Conditional Probability and Bayes' Theorem

Abstract

In this session, we shall explore topics that are fundamental in probability: *conditional probability* and *Bayes' Theorem*. Probability, in general, explores and quantifies how likely certain events are, typically expressed as either a number between 0 and 1, or as a percentage between 0% and 100%. For conditional probability, we ask how likely an event A is *given* that event B has already happened, denoted $P(A \mid B)$.

Bayes' Theorem then allows us to compute conditional probabilities more easily, especially when direct computations would be unavailable. The applications of Bayes' Theorem are too numerous to mention, including everything from medical testing and diagnostics, weather forecasting, genetics, election predictions, criminal justice, online recommendation algorithms (like those by social media, streaming media, and online retailers), spam filtering, facial recognition, and artificial intelligence.

Background needed: We assume familiarity with very basic principles from probability. Our discussion of conditional probability and Bayes' Theorem will otherwise be mostly self-contained.

0 Warmup

Exercises:

0.1 Consider a standard six-sided die, where each of the six faces of the cube has precisely one of 1, 2, 3, 4, 5, or 6 dots on it. Further, assume the die is *fair*, in the sense that the probability of each side turning face up is $1/6$.

(a) Let A denote the event "the number of dots turns up on the next roll is a multiple of 3". What is $P(A)$, the probability that A happens?

(b) Let B denote the event "a prime number[1] of dots turns up on the next roll". What is $P(B)$, the probability that B happens?

[1] *Note:* A *prime number*, for our purposes here, is a positive integer n such that the only positive integer divisor of n are 1 and n itself. For technical reasons, the number 1 is *not* a prime number.

(c) Let $P(A \cap B)$ denote the probability that both A and B hold simultaneously; that is, $P(A \cap B)$ denotes the probability that we roll a number that is both a multiple of 3 and simultaneously a prime number. What is $P(A \cap B)$?

(d) Let $P(A \mid B)$ denote the conditional probability that A happens, *given that* B has happened. Similarly, let $P(B \mid A)$ denote the conditional probability that B happens, *given that* A has happened. What is $P(A \mid B)$? What is $P(B \mid A)$? Can you explain this in the context of Exercise 0.1(a)–0.1(c)

1 Conditional Probability

We begin with an introduction to some of the relevant terms of probability and conditional probability. Our interest here is not a thorough, rigorous introduction to probability, but simply an explanation of the relevant terminology.

Definition 1.1. Assume we are conducting an experiment, which in principle can be repeated indefinitely. An *event* A is a set A of possible outcomes within the sample space of our experiment. The *probability of* A is denoted $P(A)$.

Example 1.2. Consider a fair six-sided die, and let A denote the event "a composite number turns up on the next roll". Then the only[2] such outcomes are $\{4,6\}$, so $P(A) = 2/6 = 1/3$.

Definition 1.3. Let A, B be events. Then the *conditional probability of* A *given* B, denoted $P(A \mid B)$ is the probability that A has happened, given that B has already happened. This value is defined to be

$$P(A \mid B) := \begin{cases} \dfrac{P(A \cap B)}{P(B)}, \text{ when } P(B) \neq 0 \\ \text{undefined, when } P(B) = 0, \end{cases} \tag{1.1}$$

where $P(A \cap B)$ denotes the probability that events A and B happen simultaneously.

[2] Recall that 1 is *not* prime, nor is it composite.

Exercises:

1.1 Consider a fair coin, so that for any given flip, the probability of heads is $1/2$, and the probability of tails is also $1/2$. Consider an experiment where we flip the coin three times, each independent of the other two.

(a) Let A denote the event that we have at least two heads among our three flips. What is $P(A)$?

(b) Let B denote the event that we have at least one head among our three flips. What is $P(B)$?

(c) What is $P(A \cap B)$? That is, what is the probability that after three flips, both A and B hold?

(d) What is $P(A \mid B)$? What is $P(B \mid A)$?

1.2 Misapplication of probability can result in miscarriages of justice.[3] The following example is taken from [4]. Imagine that in a very large city, law enforcement has iden-

[3] For a real-world example of *the Prosecutor's Fallacy*, consider the case of Sally Clark, a British woman who was mistakenly convicted of the murder of her two children.

tified a suspect for a crime. Law enforcement has determined that this crime was committed by a single person, not a group of accomplices.

Investigators obtained a fingerprint from the crime scene, and it in their test, it matches a particular person whose fingerprints were already on record. Their expert explains that it is very unlikely for the the fingerprint evidence from the crime scene to match the sample taken from the suspect. *On the basis of this fingerprint evidence alone*, the prosecutor concludes that the suspect is guilty of the crime.

Let G denote the event "selected person is guilty of this crime", and F denote the event "selected person's fingerprints match those from the crime scene."

(a) Express the above scenario in the language of conditional probability.

(b) Say that in the above scenario, the accuracy of the fingerprint results is high: only 1 in 100,000 will be a mismatch under this test. The crime takes place in a populous place like The Triangle, whose approximate population according to *Wikipedia* is 2,106,463. Further, assume that nobody else living in The Triangle has been excluded as a suspect, and the fingerprint test is the only evidence identifying the suspect. Explain why the prosecutor is premature in concluding that this evidence conclusively identifies the suspect as the guilty party.

1.3 Conditional probability has a number of enormously important consequences for public health. Consider the following excerpt from [5], considering statistic regarding COVID vaccination outcomes in late-2021:

> **As vaccination increases, a higher proportion of cases will appear in vaccinated people—and that's what should happen**
>
> [...]
>
> Let's work through some numbers. Assume, first, that vaccines are 60 percent effective at preventing symptomatic infections. (There's a lot of conflicting information about this, but the exact number doesn't affect this exercise much.) Vaccinated people are still less likely to get infected, but as

> their proportion of the community rises, so does the percentage of infections occurring among them. If 20 percent of people are fully vaccinated, they'll account for 9 percent of infections; meanwhile, the 80 percent of the population that's unvaccinated will account for 91 percent. Now flip that. If only 20 percent of people are unvaccinated, there will be fewer infections overall. But vaccinated people, who are now in the majority, will account for most of those infections—62 percent.
>
> That is why this particular statistic—the proportion of vaccinated people in a given outbreak—is so deeply misleading. "The better the vaccine uptake, the scarier this number will seem," wrote Lucy D'Agostino McGowan, a statistician at Wake Forest University. By extension, the safer communities become, the more it will seem like the sky is falling—if we continue focusing on the wrong statistics.
>
> "If you're trying to decide on getting vaccinated, you don't want to look at the percentage of sick people who were vaccinated," McGowan wrote. "You want to look at the percentage of people who were vaccinated and got sick."

Using the language of conditional probability, explain mathematically what the above is trying to express.

1.4 Continuing with a public health theme, imagine we have a medical test attempting to detect the presence of a particular disease. Let D denote the event "patient has the disease" and T denote the event "patient tests positive on a particular test". Further, let $\neg D$, read "not D", denote probability that the patient does not have the disease, and $\neg T$, read "not T", denote the probability that the patient does not test positive on a particular test.

For our purposes, we shall assume that having the disease is binary: either one has it or one does not. Further, we assume that the test results are also binary: the test returns either a positive result (indicating the patient has the disease) or a negative result (indicating the patient does not have the disease).

(a) The *sensitivity* or *true positive rate* is the probability that someone who has the disease will test positive on our test. Express the sensitivity as a conditional probability.

(b) The *specificity* or *true negative rate* is the probability that someone who does not have the disease will test negative on the test. Express the specificity as a conditional probability.

(c) The *precision* or *positive predictive value (PPV)* of a test is the probability that if a patient tests positive, then the patient has the disease. Express the precision as a condition probability.

(d) Consider the following hypothetical scenario:[4]

- A disease has prevalence $P(D) = 0.01 = 1\%$.
  That is, for a person randomly chosen, where we have no additional information (such as medical history, presence of symptoms, etc.), the likelihood that person has the disease is $0.01 = 1\%$.
- The sensitivity/true positive rate is $0.9 = 90\%$.
  That is, among those with the disease, the test accurately detects this 90% of the time.
- The specificity/true negative rate is $0.91 = 91\%$.
  That is, among those without the disease, the test accurately detects this 91% of the time.

Compute the precision/positive predictive value for our test. That is, given that a patient tests positive, what is the probability the patient has the disease? Is this surprising, given the sensitivity and specificity of the test?

Hint: Consider a representative sample population of 1,000, and imagine testing all of them. As a starting point, how many would have the disease? Of them, how many would test positive? What other data do we need in order to compute the precision of the test?

[4] *Note:* This demonstration of the Base Rate Fallacy, is taken from [2]—and it is not merely hypothetical. German psychologist Gerd Gigerenzer posed this as a multiple-choice question to a seminar of practicing gynecologists, asking them to compute the hypothetical probability that a patient has breast cancer. Over half the doctors estimated the probability was 9 in 10, and only a fifth provided the correct answer.

2 Bayes' Theorem

As motivated by Section 0, we can use these ideas to compute facilitate the computation of conditional probabilities.

Theorem 2.1 (Bayes'[5] Theorem)**.** *Let A, B be events such that $P(B) \neq 0$. Then*

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}. \tag{2.1}$$

One typically interprets Bayes' Theorem as *updating* our initial assessment of the probability of A based on incorporating the additional evidence provided by B being true. That is, we think of $P(A)$ as our *prior probability* before taking into account B. The conditional probability $P(A \mid B)$ is the *posterior probability* of A, now having B as additional evidence. (There are names for other elements of the formula in (2.1), but they are unimportant for our purposes.)

Also, note that in computing $P(A \mid B)$, the probability of A *given* B, our "given" probability $P(B)$ appears in the denominator of the right-hand side of (2.1). This may help keep straight where A and B appear, respectively, in the formula.

Exercises:

2.1 Recompute the conditional probabilities $P(A \mid B)$ and $P(B \mid A)$ from Exercise 1.1(d) using Bayes' Theorem. (You may already assume that $P(A)$, $P(B)$, and $P(B \mid A)$ are given.)

2.2 To compute $P(D \mid T)$ in Exercise 1.4(c) using Bayes' Rule, we must compute $P(T)$. Explain how to compute $P(T)$ from our given data.

[5] Named for Rev. *Thomas Bayes* (c. 1701–1761), an English philosopher, mathematician, and Presbyterian minister. The theorem itself was published posthumously in *An Essay towards solving a Problem in the Doctrine of Chances*, 1763.

2.3 Using Bayes' Theorem, recompute the precision/positive predictive value of the testing scenario in Exercise 1.4(d).

2.4 Draw a picture to represent Bayes' Theorem geometrically. What is $P(A)$? What is $P(B)$? What are the conditional probabilities $P(A \mid B)$ and $P(B \mid A)$?

Note: For an interactive tool, consider using "Geometric Bayes Theorem" from [3] at https://www.skobelevs.ie/BayesTheorem/.

2.5 For a given event A with probability $P(A)$, we can express $P(A)$ in terms of *odds* rather than probability. For example, an event with probability $\frac{4}{17}$ will have odds $4:(17-4) = 4:13$. An event with odds $p:q$ will have probability $\frac{p}{p+q}$.

Express Bayes' Theorem in terms of the odds of $O(A \mid B)$ expressed relative to the odds $O(A)$, $O(B)$, and $O(B \mid A)$.

References

[1] 3Blue1Brown. Bayes theorem, the geometry of changing beliefs. https://www.youtube.com/watch?v=HZGCoVF3YvM, December 22, 2019.

[2] 3Blue1Brown. The medical test paradox, and redesigning Bayes' rule. https://www.youtube.com/watch?v=lG4VkPoG3ko, December 22, 2020.

[3] Nikita Skobelevs. Geometric Bayes theorem. https://www.skobelevs.ie/BayesTheorem/, April 21, 2020.

[4] Brani Vidakovic. Handout 1. https://www2.isye.gatech.edu/isyebayes/bank/handout1.pdf, August 15, 2004. online: retrieved January 16, 2024.

[5] Katherine J. Wu, Ed Yong, and Sarah Zhang. Six rules that will define our second pandemic winter. *The Atlantic*, September 20, 2021. online: retrieved January 16, 2024.