1   Codebreaking

Continued from previous section.

1.1   Interlude: Conditional Probability

Suppose Kambili and Amaka both secretly flip two fair coins.

- Kambili announces that her second flip was heads.

- Amaka announces that she had at least one heads.

Who is more likely to have flipped two heads? In other words, if you had to make a bet about who flipped more heads, who would you bet on?

Definition 1.1: Conditional Probability

Fix a probability space. Given two events A and B, we define the **conditional probability** $\mathbb{P}[A|B]$ by

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

The intuition here is that $\mathbb{P}[A|B]$ represents how confident we are that A happens, *given that we already know* that B happens.

(Example.) Consider the example with Kambili and Amaka. Intuitively, the answer is that "both are equally likely"; that is, both Kambili and Amaka have an equal chance of getting two heads. This is not correct.

To formalize this argument, consider the following:

- The experiment that we're considering involves two coin flips, so we're working with

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

- We're interested in the event
$$A = \{HH\}.$$

In Kambili's situation, we know that her second flip was heads. So, in other words, we're restricting ourselves to the event
$$B_1 = \{HH, TH\}$$
and we have
$$\mathbb{P}[A|B_1] = \frac{\mathbb{P}[A \cap B_1]}{\mathbb{P}[B_1]} = \frac{\mathbb{P}[A]}{\mathbb{P}[B_1]} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Therefore, **the probability that Kambili has two heads** is $\frac{1}{2}$. In Amaka's situation, we only know that one of her flips was heads. In other words, we have the event

$$B_2 = \{HH, TH, HT\}.$$

Then,
$$\mathbb{P}[A|B_2] = \frac{\mathbb{P}[A \cap B_2]}{\mathbb{P}[B_2]} = \frac{\mathbb{P}[A]}{\mathbb{P}[B_2]} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Therefore, **the probability that Amaka has two heads** is $\frac{1}{3}$. In other words, Kambili is more likely to have two heads than Amaka.

(Exercise.) There are three coins in a bag. One is a normal quarter: one side is heads, the other side is tails. The second coin is almost identical except that both sides are heads; similarly, both sides of the third coin are tails. You shake the bag around to shuffle the coins. You then close your eyes, pull one coin out at random, put it down on a table, and then open your eyes. You see heads. What is the probability that the other side of the coin is also heads?

Consider the following tree diagram:
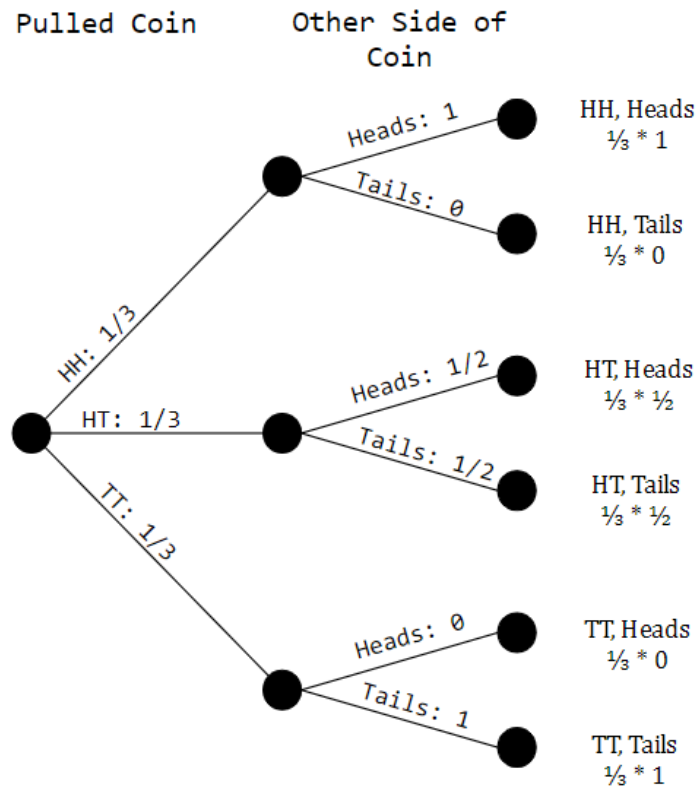
So, we want to find the probability that, given we got heads initially, the otherwise will also have heads. This gives us

$$\mathbb{P}[\text{heads}|\text{got heads}] = \frac{\mathbb{P}[\text{heads given heads}]}{\mathbb{P}[\text{got heads}]} = \frac{1/3 \cdot 1}{1/2} = \frac{2}{3}.$$

Note that one way we can think about getting heads is by thinking about the possible face we *can* get; in this case, we can either get $\{H, H, H, T, T, T\}$. We have a $\frac{3}{6} = \frac{1}{2}$ chance of getting heads.

(Exercise.) Suppose 80% of people like peanut butter, 89% like jelly, and 78% like both. Given that a randomly sampled person likes peanut butter, what's the probability that they also like jelly?

Let J be the event that someone likes jelly and B be the event that someone likes peanut butter. We know that
$$PR[B] = 0.80.$$
We also know that $\mathbb{P}[J \cap B] = 0.78$ since 78% of people like *both* peanut butter and jelly. So,
$$\mathbb{P}[J|B] = \frac{\mathbb{P}[J \cap B]}{\mathbb{P}[B]} = \frac{0.78}{0.80} = 0.975.$$

Lupus is a medical phenomenon where antibodies that are supposed to attack foreign cells to prevent infections instead see plasma proteins as foreign bodies, leading to a high risk of blood clotting. It is believed that 2% of the population suffers from this disease. A test for lupus is 98% accurate if a person actually has the disease, and 74% accurate if a person does not have the disease. There is a line from the Fox television show House that is often used after a patient tests positive for lupus: "It's never lupus." Do you think there is truth to this statement? Use appropriate probabilities to support your answer.

Consider the following tree diagram:
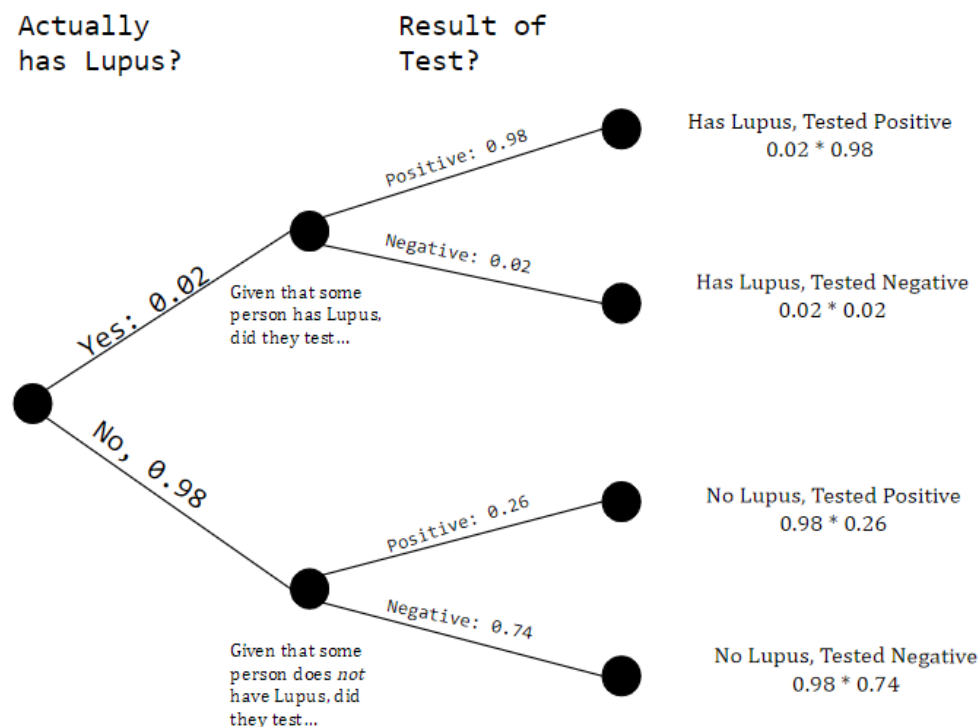
Then,
$$\mathbb{P}[\text{has lupus}|\text{tested positive}] = \frac{\mathbb{P}[\text{has lupus + tested positive}]}{\mathbb{P}[\text{tested positive}]} = \frac{0.02 \cdot 0.98}{0.02 \cdot 0.98 + 0.98 \cdot 0.26} \approx 0.07142.$$

So, there is some truth to the statement since if someone tests positive for lupus, there's only 7.14% that they have lupus.

Definition 1.2: Independent Events

Fix a probability space. We say that two events A and B are independent if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B].$$

Often, it's convenient to reformulate this definition slightly. In the case that $\mathbb{P}[B] > 0$,

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B] \iff \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \mathbb{P}[A].$$

However, notice that

$$\frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \mathbb{P}[A|B]$$

so it follows that the independence of A and B is equivalent to asserting that

$$\mathbb{P}[A|B] = \mathbb{P}[A].$$

We could interpret this statement as follows: if A and B are independent, then our confidence in A happening does not change at all even if we're told that B happened (or did not happen). More loosely, knowing whether or not B happens tells us "nothing" about whether or not A happenes.

(Exercise.) The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line, 20.7% speak a language other than English at home, and 4.2% fall into both categories. Is the event that a randomly chosen American lives below the poverty line independent of the event that the person speaks a language other than English at home?

> Define
> $$\mathbb{P}[\text{Below Poverty Line}] = 0.146,$$
> $$\mathbb{P}[\text{Speak Language Other Than English}] = 0.207,$$
> $$\mathbb{P}[\text{Below Poverty Line AND Speak Language Other Than English}] = 0.042.$$
> Using the formula in (1.2), notice how
> $$0.042 \neq 0.146(0.207) = 0.030222.$$

(Exercise.) A bag contains 5 red marbles and 3 blue marbles. Two marbles are drawn randomly from the bag: Alejandra takes the first one and Beatrice takes the second. Is the event that Alejandra's marble is blue independent of the event that Beatrice's marble is blue?

> No. If Alejandra takes the first marble and doesn't put it back, then it's possible that she took the blue marble, which affects the probability that Beatrice gets a blue marble.

Definition 1.3

Fix a probability space. Two random variables X and Y are **independent** if the events $X = a$ and $Y = b$ are independent for all pairs (a, b) where a is a value of X and b a value of Y.