Lecture 4 - Conditional Probability
6.042 - February 13, 2003

1 Conditional Probability

Suppose that we pick a random person in the world. Everyone has an equal chance of being picked. Let A be the event that the person is an MIT student, and let B be the event that the person lives in Cambridge. The situation is shown in Figure 1. Clearly, both events A and B have low probability. But what is the probability that a person is an MIT student, *given* that the person lives in Cambridge?

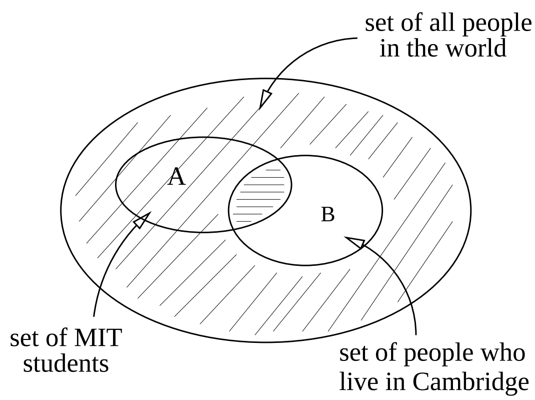

Figure 1: What is the probability that a random person in the world is an MIT student, given that the person is a Cambridge resident?

This is a conditional probability question. It can be concisely expressed in a special notation. In general, $\Pr\{A \mid B\}$ denotes the probability of event A, given that event B occurs. In this example, $\Pr\{A \mid B\}$ is the probability that the person is an MIT student, given that he or she is a Cambridge resident.

How do we compute $\Pr\{A \mid B\}$? Since we are *given* that the person lives in Cambridge, all outcomes outside of event B are irrelevant; these irrelevant outcomes are diagonally shaded in the figure. Intuitively, $\Pr\{A \mid B\}$ should be the fraction of Cambridge residents that are also MIT students. That is, the answer should be the probability that the person is in set $A \cap B$ (horizontally shaded) divided by the probability that the person is in set B. This leads us to the following definition.

Definition 1 *Let* A *and* B *be events such that* $\Pr\{B\} > 0$. *Then the* conditional probability of A given B *is denoted* $\Pr\{A \mid B\}$ *and defined by:*

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

As an example, what is $\Pr\{A \mid B\}$ when event B is a subset of event A? This is the case if A is the event that a random person is an MIT student and B is the event that the person is a 6.042 student. From the definition of conditional probability, we have:

$$\begin{aligned} \Pr\{A \mid B\} &= \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \\ &= \frac{\Pr\{B\}}{\Pr\{B\}} \\ &= 1 \end{aligned}$$

In words, the probability that a person is an MIT student, given that they are a 6.042 student is 1.

Rearranging terms in this definition gives an equation called the *product rule*:

$$\Pr\{A \cap B\} = \Pr\{B\} \cdot \Pr\{A \mid B\}$$

This is useful for calculating the probabilty that two events *both* occur. We've actually been using this formula on the sly all along. Exactly where will be clear shortly.

Conditional probabilities also play a central role in an important equation called the *Total Probability Law*:

$$\Pr\{A\} = \Pr\{A \mid B\} \cdot \Pr\{B\} + \Pr\{A \mid \overline{B}\} \cdot \Pr\{\overline{B}\}$$

Sometimes the probability of an event A depends on whether or not some other event B occurs. In such situations, one can divide the analysis into two cases, when B occurs and when it does not, and then assemble the results to find the probability of A using the Total Probability Law.

2 The Halting Problem

The *Halting Problem* is the canonical undecidable problem in computation theory that was first introduced by Alan Turing in his seminal 1936 paper, "On Computable Numbers, with an Application to the Entscheidungsproblem". The problem is to determine whether a Turing machine halts on a given blah, blah, blah. *Much more importantly*, it is the name of the Laboratory for Computer Science D-league hockey team.

In a best-of-three tournament, the Halting Problem wins the first game with probability $\frac{1}{2}$. In subsequent games, their probability of winning is determined by the outcome of the previous game. If the Halting Problem won the previous game, then they are energized by victory and win the current game with probability $\frac{2}{3}$. If they lost the previous game, then they are demoralized by deefeat and win the current game with probablity $\frac{1}{3}$. What is the probability that the Halting Problem wins the tournament, given that they win the first game?

We can cast this question in conditional probability terms. Let T be the event that the Halting Problem wins the tournament, and let F be the event that they win the first game. Our goal is then to determine a conditional probability:

$$\Pr\{T \mid F\} \quad = \quad \frac{\Pr\{T \cap F\}}{\Pr\{F\}}$$

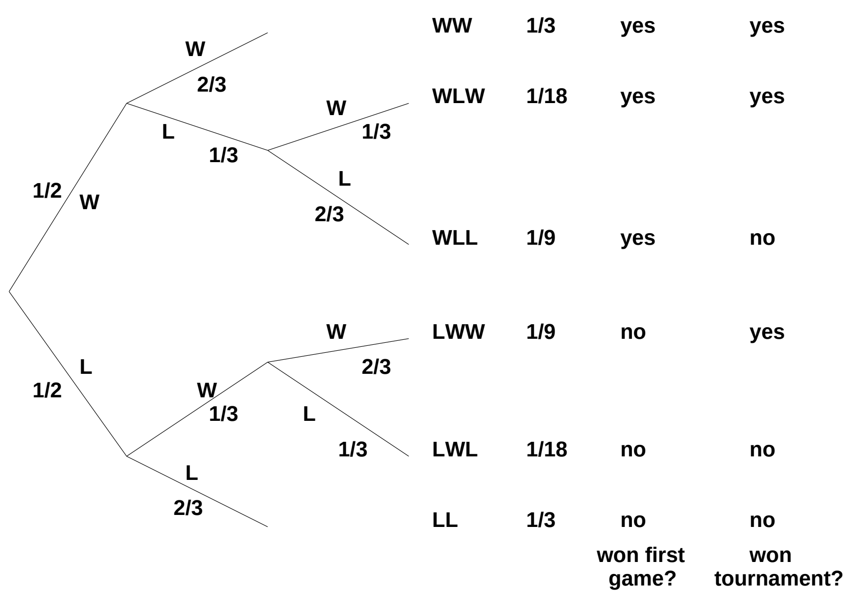

Figure 2: Tree diagram for the Halting Problem tournament.

We can evaluate the expression on the right by finding the probabilities of events $T \cap F$ and F. For this purpose, we can use a tree diagram and the usual four-step method. Such a tree diagram is shown in Figure 2. The edges are annotated with victory probabilities based on the the Halting Problem's psychology. As usual, outcomes probabilities are computed by multiplying edge probabilities along root-to-leaf paths. The event $T \cap F$ consists of two outcomes:

$$\{\ WW,\ WLW\ \}$$

The event F consists of three outcomes:

$$\{\ WW,\ WLW\,WLL\}$$

The probability of an event is the some of the probabilities of the outcomes it contains. Therefore, we conclude:

$$\begin{aligned}
\Pr\{T \mid F\} &= \frac{\Pr\{T \cap F\}}{\Pr\{F\}} \\
&= \frac{\frac{1}{3} + \frac{1}{18}}{\frac{1}{3} + \frac{1}{18} + \frac{1}{9}} \\
&= \frac{7}{9}
\end{aligned}$$

Thus, the Halting Problem is very likely to win the tournament, given that they win the first game.

Why Tree Diagrams Work

Before putting away the Halting Problem example, let's reconsider one step. *Why* is the probability of an outcome equal to the product of the edge probabilities along the path from the root to that outcome? This is a fact we've relied upon since the first lecture, but what is the mathematical justification? For example, we concluded above that:

$$\begin{aligned}
\Pr\{WW\} &= \frac{1}{2} \cdot \frac{2}{3} \\
&= \frac{1}{3}
\end{aligned}$$

The basis for this conclusion is actually the product rule. We know that the Halting Problem wins the first game with probability $\frac{1}{2}$. Furthermore, we know that they win the second game with probability $\frac{2}{3}$, given that they won the first game. In other words, we know:

$$\begin{aligned}
\Pr\{\text{win first game}\} &= \frac{1}{2} \\
\Pr\{\text{win second game} \mid \text{win first game}\} &= \frac{2}{3}
\end{aligned}$$

Based on these two facts, we use the product rule to compute the probability of outcome WW, the event that they win both the first and second games:

$$\begin{aligned}
&\Pr\{\text{win first game} \cap \text{win second game}\} \\
&\quad = \Pr\{\text{win first game}\} \cdot \Pr\{\text{win second game} \mid \text{win first game}\} \\
&\quad = \frac{1}{2} \cdot \frac{2}{3} \\
&\quad = \frac{1}{3}
\end{aligned}$$

This explains why tree diagrams work! On each edge of a tree diagram, we record the conditional probability of traversing that edge, given that we reach the parent node. The product rule implies that the probability of an outcome is the product of the edge probabilities on the path from the root to that outcome.

3 A Posteriori Probability

Probability theory sometimes allows us to compute the probability that something happened, given observed consquences. This is called an *a posteriori* probability. For example, suppose that I have two coins. One is normal and fair, but the other has heads on both sides. Suppose that I pick a coin uniformly at random, flip it, and the outcome is heads. What is he probability that I picked the fair coin?

The sample space is worked out in the tree diagram in Figure 3. Let F be the event that I picked the fair coin, and let H be the event that I flipped heads. Our goal is to compute a conditional probability:

$$\Pr\{F \mid H\} = \frac{\Pr\{F \cap H\}}{\Pr\{H\}}$$

As usual, we determine outcome probabilities by multiplying edge probabilities along root-to-leaf paths. The probability of the events $F \cap H$ and H are equal to the sum of the probabilities of the outcomes they contain. We find:

$$\begin{aligned}
\Pr\{F \mid H\} &= \frac{\Pr\{F \cap H\}}{\Pr\{H\}} \\
&= \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{2}} \\
&= \frac{1}{3}
\end{aligned}$$

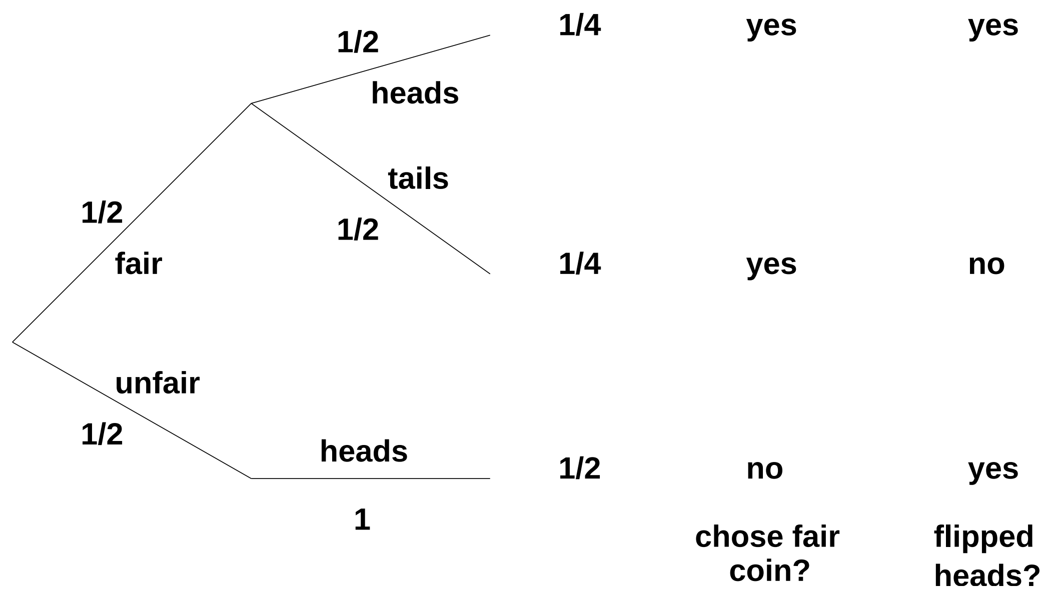

Figure 3: Tree diagram for the two coins experiment.

People are sometimes troubled by a posteriori probabilities on a philosophical level. If I flip heads, then I've *already chosen* either the normal coin or the double-headed coin. Which coin I flipped is at that point not an matter of probability; rather, it is an already determined fact.

4 The 3-Card Game

Suppose that I have three cards. One is red on both sides (RR), one is blue on both sides (BB), and one is red on one side and blue on the other (RB). I jumble the cards in a hat, pick one out, and slap it on the table. The side you can see is blue. What is the probability that the other, hidden side is also blue?

One argument says that the hidden side is blue with probability $\frac{1}{2}$. After all, there are three equally-likely outcomes: RR, BB, and RB. We want the probability that the hidden face is blue, given that the exposed face is blue. Since the exposed face is red, I could not have picked the RR card out of the hat. This leaves two equally-likely possibilites: either I picked BB or RB. In one case, the hidden face is blue and in the other it is red. Therefore, the hidden side is blue with probability $\frac{1}{2}$.

This argument is bogus. The error goes back to the definition of an outcome: *An outcome of an experiment consists of the total information about the experiment after it has been performed, including the values of all random choices.*

There are two random choices in this experiment, not one: I not only select a card from the hat, but also an orientation of that card. Thus, as the tree diagram in Figure 4 shows, there

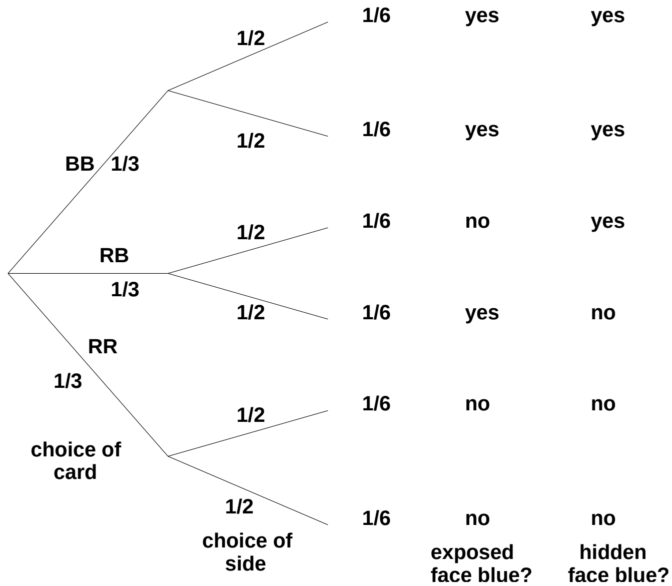

Figure 4: Tree diagram for the 3-card experiment.

are six outcomes, not three. Let E be the event that the exposed face is blue, and let H be the event that the hidden face is blue. After computing outcome probabilities, we can determine the desired conditional probability:

$$\begin{aligned}
\Pr\{H \mid E\} &= \frac{\Pr\{H \cap E\}}{\Pr\{E\}} \\
&= \frac{\frac{1}{6} + \frac{1}{6}}{\frac{1}{6} + \frac{1}{6} + \frac{1}{6}} \\
&= \frac{2}{3}
\end{aligned}$$

Therefore, if the exposed face is blue, then the hidden face is blue with probability $\frac{2}{3}$.