Data Analytics Workshop, JHU Pre-College Program
Day 6: Probability and Conditional Probability (optional notes)

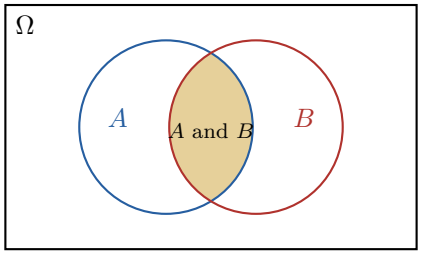
TL;DR Today is a more theoretical day, but it is what makes sense of everything else we do with data. Everything in this course has finitely many outcomes, so probability comes down to counting.

(Probability)

*The **sample space** Ω is the set of all possible outcomes, and an **event** A is a subset of Ω. If every outcome is equally likely,*

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\textit{number of outcomes in } A}{\textit{number of outcomes in } \Omega}.$$

- Events A and B are **independent** if $\mathbb{P}(A \text{ and } B) = \mathbb{P}(A)\,\mathbb{P}(B)$. Intuitively, one has no effect on the other (two coin flips). This **product rule** only works for independent events!
- The **conditional probability** of A given B is $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \text{ and } B)}{\mathbb{P}(B)}$. Knowing B happened shrinks the sample space down to B.
- **Bayes' theorem** flips a conditional probability around: $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(B \mid A)\,\mathbb{P}(A)}{\mathbb{P}(B)}$.
- In general $\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)$. Mixing these up has caused real damage, including in court cases.



Given B, only the red circle matters, and the event becomes the overlap

1 Probability and Independence

We have used the word probability a lot without saying what it is. Everything in this course involves only finitely many outcomes, so we can define probability by counting. I am going to introduce some new symbols, but bear with me; they are just symbols, nothing too scary.

Definition 1.1 (Sample space, event, probability)

*The **sample space** Ω is the set of all possible outcomes. An **event** A is a subset of Ω. If every outcome is equally likely, the **probability** of A is*

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},$$

the number of outcomes in A divided by the number of outcomes in Ω.

For a coin, $\Omega = \{H, T\}$. For a die, $\Omega = \{1, 2, 3, 4, 5, 6\}$. For a deck of cards, Ω is all 52 cards, and events could be "the card is a heart" or "the card is a three." Two notable probabilities are

$$\mathbb{P}(\Omega) = \frac{|\Omega|}{|\Omega|} = 1 \qquad \text{and} \qquad \mathbb{P}(\emptyset) = \frac{0}{|\Omega|} = 0,$$

where $\emptyset$ is the empty set (nothing happens). Since every event is somewhere between nothing and everything, every probability is between 0 and 1.

Definition 1.2 (Independence)

*Two events A and B are **independent** if*

$$\mathbb{P}(A \text{ and } B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Intuitively, independence means one event has no effect on the other. Flipping a coin today has no effect on flipping it again in thirty seconds, or in thirty years. The same is true at a slot machine: losing 100 times in a row does *not* make the next pull more likely to win, no matter how much it feels like it should. Let's check the definition against our intuition.

Example 1.1

Flip a coin twice. Let A be the event that the first flip is tails and B the event that the second flip is tails. The sample space is

$$\Omega = \{HH, HT, TH, TT\},$$

and only TT is in both A and B, so $\mathbb{P}(A \text{ and } B) = \frac{1}{4}$. Since $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$,

$$\mathbb{P}(A)\,\mathbb{P}(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = \mathbb{P}(A \text{ and } B),$$

and the events are independent, just as we expected.

Example 1.2

Flip a coin once. Let A be the event that the flip is tails and B the event that the same flip is heads. Then $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$, but no outcome is both heads and tails, so

$$\mathbb{P}(A \text{ and } B) = \mathbb{P}(\emptyset) = 0 \neq \frac{1}{4}.$$

These events are not independent. In fact they are as dependent as it gets: if A occurs, B cannot.

Example 1.3

Shuffle a deck, draw a card, and let A be the event that it is a club. Put the card back, shuffle again, draw, and let B be the event that this card is a club. There are 13 clubs, so $\mathbb{P}(A) = \mathbb{P}(B) = \frac{13}{52} = \frac{1}{4}$. There are 52^2 ways to draw two cards and 13^2 ways to draw two clubs, so

$$\mathbb{P}(A \text{ and } B) = \frac{13^2}{52^2} = \frac{1}{16} = \mathbb{P}(A)\,\mathbb{P}(B).$$

Drawing with replacement gives independent events, just like flipping a coin.

Now suppose we do not put the first card back. If the first card was a club, there are fewer clubs left, so the second draw is less likely to be a club. If it was not a club, the second draw is more likely to be one. So the first draw affects the second, and indeed

$$\mathbb{P}(A \text{ and } B) = \frac{13}{52} \cdot \frac{12}{51} = \frac{1}{17} \neq \frac{1}{16}.$$

Without replacement the events are not independent.

The **product rule** for independent events makes many probabilities easy to compute, since we can find each probability on its own and multiply.

Example 1.4

Suppose you pick an outfit from 3 shirts, 4 pairs of pants, and 2 pairs of shoes, each chosen independently at random. There are $3 \cdot 4 \cdot 2 = 24$ outfits. Rather than listing all 24, the probability of choosing shirt 1, pants 2 or 3, and shoe 1 is just

$$\frac{1}{3} \cdot \frac{2}{4} \cdot \frac{1}{2} = \frac{1}{12}.$$

The product rule only works for independent events. In Discussion 4 you read about real trials where an expert multiplied probabilities of events that were *not* independent, and got a number that was wildly wrong, with serious consequences.

2 Conditional Probability and Bayes' Theorem

Probability always depends on the conditions at play. Even independence is a condition. If I roll a die, hide it, and tell you the result is odd, the sample space shrinks from $\{1, 2, 3, 4, 5, 6\}$ to $\{1, 3, 5\}$,

and the probability that the result is 3 becomes $\frac{1}{3}$ instead of $\frac{1}{6}$. The information changed the probability.

Definition 2.1 (Conditional probability)

The probability of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \text{ and } B)}{\mathbb{P}(B)}.$$

Picture A and B as two overlapping circles inside Ω, like the first page, and think of probability as area. $\mathbb{P}(A)$ is the area of A compared to all of Ω. Knowing B is true means we only look inside B, so B becomes our new sample space and the only part of A we can land in is the overlap. That is exactly the formula: the overlap, compared to B.

Being in the overlap is the same as having A occur, and then having B occur given that A did. So $\mathbb{P}(A \text{ and } B) = \mathbb{P}(B \mid A)\,\mathbb{P}(A)$. Substituting this into the top of the definition gives the following.

Theorem 2.2 (Bayes' theorem)

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

This can look like it says nothing new, but often we know three of these four quantities and not the fourth, and Bayes' theorem hands us the last one. It also suggests a way to think about drawing conclusions from data: to make an inference, we must take into account prior information.

Example 2.1

Alice and Bob have two children, and each child is equally likely to be a boy or a girl. Given that one of them is a boy, what is the probability that the other is a boy?

This is not the same as asking about the next child after a first boy. That probability is $\frac{1}{2}$, since the first child has no effect on the second. Here both children are already born. The possibilities are

$$\{BB, BG, GB, GG\}.$$

Knowing one child is a boy removes GG, leaving $\{BB, BG, GB\}$. Only BB has the other child also a boy, so the answer is $\frac{1}{3}$, not $\frac{1}{2}$ as most people guess! With the formula: $\mathbb{P}(\text{both boys} \mid \text{at least one boy}) = \frac{1/4}{3/4} = \frac{1}{3}$.

Remark 1

In general, $\mathbb{P}(A \mid B)$ and $\mathbb{P}(B \mid A)$ are very different. The probability that someone speaks English, given that they are President of the United States, is essentially 1. The probability that someone is President of the United States, given that they speak English, is about one in a

billion. Mixing these two up in court is called the prosecutor's fallacy, and it shows up in the Discussion 4 readings. It is also exactly the p-value mistake from Day 5.

3 Practice

Problem 3.1

Roll two dice, one red and one blue. Let A be the event that the red die shows a 3.

1) *Let B be the event that the total is 7. Are A and B independent?*
2) *Let C be the event that the total is 8. Are A and C independent?*
3) *What is $\mathbb{P}(C \mid A)$?*

Solution:

There are 36 equally likely outcomes, and $\mathbb{P}(A) = \frac{6}{36} = \frac{1}{6}$.

1) $\mathbb{P}(B) = \frac{6}{36} = \frac{1}{6}$, and the only way to get both is red 3, blue 4, so $\mathbb{P}(A \text{ and } B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$. Independent! Whatever the red die shows, there is exactly one blue value that makes 7.
2) $\mathbb{P}(C) = \frac{5}{36}$, and $\mathbb{P}(A \text{ and } C) = \frac{1}{36}$ (red 3, blue 5). But $\frac{1}{6} \cdot \frac{5}{36} = \frac{5}{216} \neq \frac{1}{36}$, so these are *not* independent. (If the red die showed a 1, a total of 8 would be impossible.)
3) $\mathbb{P}(C \mid A) = \frac{1/36}{1/6} = \frac{1}{6}$, which is bigger than $\mathbb{P}(C) = \frac{5}{36}$. Knowing the red die is a 3 makes an 8 slightly more likely.