{m} A_i = S$.*

Definition 2 (Mutually exclusive and collectively exhaustive events) *Let S be the sample space, and let $A_i, i = 1, \ldots, m$ be some events. Then, events A_i are mutually exclusive and collectively exhaustive if $\cup_{i=1}^{m} A_i = S$ and $A_i \cap A_j = \emptyset$ for any two sets $A_i, A_j, i \neq j$.*

An example of a series of mutually exclusive and collectively exhaustive events is given in Figure 3, where $S = A_1 \cup A_2 \cup A_3 \cup A_4$ and, as is shown, $A_1 \cap A_2 = A_1 \cap A_3 = A_1 \cap A_4 = A_2 \cap A_3 = A_2 \cap A_4 = A_3 \cap A_4 = \emptyset$.

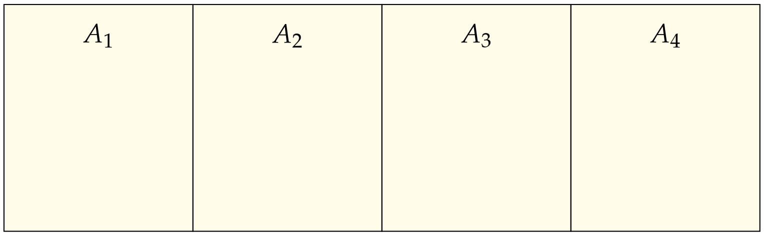

Figure 3: Four collectively exhaustive and mutually exclusive events. For example, they could represent numbers of unique website visitors in a given day. A_1 could then be up to 1000 visitors, A_2 could represent between 1001 and 3000 visitors, A_3 between 3001 and 5000 visitors, and A_4 5001 or more visitors.

Credit or no credit?

In the student example, the final project can be viewed as three collectively exhaustive and mutually exclusive events, since the student can only have received a score of 50, 60, or 70. On the other hand, the final exam score has two collectively exhaustive and mutually exclusive events (a score of 80 or 90).

In the case of $m > 2$ mutually exclusive and collectively exhaustive events, the law of total probability becomes:

$$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \ldots + P(A_m) \cdot P(B|A_m) = \\ = \sum_{i=1}^{m} P(A_i) \cdot P(B|A_i). \quad (3)$$

Credit or no credit?

Let's go back to the table of scenarios the student had prepared. Let us rewrite the eventualities:

- If they receive a 50 in the final project, then they definitely do not get credit.
- If they receive a 60 in the final project, then they have a 50% of getting credit.
- If they receive a 70 in the final project, then they definitely get credit.

Let A_1, A_2, A_3 be the events of getting a 50, 60, or 70 in the final project and let C be the event of receiving credit in the class. Then, in probability terms, we have:

$$P(C) = P(A_1) \cdot P(C|A_1) + P(A_2) \cdot P(C|A_2) + P(A_3) \cdot P(C|A_3) = \\ = 0.3 \cdot 0 + 0.6 \cdot 0.5 + 0.1 \cdot 1 = \\ = 0.4.$$

> You have just booked a two-leg (two-flight) trip. The flights are very close to one another and you are worried you will miss your second flight in the case of a delay. If the first flight is not delayed (which happens 60% of the time), you will be certainly fine (won't miss the flight); if the first flight is delayed up to 30 minutes (which happens 20% of the time), you might miss the second flight with probability 50%; finally, if the first flight is delayed by more than 30 minutes (which happens 20% of the time), you will definitely miss the second flight. What is the probability you miss the second flight?

Bayes' theorem

States of the world

Consider the following paradigm. You wake up in the morning, and unbeknownst to you the worlds is at a certain state. Let's call this state "good" or "bad". In a "good" world, 90% of everyone you talk to is happy and smiling and welcoming. In a "bad" world, only 5% of the people you talk to are happy and smiling and welcoming. Unfortunately, you have no idea which state the world is in today. What could you do to find out?

The above paradigm, as far-fetched as it sounds, applies in multiple aspects of our life. A student could have studied and can answer a multiple choice question correctly, or could have gotten lucky and could give the correct answer by chance. A diagnostic test could come back positive, and this could mean that the patient is indeed positive, or it could be a mistake (referred to as a false positive). Even worse, a diagnostic test could come back negative, when the patient is unfortunately positive (this is called a false negative). In all the above cases, there is a "state of the world" that we are querying through tests, whose outcomes we read.

States versus outcomes

We contrast states to outcomes as follows. We consider that a state is fleeting and unknowable; hence, we perform a test and make an observation of its outcome. However, the outcome of the test does not necessarily reveal the state, as no test is perfect.

The Mantoux test

In the Motivation section, we saw the Mantoux test. The states of the world (unknowable for certain) are whether a kid has TB or not. The test here is the Mantoux test. Its outcomes are positive or negative.

We state two key observations:

- **The test outcome is not equivalent to the state**. A positive Mantoux test does not always mean a person with TB. A low score in a test does not always mean a student who did not study. A good review does not necessarily mean you will like a movie.
- **Looking for something rate, we will encounter many false positives**. Think of what happens when searching in a vast desert for an oasis. Most times, the oasis is a mirage.

A two-state example

We present a two-state example, adapted by Daniel Kahneman's "Thinking, Fast and Slow" book.

Farmer or librarian?

You sit next to someone in a flight and you start talking. The person tells you that they are from the USA, they discuss with you how much they enjoy reading books in their free time, and that they enjoy learning about other cultures. They then ask you: "We've been talking for a while. Guess what my occupation is. Do you think I work as a librarian or as a farmer?"

Our mind can create some connections based on what we know and what we *think we know*. We know that there are more farmers than librarians in the USA (roughly 3 million farmers compared to 300,000 librarians). We also *think we know* that librarians probably enjoy books and learning about other cultures. We may jump to a conclusion, but if we do the math, we will see that our mind can rely too much on prior beliefs rather than context.

In summary: we formulate a hypothesis (for the state of the world) and, then, given evidence (outcomes of a test) that we leverage, we check to see if we are right or wrong.

Definitions

Let's collect here some definitions and notations that will be useful throughout the derivation of *Bayes' theorem*:

- $S_i, i = 1, \ldots, n$: n states of the world, which are mutually exclusive and collectively exhaustive.
- $O_j, j = 1, \ldots, m$: m outcomes of a test we administer trying to understand the true state of the world.

We also need to define our beliefs for what the state of the world is. These are called *prior probabilities*, as they reflect prior beliefs and biases [3] (before we see the outcomes of the test):

- $P(S_i)$: prior probability of state S_i.

Additionally, we define *likelihood probabilities* that represent the probability we see a certain outcome of the test when the world is in a certain state [4]:

- $P(O_j|S_i)$: likelihood probability of seeing outcome O_j given that we are in state S_i.

Moreover, we have already established that $P(O_j \cap S_i)$ is the probability that we both experience outcome O_j and we are in state S_i. This is called a *joint probability*:

- $P(O_j|S_i)$: joint probability of both seeing outcome O_j and we are in state S_i.

Finally, we may calculate the probability that a random test returns a certain outcome O_j. This is referred to as a *marginal probability* [5]:

- $P(O_j)$: marginal probability of seeing outcome O_j.

Don't lose sight of what we are searching for! This would be $P(S_i|O_j)$: the *posterior probability* of being in state S_i given that we observed outcome O_j in the test.

We may now proceed to the derivation of Bayes' theorem.

Derivation and use

We break down the derivation in steps.

[3] Examples include the probability that a random student has studied or not, or the probability that a random person is a librarian or a farmer.

[4] Examples include the probability that a student does well in an exam given that they have studied or that a person works as a librarian given that they enjoy reading books

[5] An example would be the probability of an "A" in an exam, or the probability of a positive Mantoux test.

Derivation

This will be filled in after the lecture.

Bayes' theorem states that the posterior probability $P(S_i|O_j)$ depends on our prior probabilities $P(S_i)$ and our joint probabilities $P(O_j|S_i)$.

$$\boxed{P(S_i|O_j) = \frac{P(S_i) \cdot P(O_j|S_i)}{\sum\limits_{i=1}^{n} P(S_i) \cdot P(O_j|S_i)}.}$$

For a visual representation of Bayes' theorem, check Figure 4.

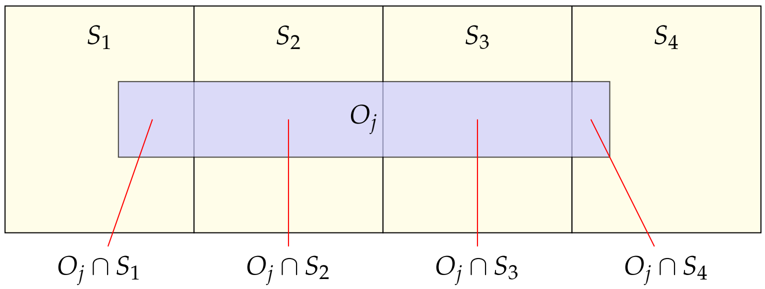

Figure 4: Consider 4 mutually exclusive and collectively exhaustive states S_1, S_2, S_3, S_4. When outcome O_j happens, note how the probabilities of each state change. For example, given O_j, we have $P(S_1|O_j) = \frac{P(O_j \cap S_1)}{P(O_j)}$. Replacing $P(O_j \cap S_1)$ by $P(S_1) \cdot P(O_j|S_1)$ and $P(O_j)$ by $P(S_1) \cdot P(O_j|S_1) + P(S_2) \cdot P(O_j|S_2) + P(S_3) \cdot P(O_j|S_3) + P(S_4) \cdot P(O_j|S_4)$ gives the result.

The Mantoux test

Going back to the Mantoux test, let's fill in the information we need to answer the question: "what is the probability a kid has TB given that the test came back positive?"

- S_1: kid has TB; S_2: kid does not have TB.
- O_1: positive Mantoux test; O_2: negative Mantoux test.
- $P(S_1) = 0.0005$; $P(S_2) = 0.9995$.
- $P(O_1|S_1) = 0.99$; $P(O_1|S_2) = 0.01$; $P(O_2|S_1) = 0.01$; $P(O_2|S_2) = 0.99$.

Using the Bayes' theorem, we have:

$$P(S_1|O_1) = \frac{P(S_1) \cdot P(O_1|S_1)}{P(S_1) \cdot P(O_1|S_1) + P(S_2) \cdot P(O_1|S_2)} = $$
$$= \frac{0.0005 \cdot 0.99}{0.0005 \cdot 0.99 + 0.9995 \cdot 0.01} = 0.0472.$$

We deduce that a positive Mantoux test implies a 4.72% chance of actually having TB.

Answer the probability question in the pilot episode example of the motivation.

Another visual representation of Bayes' theorem

Assume that we get a representative population from two states: Kansas and New York. Seeing as Kansas is about 9 times smaller than New York, we pick 90 people from New York and 10 from Kansas, represented pictorially in Figure **??**.

Statistically 20% of the population of New York works in an agriculture-related job. The same percentage is 80% for Kansas. Without loss of generality, we show that with the following Figure where

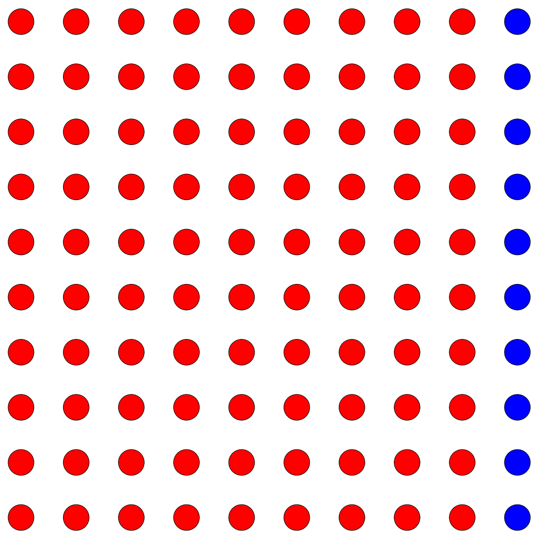

Figure 5: A visual representation of the population picked. In red, we have the residents of the state of New York; in blue the residents of the state of Kansas.

farmers are shaded in green (light green for New York, dark green for Kansas).

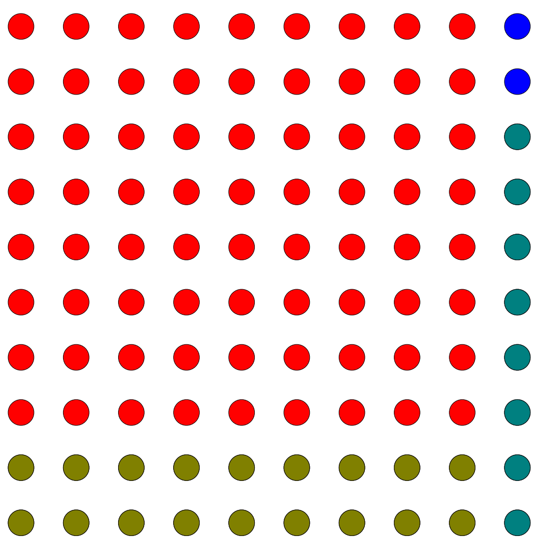

Figure 6: A visual representation of 100 registered voters. The voters in shades of green (darker green for red, lighter green for blue) are voters from Party A and Party B that overwhelmingly agree with a given statement.

Finally, assume the person next to you is flying from Kansas to New York (and has made it clear that they are either from Kansas or New York) and works in a farm. While your original bias may be that the person has to be in Kansas (look at the percentage of agriculture-related jobs for the Kansas population!), Bayes' theorem states that the probability is only $8/26 = 0.31$. Formally:

$$P(Kansas|farmer) = \frac{P(Kansas \cap farmer)}{P(farmer)} = \frac{0.1 \cdot 0.8}{0.1 \cdot 0.8 + 0.9 \cdot 0.2} = 0.31.$$