Lecture 18: conditional probability, independence, and Bayes' theorem

November 18, 2026

Recall from last time: for events A and B,

$$P(A \cup B) = P(A) + P(B) \qquad \text{if } A \text{ and } B \text{ are disjoint,}$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad \text{in general.}$$

Another fact is that, for $\overline{A}$ the complement of A,

$$P(\overline{A}) = 1 - P(A).$$

Probability is useful for modeling uncertainty about the world. Sometimes, though, we observe something new about the world, and that should update our information. So we often want to ask questions of the form: if event A happened, what is the probability that event B happened?

- If my COVID test comes back positive, what is the probability that I have COVID?
- If I got a B on midterm 2, what is the probability that I pass the class?
- If a Democrat wins the senate seat in South Carolina, what are the odds the Democrats control a majority in the senate?
- If I miss the M60 bus to LGA, what is the probability that I miss my flight?

In each of these the two events — "have COVID" and "test positive," "miss the bus" and "miss the flight" — are related in some way.

1 Independence

Here are some toy examples, using the uniform spaces from last time. If I flip 10 coins, these two events are related: A = I flip at least 9 heads, B = I flip exactly 10 heads. (In order to flip 10, I need to flip at least 9).

If I roll two dice, these two events are related: A = I roll at least one 6, B = the dice sum to 10.

Some events are not related. If I flip 10 coins, the events $\{A$ = the first coin is H, B = the last 3 coins are all T$\}$ are unrelated.

In probability we call this idea of relatedness **dependence**.

Definition 1.1. *Two events A and B are **independent** if*

$$P(A \cap B) = P(A) \cdot P(B).$$

*Otherwise they are **dependent**.*

Let's check each of the three examples against the definition.

Example 1.2. I flip 10 coins. Are "at least 9 heads" and "exactly 10 heads" independent?

The space is uniform with $|\Omega| = 2^{10}$. Exactly one outcome has all 10 heads, and 10 outcomes have exactly 9 heads (one for each coin that could be the tail), so

$$P(\text{at least 9 heads}) = \frac{11}{2^{10}}, \qquad P(\text{exactly 10 heads}) = \frac{1}{2^{10}}.$$

Every outcome with exactly 10 heads has at least 9 heads, so the intersection of the two events is just "exactly 10 heads," giving $P(A \cap B) = 1/2^{10}$. Is

$$\frac{1}{2^{10}} \stackrel{?}{=} \frac{11}{2^{10}} \cdot \frac{1}{2^{10}}?$$

No. So these two events are dependent.

Example 1.3. I roll two dice. Are "at least one 6" and "the dice sum to 10" independent?

Here Ω is the 36 ordered pairs of rolls. There are 11 pairs with at least one 6 (six with a 6 first, six with a 6 second, minus the double-counted $(6,6)$), and three pairs summing to 10: $(4,6)$, $(5,5)$, $(6,4)$. So

$$P(\text{at least one 6}) = \frac{11}{36}, \qquad P(\text{sum is 10}) = \frac{3}{36}.$$

Two of those three pairs contain a 6, namely $(4,6)$ and $(6,4)$, so $P(A \cap B) = 2/36$. And

$$\frac{2}{36} \neq \frac{11}{36} \cdot \frac{3}{36},$$

so these are dependent too.

Example 1.4. I flip 10 coins. Are "the first coin is H" and "the last 3 coins are all T" independent?

The first coin is H in half of the outcomes, and the last three coins are all T in $1/2^3$ of them, so

$$P(\text{first is H}) = \frac{1}{2}, \qquad P(\text{last 3 are T}) = \frac{1}{8}.$$

An outcome in both events has its first coin and its last three coins all determined, leaving 6 free coins, so $P(A \cap B) = 2^6/2^{10} = 1/16$. And

$$\frac{1}{16} = \frac{1}{2} \cdot \frac{1}{8},$$

so these two events are independent.

2 Independence versus mutual exclusivity

Independence and disjointness — which we also call being **mutually exclusive** — are not the same thing. A pair of events can be one, the other, both, or neither.

Example 2.1. For each pair of events E and F, are they independent, mutually exclusive, both, or neither?

(a) Flip a coin. E = get H, F = get T.

(b) Flip 2 coins. E = the first is H, F = the second is T.

(c) Flip a coin. E = get H, F = get H.

(d) Roll two dice. E = the sum is 4, F = the first roll is 2.

(e) Roll two dice. E = the sum is 7, F = the first roll is 2.

(f) Roll two dice. E = roll a 6, F = roll a 3.

Example 2.2. Can you find a pair of events that is independent *and* mutually exclusive?

Take $E = \Omega$ and $F = \varnothing$. They are disjoint, and

$$P(E \cap F) = P(\varnothing) = 0 = 1 \cdot 0 = P(E) \cdot P(F),$$

so they are also independent.

3 Conditional probability

When events are dependent, knowing the outcome of one gives us information about the other. If I tell you that I rolled at least one 6, that increases the probability that the total is 10. If I tell you that I rolled at least one 1, that decreases the probability that the total is 10 — to 0, in fact.

Conditional probability is the notation for this.

Definition 3.1. *The **probability of** A **given** B is*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

So, given that I rolled at least one 6, the probability that the sum is 10 is

$$P(\text{sum is 10} \mid \text{at least one 6}) = \frac{P(\text{at least one 6} \ \cap \ \text{sum is 10})}{P(\text{at least one 6})} = \frac{2/36}{11/36} = \frac{2}{11}.$$

Compare that to the unconditional $P(\text{sum is 10}) = 3/36 = 1/12$. Learning about the 6 made the sum of 10 more likely.

Fact 3.2. *If A and B are independent, then $P(A \mid B) = P(A)$.*

Proof. Since A and B are independent, $P(A \cap B) = P(A)P(B)$. Then

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A). \qquad \square$$

That is: if the two events are independent, then learning B tells you nothing about A.

Fact 3.3. $P(A \cap B) = P(A \mid B) \cdot P(B)$.

Proof. Rearrange the definition of $P(A \mid B)$. □

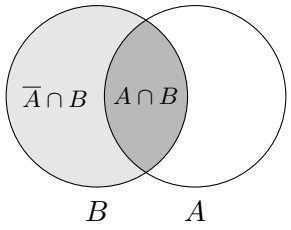
Fact 3.4. $P(\overline{A} \mid B) = 1 - P(A \mid B)$.

Proof. By definition,

$$P(\overline{A} \mid B) = \frac{P(\overline{A} \cap B)}{P(B)},$$

so let's rewrite the numerator. Every outcome of B either is in A or is not, so B splits into two disjoint pieces:

$$B = (\overline{A} \cap B) \ \cup \ (A \cap B).$$



Since the two pieces are disjoint, last lecture's rule for unions gives

$$P(B) = P(\overline{A} \cap B) + P(A \cap B), \qquad \text{so} \qquad P(\overline{A} \cap B) = P(B) - P(A \cap B).$$

Plugging that back in,

$$P(\overline{A} \mid B) = \frac{P(B) - P(A \cap B)}{P(B)} = 1 - \frac{P(A \cap B)}{P(B)} = 1 - P(A \mid B).$$ □

4 Covid testing

Here is a more realistic example. Suppose that

- 2% of people have Covid, and
- our Covid tests have a 10% false positive rate and a 5% false negative rate. That is, if you have Covid the test says "negative" 5% of the time, and if you do not have Covid it says "positive" 10% of the time.

Example 4.1. I take a test and it comes back positive. What is the probability that I have Covid?

Let A be the event that I have Covid and B the event that I test positive. The facts we are given are

$$P(A) = 0.02, \qquad P(\overline{A}) = 0.98, \qquad P(\overline{B} \mid A) = 0.05, \qquad P(B \mid \overline{A}) = 0.1,$$

where $P(\overline{B} \mid A)$ is $P(\text{negative} \mid \text{COVID})$ and $P(B \mid \overline{A})$ is $P(\text{positive} \mid \text{no COVID})$. What we want is

$$P(\text{COVID} \mid \text{positive}) = P(A \mid B) = \frac{P(A \cap B)}{P(B)},$$

so we need two things: $P(A \cap B)$ and $P(B)$.

What is $P(A \cap B)$? This is P(positive and COVID). Using Fact 3.3 and then Fact 3.4,

$$\begin{aligned} P(A \cap B) &= P(\text{positive} \mid \text{COVID}) \cdot P(\text{COVID}) && \text{(Fact 3.3)} \\ &= \big(1 - P(\text{negative} \mid \text{COVID})\big) \cdot P(\text{COVID}) && \text{(Fact 3.4)} \\ &= (1 - 0.05) \cdot 0.02 \\ &= 0.95 \cdot 0.02 = 0.019. \end{aligned}$$

What is $P(B)$? Every positive test either comes from someone with COVID or from someone without, so $B = (B \cap A) \cup (B \cap \overline{A})$, and those two pieces are disjoint. So

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap \overline{A}) && \text{(union of disjoint events)} \\ &= 0.019 + P(B \cap \overline{A}) && \text{(computed just above)} \\ &= 0.019 + P(B \mid \overline{A}) \cdot P(\overline{A}) && \text{(Fact 3.3)} \\ &= 0.019 + 0.1 \cdot 0.98 \\ &= 0.019 + 0.098 = 0.117. \end{aligned}$$

Putting them together.

$$P(\text{COVID} \mid \text{positive}) = \frac{0.019}{0.117} \approx 0.162.$$

So you just tested positive, and the chance you have COVID is only about 16%.

5 Bayes' theorem

Theorem 5.1 (Bayes' theorem)**.** $P(A \mid B) = \dfrac{P(B \mid A) \cdot P(A)}{P(B)}$.

Proof. By Fact 3.3, applied in each order,

$$P(B \mid A)P(A) = P(B \cap A) = P(A \cap B) = P(A \mid B)P(B).$$

Dividing both sides by $P(B)$ gives the theorem. □

This is, in theorem form, what we just did for the COVID test: we computed $P(A \cap B)$ as $P(B \mid A)P(A)$, and we computed $P(B)$ by splitting it into two cases.

Example 5.2. I have two coins. One is red on both sides; the other is blue on one side and red on the other. I pull out a random coin, look at one random side, and see red. What is the probability that the other side is blue?

$$\begin{aligned} P(\text{other side blue} \mid \text{this side red}) &= \frac{P(\text{this side red} \mid \text{other side blue}) \cdot P(\text{other side blue})}{P(\text{this side red})} \\ &= \frac{1 \cdot 1/4}{3/4} = \frac{1}{3}. \end{aligned}$$

6 The law of total probability

Theorem 6.1 (Law of total probability)**.** *If the events $E_1, E_2, \ldots$ partition Ω, then for any event F,*

$$P(F) = \sum_i P(F \mid E_i) \cdot P(E_i).$$

We most often use this to compute the denominator in Bayes' theorem — it is the general version of the case-splitting we did for $P(B)$ in the COVID example.

7 (Bonus) Monty Hall

You participate in a game show, where are three doors. I, the host, put a car behind one of the doors at random, and goats behind the other two. (The car is worth a bazillion dollars; the goats are worth negative money, because it costs money to take care of a goat.)

You pick a door. I then open a door that is *not* the one you picked and that has a goat behind it. You may now change your pick or keep it.

Example 7.1. Should you switch?

Partition on which door you picked first.

- With probability 1/3 you picked the car. I open a goat door, and switching loses.
- With probability 1/3 you picked goat 1. I have to open goat 2, and switching wins.
- With probability 1/3 you picked goat 2. I have to open goat 1, and switching wins.

The three cases are disjoint and cover everything, so switching wins with probability 2/3. You should always switch.