Intermediate Math Circles
November 15, 2017
Probability III

Example 1:

You have 3 bins in which there are coloured balls. The balls are identical except for their colours. The contents of the containers are:

- Bin 1: 3 green balls, 6 yellow balls, 1 white ball
- Bin 2: 4 black balls, 3 white balls, 2 green balls, 1 orange ball
- Bin 3: 6 brown balls, 2 green balls, 2 purple balls

The probability of choosing Bin 1 is 50%, the probability of choosing Bin 2 is 30%, and the probability of choose Bin 3 is 20%. Once a random bin is chosen, a ball is drawn out.

a) What is P(ball is green)? What is P(ball is yellow)? What is P(ball is white)?

Solution:

$$\begin{aligned} P(\text{green}) &= P(\text{green} \cap \text{bin 1}) + P(\text{green} \cap \text{bin 2}) + P(\text{green} \cap \text{bin 3}) \\ &= P(\text{bin 1})P(\text{green}|\text{bin 1}) + P(\text{bin 2})P(\text{green}|\text{bin 2}) + P(\text{bin 3})P(\text{green}|\text{bin 3}) \\ &= (0.5)\left(\frac{3}{10}\right) + (0.3)\left(\frac{2}{10}\right) + (0.2)\left(\frac{2}{10}\right) \\ &= 0.15 + 0.06 + 0.04 \\ &= 0.25 \end{aligned}$$

$$\begin{aligned} P(\text{yellow}) &= P(\text{yellow} \cap \text{bin 1}) \\ &= P(\text{bin 1})P(\text{yellow}|\text{bin 1}) \\ &= (0.5)\left(\frac{6}{10}\right) \\ &= 0.3 \end{aligned}$$

$$\begin{aligned} P(\text{white}) &= P(\text{white} \cap \text{bin 1}) + P(\text{white} \cap \text{bin 2}) \\ &= P(\text{bin 1})P(\text{white}|\text{bin 1}) + P(\text{bin 2})P(\text{white}|\text{bin 2}) \\ &= (0.5)\left(\frac{1}{10}\right) + (0.3)\left(\frac{3}{10}\right) \\ &= 0.05 + 0.09 \\ &= 0.14 \end{aligned}$$

b) There are fewer yellow balls than green, and all of the yellow balls are in one bin, but the probability of drawing a yellow ball is greater than the probability of drawing a green ball. How do you explain this?

Recall the definition of conditional probability of A given B:

$$\begin{aligned} P(A|B) &= \frac{P(A\cap B)}{P(B)} \\ P(A|B)P(B) &= P(A\cap B) \end{aligned}$$

Now consider the related conditional probability of B given A:

$$\begin{aligned} P(B|A) &= \frac{P(B\cap A)}{P(A)} \\ P(B|A)P(A) &= P(B\cap A) \end{aligned}$$

Since $P(A\cap B) = P(B\cap A)$,

$$P(A|B)P(B) = P(B|A)P(A)$$

Rearranging, we get

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} \end{aligned}$$

This is a special case of Bayes' Theorem

Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$

Example 2:

Let's return to our previous example with the bins.

- Bin 1: 3 green balls, 6 yellow balls, 1 white ball
- Bin 2: 4 black balls, 3 white balls, 2 green balls, 1 orange ball
- Bin 3: 6 brown balls, 2 green balls, 2 purple balls

$P(\text{Bin 1}) = 0.5,\ P(\text{Bin 2}) = 0.3,\ P(\text{Bin 3}) = 0.2$

Suppose now that you do not know which bin has been selected. Someone tells you that a green ball has been selected. What is the probability that Bin 1 was chosen?

Solution:

We want to use Bayes' Theorem here, so we have to decide what is our A and what is our B. The proposition that we are trying to verify is that "Bin 1 was chosen", so we let A = Bin 1. The evidence that we have is that a green ball was selected, so we have B = green. Using these, we get

$$\begin{aligned} P(\text{bin 1}|\text{green}) &= \frac{P(\text{green}|\text{bin1})P(\text{bin 1})}{P(\text{green})} \\ &= \frac{\left(\frac{3}{10}\right)(0.5)}{(0.25)} \\ &= 0.6 \end{aligned}$$

Where did these values come from? We know that $P(\text{bin 1}) = 0.5$ and we can easily compute $P(\text{green}|\text{bin 1}) = 0.3$. For the denominator, we want to compute $P(\text{green})$. Our first instinct might be to divide the number of green balls in all three bins by the total number of balls in all of the bins. This isn't going to work because the probability of picking balls out of bin 1 is greater than picking them out of bins 2 and 3 and we have to take this into account.

$$\begin{aligned} P(\text{green}) &= P(\text{green} \cap \text{bin 1}) + P(\text{green} \cap \text{bin 2}) + P(\text{green} \cap \text{bin 3}) \\ &= P(\text{bin 1})P(\text{green}|\text{bin 1}) + P(\text{bin 2})P(\text{green}|\text{bin 2}) + P(\text{bin 3})P(\text{green}|\text{bin 3}) \\ &= (0.5)\left(\frac{3}{10}\right) + (0.3)\left(\frac{2}{10}\right) + (0.2)\left(\frac{2}{10}\right) \\ &= 0.15 + 0.06 + 0.04 \\ &= 0.25 \end{aligned}$$

So what is our calculation telling us? It tells us that if we know that a green ball was selected, then we should change our belief that bin 1 was chosen from 50% to 60%.

Exercises

1. What is the probability that bin 2 was chosen? Bin 3?

2. What if you were told that a white ball had been selected. What is the probability that bin 1 was chosen? Bin 2 ? Bin 3? Give an intuitive explanation for the value you found for bin 3 being chosen.

Example 3:

During a shift, the probability of a worker making an error on the production line is 0.05. The probability that an accident will occur when there is a worker error is 0.4 and the probability that there is an accident when there is no worker error is 0.2. Find the probability of a worker error given that there is an accident.

Solution:

Let A represent the event of an accident.
Let E represent the event of worker error.
We are given

$$P(E) = 0.05 \qquad P(A|E) = 0.4 \qquad P(A|E^C) = 0.2$$

We want to find

$$P(E|A)$$

Using Bayes' Theorem, we have

$$\begin{aligned} P(E|A) &= \frac{P(A|E)P(E)}{P(A)} \\ &= \frac{P(A|E)P(E)}{P(A|E)P(E) + P(A|E^C)P(E^C)} \\ &= \frac{(0.4)(0.05)}{(0.4)(0.05) + (0.2)(0.95)} \\ &= \frac{0.02}{0.21} \\ &\approx 0.095 \end{aligned}$$

Exercises

1. For two events, A and B, $P(A) = 0.6$, $P(B|A) = 0.7$ and $P(B|A^C) = 0.6$. Find each of the following
   (a) $P(A|B)$
   (b) $P(A^C|B)$

2. A coin, which you are not allowed to examine, is either a fair coin ($P(\text{heads}) = 0.5$) or has 2 heads. Your initial opinion is $P(\text{fair}) = 0.9$. The coin is flipped and heads comes up. What is your opinion now?

Now that we know what Bayes' Theorem is, we want to explore some of the ways that it can be used in real-life situations. Often the results are surprising and seem to contradict common sense.

Bayes' Theorem can be effectively used in assessing the results of any kind of test where there is not complete accuracy in the outcome of the test. These kinds of tests are often given in medical context. Before we turn to the problem, we have to formally define what we mean when we say that the outcome of a test isn't completely accurate.

False Positives and False Negatives

We are going to assume that a test has only two outcomes, which will be called "Positive" (P) and "Negative" (N). We are also going to assume that there are only two possible states that we can be in; we either have what the test is checking for, or we don't. We will call these states "Sick" (S), if we have it, and "Healthy" (H) if we don't. There are now four possible combinations of the two states and the two outcomes.

(N,H): The test says we don't have the disease and we don't
(P,S): The test says we have the disease and we do
(P,H): The test says we have the disease but we don't
(N,S): The test says we don't have the disease but we do

In the first two cases, the test is doing what is supposed to do: if we're healthy, the test says that we are healthy, which if we are sick then the test confirms this. In the second two cases, however, the test is not working, but its not working in different ways. Given (P,H), the test comes back positive but we are healthy. This means that we think we have the disease even though we don't. This has widespread ramifications, emotionally, physically, and financially. When this happens, we say that the test has produced a *false positive*, since the outcome of the test is positive, but that is not the state that we are in. Similarly, given (N,S), the test comes back negative, but we are sick. This means that we don't think that we have the disease even though we do. This has even more dangerous ramifications because we might fail to get treatment that we need to regain our health because we don't think we are sick. When this happens, we say that the test has produced a *false negative*, because the outcome of the test is negative but that is not the state that we are in. To keep the terminology consistent, we refer to (N,H) as a *true negative* and (P,S) as a *true positive*.

Suppose you go to the doctor to have a check-up. You feel perfectly healthy and look perfectly healthy. When you get to the doctor's office, he tells you that there is a pretty serious infection going around, and it seems like 1 in 5000 people in the Waterloo area have contracted the disease. A reasonable description of your knowledge would be the prior probabilities

$$P(\text{healthy}) = 0.9998 \qquad P(\text{sick}) = 0.0002$$

The doctor tells you that he has a test for the disease, and that the test has about 5% false positives and about 2% false negatives, so that

$$P(P|H) = 0.05 \qquad P(N|H) = 0.95 \qquad P(P|S) = 0.98 \qquad P(N|S) = 0.02$$

You decide to take the test. A few days later the doctor calls you with your results.

1. If the doctor said the test came back negative, what is the probability that you are healthy?
2. If the test came back positive, what is the probability that you are sick?

The Monty Hall Problem

This is a famous application of Bayes' Theorem to a probabilistic problem which explains the counterintuitive results.

The Problem

Suppose that you are a contestant on a game show. There are three doors, which are closed, and which are numbered 1, 2, and 3. Behind one of the doors is a brand new sports car. Behind the other two doors are goats. You have one chance to pick a door, and if you pick the door with the sports car behind it, you win the sports car. If you pick a door with a goat, you win the goat. Say that you pick door 1, and announce it to the host. Before he opens door 1, he opens door 3 and reveals a goat. The host then turns to you and asks if you want to change your choice to door 2.
Is it to your advantage to switch?

Before the host opens door 3, you figure that you have a $\frac{1}{3}$ chance of winning the sports car. When he opens door 3 and shows you a goat, you now know that the sports car is either behind door 1 or door 2. You might think to yourself that since you now know that there are only two possibilities (door 1 and door 2), that the probability of the sports car being behind door 1 is $\frac{1}{2}$ and the probability of it being behind door 2 is $\frac{1}{2}$, so there is no advantage in switching. In essence, once the host opens door 3, you think to yourself that the sample space has shrunk from three options to two options, and since the information provided by opening door 3 doesn't change the location of the car, each option is still equally likely.

You're standing there thinking and thinking. The host can tell that you're working through the various probabilities, and he asks you to share your thoughts with him. You explain that you're not going to switch because you figure that the probabilities of being behind door 1 and door 2 are both $\frac{1}{2}$ and you'll stick with your gut instinct which told you to pick door number 1. The host listens to you, nodding as you explain your reasoning. When you're finished, he says "What if I told you that the probability of the car being behind door 2 is $\frac{2}{3}$?" Do you believe him? Would you switch?

Let's consider the probabilities of the different options and investigate this more rigorously.

Door # Picked	Door # Car Behind	Door Host Opens	Probability Host Opens
1	1	2	0.5
		3	0.5
	2	3	1
	3	2	1
2	1	3	1
	2	1	0.5
		3	0.5
	3	1	1
3	1	2	1
	2	1	1
	3	1	0.5
		2	0.5

These probabilities are all that we have to work with, but it turns out that they are all we need.

Exercise

1. Suppose that you pick door 1. Using Bayes' Theorem, find the probability that the car is being door 2 if the hosts opens door 3.

Problem Set

1. If you take a bus to work in the morning, there is a 30% chance that you will arrive late. When you go by bicycle there is a 10% chance you will arrive late. 70% of the time you go by bike, and 30% of the time you go by bus. Given that you arrive late, what is the probability that you took the bus?

2. A box contains 4 coins – 3 fair coins and 1 biased coin for which $P(\text{heads}) = 0.6$. A coin is picked at random and tossed 5 times. It shows 3 heads. Find the probability that the coin is fair.

3. At a police spot check, 10% of cars stopped have defective break lights and a faulty muffler. 15% of the cars have a defective break lights, but the muffler was ok. If a car is stopped and has defective break lights, what is the probability that the muffler is also faulty?

4. For one deck building project, a contractor buys 70% of the lumber from supplier A, and 30% of the lumber from supplier B. Typically, 85% of the lumber from supplier A arrived undamaged, while 90% of the lumber from supplier B arrived undamaged. If the contractor pulls one board out and its damaged, what is the probability that the damaged board is from

   (a) Supplier A?

   (b) Supplier B?

Problem Set Answers

1. $\frac{3}{7}$

2. 0.73065

3. 0.4

4. (a) 0.778

   (b) 0.222

Problem Set Solutions

1. Let B represent that you take the bus, and let L represent that you are late. Then

$$\begin{aligned} P(B|L) &= \frac{P(L|B)P(B)}{P(L|B)P(B) + P(L|B^C)P(B^C)} \\ &= \frac{(0.3)(0.2)}{(0.3)(0.2) + (0.1)(0.8)} \\ &= \frac{3}{7} \end{aligned}$$

2. Let F represent that the coin chosen is fair and let H represent that 3 heads were shown. Then

$$\begin{aligned} P(F|H) &= \frac{P(H|F)P(F)}{P(H|F)P(F) + P(H|F^C)P(F^C)} \\ &= \frac{\binom{5}{3}\left(\frac{1}{2}\right)^5 \times \left(\frac{3}{4}\right)}{\binom{5}{3}\left(\frac{1}{2}\right)^5 \times \left(\frac{3}{4}\right) + \binom{5}{3}(0.6)^3(0.4)^2 \times \left(\frac{1}{4}\right)} \\ &= 0.73065 \end{aligned}$$

3. Let B represent the event that a break lights are defective and let M represent the event that a muffler is defective.

$$\begin{aligned} P(M|H) &= \frac{P(M \cap H)}{P(H)} \\ &= \frac{P(M \cap H)}{P(M \cap H) \cup P(M^C \cap H)} \\ &= \frac{0.1}{0.1 + 0.15} \\ &= 0.4 \end{aligned}$$

4. (a) Let A represent supplier A and D represent that wood is damaged.

$$\begin{aligned} P(A|D) &= \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|A^C)P(A^C)} \\ &= \frac{(0.15)(0.7)}{(0.15)(0.7) + (0.1)(0.3)} \\ &= \frac{0.105}{0.135} \\ &= 0.778 \end{aligned}$$

(b) Let B represent supplier B and D represent that wood is damaged.

$$\begin{aligned} P(B|D) &= \frac{P(D|B)P(B)}{P(D|B)P(B) + P(D|B^C)P(B^C)} \\ &= \frac{(0.1)(0.3)}{(0.1)(0.3) + (0.15)(0.7)} \\ &= \frac{0.03}{0.135} \\ &= 0.222 \end{aligned}$$