6.1200J/18.062J *Mathematics for Computer Science* Thursday 25$^{\text{th}}$ April, 2024
Massachusetts Institute of Technology, Spring 2024
Z. Abel, B. Chapman, E. Demaine
revised Sunday 28$^{\text{th}}$ April, 2024

Lecture 19: Conditional Probability

1 Probability Rules

In the last lecture, we learned how to compute probabilities using the Tree Method. In this lecture, we will see how some of our tools for reasoning about sizes of sets carry over naturally to the world of probability, and we will learn how to express mathematically statements like "if the prize is behind door A, what is the probability that Monty opens door B?"

Recall:

Definition 1. *We define the* probability of an event A *as*

$$\Pr[A] := \sum_{\omega \in A} \Pr[\omega]$$

An immediate consequence:

Proposition 1 (Sum Rule)**.** *If A and B are disjoint events, then*

$$\Pr[A \cup B] = \Pr[A] + \Pr[B]$$

Corollary 2 (Complement Rule)**.**

$$\Pr[\bar{A}] = 1 - \Pr[A]$$

Corollary 3 (Difference Rule)**.**

$$\Pr[A \setminus B] = \Pr[A] - \Pr[A \cap B]$$

Corollary 4 (Principle of Inclusion-Exclusion)**.**

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$$

Corollary 5 (Union Bound)**.**

$$\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$$

Corollary 6 (Monotonicity Rule)**.** *If $A \subseteq B$ are events, then*

$$\Pr[A] \leq \Pr[B]$$

The Sum Rule and Union Bound generalize:

Proposition 7 (Sum Rule)**.** *If A_i are (pairwise) disjoint events, then*

$$\Pr\left[\bigcup_{i\in\mathbb{N}} A_i\right] = \sum_{i\in\mathbb{N}} \Pr[A_i]$$

Proposition 8 (Union Bound)**.**

$$\Pr\left[\bigcup_{i\in\mathbb{N}} A_i\right] \leq \sum_{i\in\mathbb{N}} \Pr[A_i]$$

PIE generalizes to finitely many events in the same way as for counting:

Proposition 9 (Principle of Inclusion-Exclusion)**.** *If I is a finite index set, then*

$$\Pr\left[\bigcup_{i\in I} A_i\right] = -\sum_{\emptyset\subset J\subseteq I} (-1)^{|J|} \Pr\left[\bigcap_{j\in J} A_j\right]$$

2 Conditional Probability

In the last lecture, we studied the Monty Hall problem. In the analysis of this problem, it was useful to make statements such as "if the car is behind door 1, the contestant chooses door 1 with probability 1/3." How do we express this mathematically in the theory of probability?

Definition 2. *For two events A, B, the* conditional probability of A given B *is*

$$\Pr[A \mid B] = \frac{\Pr[A\cap B]}{\Pr[B]}.$$

This expression can be rewritten to obtain the "product rule" for joint probabilities:

Corollary 10 (Product Rule)**.**

$$\Pr[A\cap B] = \Pr[A \mid B]\Pr[B]$$

This can be extended to multiple events:

$$\Pr[A\cap B\cap C] = \Pr[A \mid B\cap C]\Pr[B\cap C] = \Pr[A \mid B\cap C]\Pr[B \mid C]\Pr[C].$$

The product rule is the justification for the "tree method" of computing probabilities from the last lecture: the numbers on the edges of the tree are the terms in the product. This means in particular that the numbers on the edges of the tree (except at the highest level) are conditional probabilities!

Another extension of the product rule that is useful is

$$\Pr[A\cap B \mid C] = \Pr[A \mid B\cap C]\Pr[B \mid C].$$

This can be obtained by dividing both sides of the previous product rule by $\Pr[C]$.

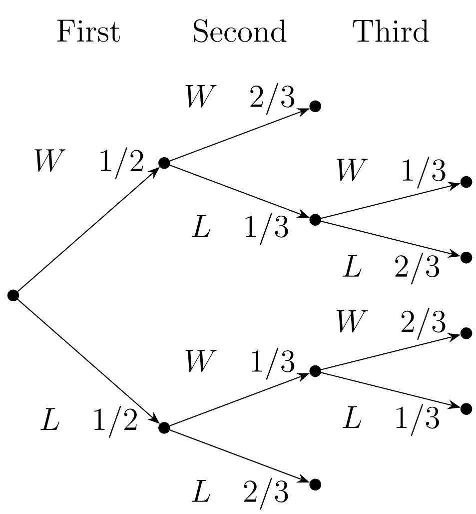
3 Example 1: tournament

Suppose Ash and Gary have a series of battles against each other, and the first to win two battles wins the series. The probabilities of victory have the following behavior:

1. The first battle is a toss-up: 1/2 probability of either trainer winning.
2. If a trainer has won the previous battle, he has a 2/3 chance of winning the next one.
3. There are no draws.

Let A be the event that Ash wins the series, and B the event that he wins the first battle. What is $\Pr[A \mid B]$?

To compute this, let's use the Tree Method.



ω	$\Pr[\omega]$	A?	B?
WW	1/3	×	×
WLW	1/18	×	×
WLL	1/9		×
LWW	1/9	×	
LWL	1/18		
LL	1/3		

$$\begin{aligned}\Pr[A \mid B] &= \frac{\Pr[A \cap B]}{\Pr[B]} \\ &= \frac{1/3 + 1/18}{1/2} \\ &= \frac{7}{9}.\end{aligned}$$

So far, this is just mechanical computation.

4 Bayes' rule

What about $\Pr[B \mid A]$? This can be calculated in exactly same manner to get 7/9. But what does it *mean*? Such a conditional probability is expressing an *inference*: what is the chance that Ash won the first battle, *given* that we later *observe* that he won the whole series?

It is often the case that we have a "model" that makes it easy to compute "forward" conditional probabilities $\Pr[A \mid B]$, but we would really like to know the "backward" probability $\Pr[B \mid A]$ in order to *infer* something that we do not have direct observational access to. In general $\Pr[B \mid A] \neq \Pr[A \mid B]$, but they are related by a simple formula called *Bayes' rule*

$$\Pr[B \mid A] = \frac{\Pr[A \mid B]\Pr[B]}{\Pr[A]}.$$

This can be derived from the product rule, but this form is so useful that it's worth committing to memory. An especially useful consequence of Bayes' rule is the expression for the *ratio* of the conditional probabilities of two events B, C given A:

$$\frac{\Pr[B \mid A]}{\Pr[C \mid A]} = \frac{\Pr[A \mid B]\Pr[B]}{\Pr[A \mid C]\Pr[C]}.$$

5 Example 2: Biased and fair coins

We will now explore an application of Bayes' rule. First, let's name the terms appearing in it:

$$\Pr[B \mid A] = \frac{\Pr[A \mid B] \cdot \Pr[B]}{\Pr[A]}.$$

We refer to $\Pr[A \mid B]$ as the *likelihood (of A given B)*, and $\Pr[B]$ as the *prior probability of B*. The left-hand side $\Pr[B \mid A]$ is the *posterior probability of B*.

Suppose I have a biased coin (which always comes up heads when I flip it), and a fair coin (which comes up heads half the time, and tails half the time). Suppose I pick a coin with uniform probability, and flip it, observing heads. What is the chance that the coin I picked was fair?

Mathematically, let H denote the event of seeing heads, F denote the event of picking a fair coin, and B the event of picking a biased coin. Then we have

$$\begin{aligned}\frac{\Pr[F \mid H]}{\Pr[B \mid H]} &= \frac{\Pr[H \mid F]\Pr[F]}{\Pr[H \mid B]\Pr[B]} \\ &= \frac{1/2 \cdot 1/2}{1 \cdot 1/2} \\ &= 1/2.\end{aligned}$$

Thus, the chance that the coin was fair is 1/3, and the chance that it was biased is 2/3.

Now observe that if the prior probability changes, this changes as well. For instance, if $\Pr[F] = 1 - \epsilon$, we see that we get a ratio of

$$\frac{1/2 \cdot (1-\epsilon)}{1 \cdot \epsilon}.$$

This ratio goes to ∞ as $\epsilon \to 0$, so $\Pr[F \mid H] \to 1$ as $\epsilon \to 0$.

6 Example 3: COVID testing

In the next few examples, we're going to see some examples of counterintuitive behavior arising from Bayes' rule, where intuitive reasoning underestimates the effect of the prior.

Suppose 10% of the MIT community has COVID, and everybody is required to take a COVID test. The tests have a *false positive* rate of 0.3, and a *false negative* rate of 0.1. If I test positive, what's the chance I have COVID?

- **Events:** H I am healthy, S I am sick, $+$ I test positive, $-$ I test negative.
- **Probabilities:** $\Pr[H] = 0.9$, $\Pr[+ \mid H] = 0.3$, $\Pr[- \mid S] = 0.1$. From these we deduce the complements: $\Pr[S] = 0.1, \Pr[- \mid H] = 0.7, \Pr[+ \mid S] = 0.9$.
- **Conditional probability:** We want to calculate $\Pr[S \mid +]$. Use Bayes' rule, again for the odds:

$$\begin{aligned}\frac{\Pr[S \mid +]}{\Pr[H \mid +]} &= \frac{\Pr[+ \mid S]\Pr[S]}{\Pr[+ \mid H]\Pr[H]} \\ &= \frac{0.9 \cdot 0.1}{0.3 \cdot 0.9} \\ &= \frac{1}{3}.\end{aligned}$$

  So I have 1/4 chance of being sick and a 3/4 chance of being healthy!

Even though the test looks pretty good on paper, the *base rate* (the prior probability that I'm sick) is the dominant effect here: I should still think I'm more likely than not healthy, even if I get a positive test.

Of course, in the real world, we don't test everybody, so this is not realistic. We only test people with symptoms, so in reality, we care about $\Pr[S \mid +, \text{has symptoms}]$. This is a much higher quantity in general!

7 Example 4: Simpson's paradox

An analysis of 1973 UC Berkeley graduate admissions data revealed the following paradoxical facts: the admissions rate was higher for men than for women for the university as a whole (that is, the fraction of men applicants who were admitted was higher than the fraction of women applicants who were admitted). However, for *each department*, the admissions rate for men was lower than it was for women. How could this be? Was the data wrong?

It turns out that these facts are both consistent with each other. Suppose there are only two departments: EE and CS. Define the events A for a student being admitted, M/F for the student being male or female, EE/CS for the student applying to the EE or CS departments. The statement about university-wide admissions rates is

$$\Pr[A \mid M] > \Pr[A \mid F].$$

The statement about per-department admissions rates is

$$\Pr[A \mid F \cap CS] \geq \Pr[A \mid M \cap CS]$$
$$\Pr[A \mid F \cap EE] \geq \Pr[A \mid M \cap EE].$$

How do we square these with each other?

$$\begin{aligned}
\Pr[A \mid F] &= \frac{\Pr[A \cap F]}{\Pr[F]} \\
&= \frac{\Pr[A \cap F \cap CS] + \Pr[A \cap F \cap EE]}{\Pr[F]} \\
&= \frac{\Pr[A \mid F \cap CS]\Pr[CS \mid F]\Pr[F] + \Pr[A \mid F \cap EE]\Pr[EE \mid F]\Pr[F]}{\Pr[F]} \\
&= \Pr[A \mid F \cap CS]\Pr[CS \mid F] + \Pr[A \mid F \cap EE]\Pr[EE \mid F] \\
\Pr[A \mid M] &= \Pr[A \mid M \cap CS]\Pr[CS \mid M] + \Pr[A \mid M \cap EE]\Pr[EE \mid M].
\end{aligned}$$

So to make the overall rates favor men, we can adjust the values of $\Pr[CS \mid M]$ and $\Pr[CS \mid F]$ accordingly. Here's the intuition: suppose both CS and EE both mildly favor women, but CS is much more popular with women, and is also much harder to get into (for everyone). Then the value of $\Pr[A \mid F]$ will be much smaller than $\Pr[A \mid M]$, simply because of the "base rate" of students applying to different departments, *not* because of a gender difference in *conditional* acceptance probabilities. As an extreme example, suppose 100 men and 100 women apply. 99 women and 1 man apply to CS. 99 men and 1 woman apply to EE. CS is super snobby and accepts 1 applicant (a woman). EE has no standards and rejects 1 applicant (a man). Now admissions rates are:

- > 0 for women and 0 for men in CS,
- 1 for women and < 1 for men in EE,
- 2% for women and 98% for men overall.

We could interpret this as saying that the gender bias in admissions is not caused by a direct preference for men over women at the level of individual application readers, but rather by other aspects of the system that determine how likely students are to apply to each department.

8 Example 5: O. J. Simpson

O. J. Simpson was a retired football player who was accused, and later acquitted, of the murder of his wife, Nicole.
Question: Was O. J.'s history of abuse towards his wife was admissible into evidence?

Prosecution: Abusers are 10× more likely than randos to be murderers. Therefore, abuse is likely precursor to murder, and should be taken into account.

Defense: Probability of abusive husband murdering wife is $\sim 1/2500$. Therefore, abuse history has negligible probative value. It would, however, bias the jury against Simpson, so should be barred.

Who is right? Both are attempting to reason about conditional probability, specifically the conditional probability that a husband murders his wife, given that he abuses her. Let's make precise some events.

- Let A be the event [Husband abuses Wife].
- Let G be the event [Husband murders Wife].
- Let M be the event [Wife is murdered].

Prosecution argued that $\frac{\Pr[G \mid A]}{\Pr[G \mid \bar{A}]}$ is high, so knowing A dramatically increases the posterior probability of G.

Defense argued that $\Pr[G \mid A]$ is low, so knowing A cannot dramatically increase the posterior probability of G.

Both neglected the fact that Nicole *was* murdered; the relevant probability is $\Pr[G \mid A \cap M]$, which as it turns out, was around 80%: 80% of abusive widowers with murdered wives are the murderers.

Probability and conditional probability are used and misused all the time, and even experts make (very public) mistakes. If in doubt, make everything precise and fall back on the fundamentals!