Unit 3 - Conditional Probability and Bayes Theorem

Adam Petrie
Department of Business Analytics
University of Tennessee

June 12, 2021

Overview

The philosophy of business analytics can be summed up as "information matters"

Once new information is available for a problem, optimal solutions may change.

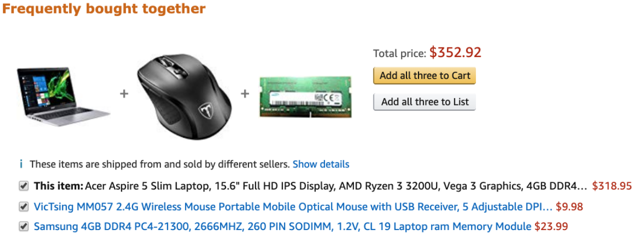
Information Matters

The probability some randomly picked senior will donate to UT in the next five years might be 15%. But what if we knew she graduated without student loans, was a member of a sorority, and was a business analytics major?

Information Matters

Fraudulent credit card purchases are a headache for companies like Chase. It might be that 1 out of every 100000 charges are fraudulent. But what if a charge is for a big screen TV from a store in Alabama, and the owner of the card mostly purchases gas and groceries in Florida?

Information Matters



Cross-selling can greatly increase revenue. Perhaps the probability some random customer buy a wireless mouse is 0.15%. Among people with a laptop and RAM in their cart, this probability might be 12%.

Updating our probabilities

Question: how can we go from the overall fraction of time an event occurs *in general* (probability a transaction is fraudulent in general) to the fraction of time an event occurs *in a more specific scenario* (transaction was out of state and in an usual product category)?

Updating our probabilities

- **Prior Probability**: the overall fraction of time an event occurs in general (transaction is fraudulent)
- **Posterior Probability**: the fraction of time an event occurs in a more specific scenario (transaction is out of state in an usual product category)

Issue: when working with large datasets, estimating these is often easy (look at just the out-of-state purchases in unusual categories; find the fraction of those that were fraudulent). If we don't have that luxury, we have to use theory because we are *terrible* at intuitively coming up with the right answer.

Notation for updating our probabilities

Let A be the event of interest (e.g., the probability that a transaction is fraudulent) and let B represent the set of information we want to incorporate when calculating the probability (e.g., usual purchase in an unusual location).

- $P(A)$ is the **prior** probability that A occurs (before incorporating knowledge of B)
- $P(A|B)$ is the **posterior** probability of A *given* B (the vertical line $|$ is pronounced "given"). In other words, it is the updated probability that A occurs after incorporating the information given by B.

$P(A|B)$ is called a **conditional probability** because we are calculating the probability of A with the condition that we must account for the fact that B has occurred.

Classic example

A medical test comes back positive for a rare disease. What's the probability the person *actually* has it?

Classic example - background

The (prior) probability that a randomly selected individual has this particular rare disease is 0.04%. Tests are quite accurate! If an individual has the disease, there's a 99.9% chance the test comes back positive. If the individual doesn't have the disease, there's a 98% chance the test comes back negative. Given the test reads positive, what's the (posterior) probability they have the disease?

a) 0.04%
b) 2%
c) 15%
d) 40%
e) 60%
f) 85%
g) 98%
h) 99.9%

A (Subtly Sexist) Example

A friend shows you a picture of a woman on his phone. You think she is cute, so you ask what she does for a living. Your friend replies that she's either a waitress or a librarian, but he's not sure which. Which is more likely: she is a waitress or she is a librarian?

A (Subtly Sexist) Example

You then ask for more details about her. Your friend tells you that she typically dresses conservatively and wears glasses. Also, she is fairly quiet, quite intelligent, and likes to read.

Now which is more likely: she is a waitress or she is a librarian? We'll try to answer this later.

Conditional Probability

The study of **conditional probability** allows us to update the chance that some event will occur when new information arrives.

- Given a test comes back positive, what is the probability the person has the disease?
- Given a woman is quiet, intelligent, wears glasses, likes to read, and dresses conservatively, what is the probability she is a librarian?
- Given a UT alumnus is 45 years old, drives a red sports car, has no pets, and is unmarried, what is the probability that she will donate in the next 5 years?

Conditional Probability

Warning: people are *notoriously* bad at guessing conditional probabilities. For whatever reason, our intuition seems to work in opposition to the rules of conditional probability, and it takes training and practice to learn how to approach these problems.

When it comes to conditional probability problems, you need to sit down, take it slow, and "math it out".

Monty Hall Problem

The setup:

- There are 3 doors. Behind two are goats and behind one is a car.
- You pick a door, then the host (who knows where the car is) opens one of the *other* doors to reveal a goat and offers you a choice – stick with the one you picked or switch.

Given this new information, should you switch? Does it matter?

https://math.ucsd.edu/~crypto/Monty/monty.html

vos Savant and the media furor

In 1990, Marilyn vos Savant (used to be listed in Guinness Book of World Records for highest IQ), wrote in her first column "Ask Marilyn" on the Monty Hall problem that the player should **always switch**.

vos Savant and the media furor

She received thousands of letters from her readers. The vast majority of people, including many from readers with PhDs, disagreed with her answer. During 1990-1991 three more of her columns in Parade were devoted to the paradox and the discussion was replayed in other venues and reported in major newspapers such as the New York Times.

After further elaboration, still only about 56% of the general public and 71% of academics accepted that the correct strategy is to always switch.

Switch or not switch

My (and probably your) intuition suggests that once the goat is revealed behind one of the doors, there is an equal chance of the car being behind either of the two remaining doors. After all, there are two options, and we know that the car was placed randomly (each of the three doors were equally likely contain the car).

Thus, our gut suggests it's 50-50 and that it doesn't matter whether you switch or not. Why not "trust your gut" and stick to the door you picked. It would be terrible to switch and lose a car that was in your grasp!

Switch or not switch

Wrong!

Just because there are two options doesn't mean the chances are 50-50.

Initially, the car's location is chosen at random. Therefore, because there are 3 options, the probability of it being behind any given door is 1/3.

However, after one door is opened, the remaining 2 options are *not necessarily* equally likely just because there are two options. We have additional information (what's behind one of the doors), so the probability of the car being behind each of the remaining doors *needs to be updated* from their prior probabilities of 1/3 each to their posterior probabilities that reflect the revealed goat.

Solution

You should **always** switch! Why?

- If you originally picked the door with the car and switch, you lose.
- If you originally picked a door with a goat, then Monty is *forced* to reveal a goat behind one of the two remaining doors. If you switch, you win!
- Therefore, if you picked the door with the car and switch you lose. If you picked a door with a goat and switch you win. There is a 2/3 chance that you picked a door with a goat, so there's a 2/3 chance that by switching you win!

Still not convinced? Another explanation

The **prior** probability of the car being behind *your* door is $1/3$ (the car is placed at random behind 1 of the 3 doors).

Monty revealing a goal behind one of the *other* doors doesn't change that fact. The probability that the car is behind the door you picked is still $1/3$.

What *does* change is the probability that the car is behind the one unopened door. Since the probability that the car is behind the door you picked is always $1/3$, the probability it is behind the unopened door is $1 - 1/3 = 2/3$.

Why does the solution make most people feel uncomfortable?

Since Monty can *always* open a door to reveal a goat no matter what door you initially picked, it *seems* like you're not really gaining any additional information. Thus, it *seems* like the probabilities should stay the same. Partially correct!

- The additional info *didn't* change the probability that the car is behind the door you picked. It's still $1/3$.
- The additional info *did* change the probability that the car is behind the one unopened door from $1/3$ to $2/3$.

Monty Hall simulation

Canvas has a file `montyhall.r` that allows you to play this game, but with a twist – there are now four doors! After you pick your door, Monty reveals what's behind *two* of the other doors, then you decide whether to stay with your pick or switch.

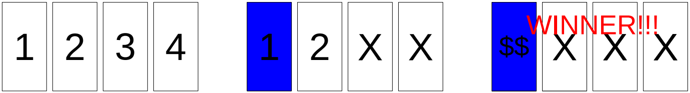
The function `letsmakeadeal()` allows you to interactively play.

The function `letsmakeadealsimulation()` will automatically pick door 1 and then *will never switch*. What do you think the probability of winning will be if you don't switch?

Monty Hall simulation

The `letsmakeadealsimulation()` always picks door 1 and never switches. It outputs either `win` or `lose`.

```
letsmakeadealsimulation()
## Which door (1-4) do you pick?
## Do you stick with door 1 or do you switch to 2 ?  Enter your door number
## [1] "win"
## [1] "win"
```



"Intuitive" Approach for Calculating Conditional Probabilities

Before getting into the math . . .

There is a solid and elegant foundation of probability theory that allows us to compute conditional probabilities. However, even for people comfortable with mathematical notation, it's by no means straightforward to set up the relevant equations (without a lot of practice).

Further, the complexity of the equations tends to obscure just *why* the calculations work!

Let's solve some conditional probability problems without introducing the equations just yet.

Probability of having a disease given a positive test result (visualized)

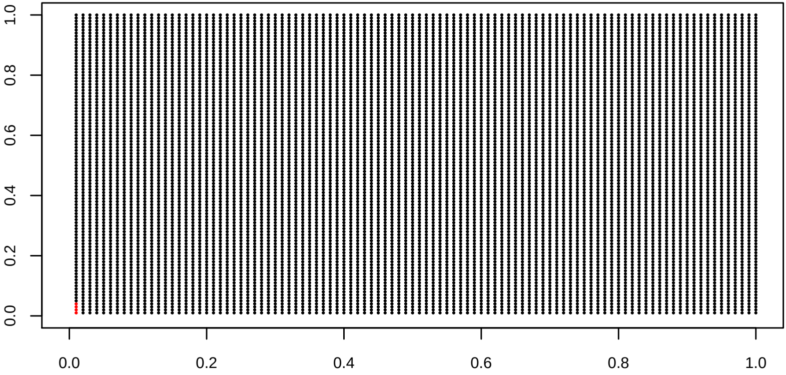
A disease affects 0.04% of population. If you have the disease, the test comes back positive 99.9% of the time. If you don't, the test comes back negative 98% of the time. You just took the test and it came back positive. What is the probability that you have it?

$$P(infected|test+) = ???$$

Let's study the problem visually. Let's create a 100 by 100 grid of dots. Each dot will represent one person (10000 total). Let's color code the people by their disease status: red if they have it (out of 10000, we'd expect 0.04% or 4 of them to have it) and black if they don't.

Probability of having a disease given a positive test result (visualized)

The 4 dots with the disease are in the bottom left. You are equally likely to be any one of these 10000 dots. Thus, your (prior) probability of having the disease is $4/10000 = 0.04\%$.



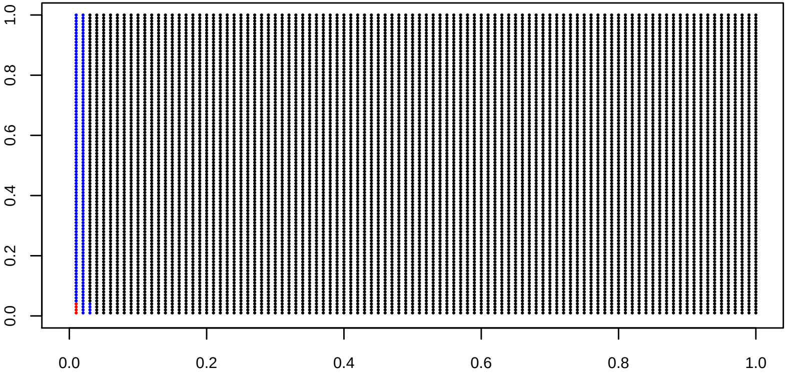
Probability of having a disease given a positive test result (visualized)

Let's give all 10000 people the test for the disease and color-code the dots based on the results.

- When a person has the disease, the probability that the test is positive is 99.9%, so let's assume all four people who have the disease test positive (keeping their color red).
- When a person doesn't have the disease, the probability that the test is negative is 98% (and positive 2%). So of the 9996 people without the disease, 2% of them or 199.92 (let's just say 200) are expected to test positive. Let's color code them blue.

Probability of having a disease given a positive test result (visualized)

The people that have tested positive for the disease (the 4 infected, and 200 additional "false positives") are colored red and blue.



Probability of having a disease given a positive test result (visualized)

When your test comes back positive, you now that you are *not* a black dot, but rather one of the 204 colored dots.

You are equally likely to be any one of these colored dots, but you don't know which.

The (posterior) probability that you have the disease given that the test came back positive is now $4/204 \approx 2\%$.

Probability of having a disease given a positive test result (visualized)

Pleasantly surprised?

The coming back positive is *strong* evidence you have the disease, and indeed it increased the probability from 0.04% to about 2%. You're 50 times more likely to have the disease given the test came back positive than you were before.

However, this disease is just *so* rare that the 50-fold increase in probability still leaves it as a long-shot that you have the disease.

COVID-19 antibody tests

This is why COVID-19 antibody tests were nearly worthless (as of July 30, 2020). Even if the tests were highly accurate like in our example (they aren't), the relatively low prior probability of having been infected with COVID-19.

- Cases in Tennessee on July 30: about 100,000
- Population of Tennessee: 6.8 million
- Prior probability (ignoring the high non-random nature of who has it): ≈ 0.015

Words of encouragement about conditional probability

Note that the final probability calculation was quite straightforward – the fraction of the sample space (colored dots) that contained the event of interest (red dots).

The only thing is that the sample space used in the probability calculation *is not the sample space we started with* (the 10,000 total dots).

> Conditional probability is difficult to compute intuitively (and often has answers that surprise us) because our minds are inherently *bad* at grasping what the appropriate sample space to use in the calculation ends up looking like after we condition a probability on a set of events.

Table Method - Probability of having a disease given a positive test result

Let's redo this calculation using what I call the "table" method (credit to Scott Lawson, former UT undergraduate and Class of 2020 MSBA). Similar to the visualization solution, the table method envisions a *very* large number of trials (people in this case), then figures out the appropriate sample space to use for the conditional probability calculation using basic reasoning.

Table Method - Probability of having a disease given a positive test result

Disease affects 0.04% of people. If you have the disease, the test comes back positive 99.9% of the time. If you don't, the test comes back positive 2% of the time. What is the probability that you have the disease if your test came back positive? Let's make a table showing the expected number of people in the each cells for a population of 10 million people.

	Test Positive	Test Negative	Total
Have Disease	a		
Don't Have Disease			
Total	b		10,000,000

The cell marked b gives the number of individuals in the sample space we'll use in the calculation (all people whose test comes back positive). The probability we want is a/b, the fraction of positive test results that are attributed to people with the disease.

Table Method - Probability of having a disease given a positive test result

Since the disease affects 0.04% of these 10 million people, we expect to have ($0.0004 \cdot 10^7$) 4000 people with the disease, and 9,996,000 without. This allows us to fill in the row totals.

	Test Positive	Test Negative	Total
Have Disease	a		4000
Don't Have Disease			9,996,000
Total	b		10,000,000

Table Method - Probability of having a disease given a positive test result

When someone has the disease, the test comes back positive 99.9% of the time (and negative 0.1% of the time). Thus of the 4000 people with the disease, we expect ($4000 \cdot 0.001$) 3996 tests to come back positive and 4 to come back negative.

	Test Positive	Test Negative	Total
Have Disease	3996	4	4000
Don't Have Disease			9,996,000
Total	b		10,000,000

Table Method - Probability of having a disease given a positive test result

When someone doesn't have the disease, the test comes back positive 2% of the time (and negative 98% of the time). Thus of the 9,996,000 people with the disease, we expect $(0.02 \cdot 9996000)$ 199,920 to come back positive and 9,796,080 to come back negative.

	Test Positive	Test Negative	Total
Have Disease	3996	4	4000
Don't Have Disease	199,920	9,796,080	9,996,000
Total	b		10,000,000

Table Method - Probability of having a disease given a positive test result

We can now get the column totals giving the number of tests we expect to come back positive and negative.

	Test Positive	Test Negative	Total
Have Disease	3996	4	4000
Don't Have Disease	199,920	9,796,080	9,996,000
Total	203,916	9,796,084	10,000,000

We have our answer! The relevant sample space is the 203916 people whose tests are positive. The fraction that have the disease, and thus $P(disease|test+)$ is $3996/203916 = 0.0195963$

Table Setup Tips

To set up the table:

- Label the rows after your event of interest (and its competing alternatives; it's ok to have "A" and to lump everything else into "not A"). The problem will give you prior probabilities for each of them (they will sum to 1).
- Label the columns based on what you're using in the conditioning – list all ways this info could turn out. It's ok to have "B" and lump everything else into "not B".
- Consider a large enough number of trials/individuals so that you won't have to do any rounding once cell total calculations take place. Put that number in the lower right (grand total).

Table Completion Tips

To complete the table:

- Calculate the row totals using the prior probabilities of your event of interest and its alternatives.
- For each row, fill in the cell totals based on the row total and the additional probabilities that are given to you.
- Fill in the column totals.
- Identify the sample space you need to use in the calculation (column total correspond to the specific event on which you are conditioning), then calculate the fraction of that sample space that contains your event of interest (simple ratio).

Technically, you can usually skip filling in cell totals for all except the one column that corresponds to the specific event on which you are conditioning.

Table Method - Monty Hall (you pick door A)

Without loss of generality, assume you picked door A. What is the probability you win if you stick with your original choice of A? Let's use a table to see what happens over 300 games.

- The rows will be labeled Car Behind A and Goat Behind A (our event of interest and its competing alternative for what lurks behind our door).
- The columns will be labeled Show A, Show B, Show C (the info on which we'll be conditioning). Although Monty will never show A because that's what we've picked, it's included for completion.
- We'll put 300 in the lower right for the grand total.

(We Picked A)	Show A	Show B	Show C	Total
Car Behind A				
Goat Behind A				
Total				300

Table Method - Monty Hall (you pick door A)

The prior probability of us having picked the car is 1/3 (since it is distributed randomly; remember we are always picking A). The prior probability of us having picked a goat is 2/3. Thus, of the 300 games, we expect the car behind our door in 100 of them and a goat behind our door in 200. We fill in the row totals.

(We Picked A)	Show A	Show B	Show C	Total
Car Behind A				100
Goat Behind A				200
Total				300

Table Method - Monty Hall (you pick door A)

For those 100 times the car is behind door A, Monty has a choice of showing what's behind B or C (since they are both goats). We assume he picked randomly, so we expect 50 times he picks door B and 50 times he picks door C.

(We Picked A)	Show A	Show B	Show C	Total
Car Behind A	0	50	50	100
Goat Behind A				200
Total				300

Table Method - Monty Hall (you pick door A)

For those 200 times a goat is behind door A, Monty will show B (if the car is behind C) or will show C (if the car is behind B). Since it's equally likely for the car to be behind either of these doors, so we expect each option to occur 100 times.

	Show A	Show B	Show C	Total
Car Behind A	0	50	50	100
Goat Behind A	0	100	100	200
Total				300

Table Method - Monty Hall (you pick door A)

Filling in the column totals, we have our answer. If Monty shows what's behind door B, then our sample space is the 150 games in the "Show B" column. A car is behind our door in 50 of them, and a goat is behind our door in 100 of them. Thus, the probability we win by staying with door A is $P(car\ A|ShowB) = 50/150 = 1/3$, and the probability we win by switching is $1 - 1/3 = 2/3$. The same conclusion is reached if Monty shows what's behind door C.

	Show A	Show B	Show C	Total
Car Behind A	0	50	50	100
Goat Behind A	0	100	100	200
Total	0	150	150	300

Table Method - Digit Recognition

Post offices have machines that automatically pick out the numbers in zip codes. Imagine that all digits are equally likely to appear and the classifier claims the digit is a 3. What is the probability the digit is actually a 3, i.e. P(actual 3 | claim 3)? Imagine the classifier is 91% accurate, and when a misclassification occurs it is more likely to classify it as an 8 than as any other digit.

- If the digit is a 3, the machine claims it is a 3 91% of the time.
- If the digit is an 8, the machine claims it is a 3 2% of the time.
- If the digit is a 0-2, 4-7, or 9, the machine claims it is a 3 0.8% of the time.

Digit Recognition Table

Let's create a table of 10000 digits to get this probability. Each are equally likely, so the prior probability of each digit 0-9 is 10%. Thus, we expect 1000 of each. Ultimately we want the ratio a/b (of all the times the machine claims 3, what fraction of the time is it actually a 3).

	Claim 0	Claim 1	Claim 2	Claim 3	Claim 4	etc.	Total
Actual 0							1000
Actual 1							1000
Actual 2							1000
Actual 3				a			1000
Actual 4							1000
Actual 5							1000
Actual 6							1000
Actual 7							1000
Actual 8							1000
Actual 9							1000
Total				b			10000

Digit Recognition Table

Big table, so let's only fill in what we need – the single column corresponding to the event on which we are conditioning (machine claims it is a 3). Thus we need the number of 0s that the machine will claim are a 3, the number of 1s that the machine will claim are a 3, etc.

	Claim 0	Claim 1	Claim 2	Claim 3	Claim 4	etc.	Total
Actual 0							1000
Actual 1							1000
Actual 2							1000
Actual 3				a			1000
Actual 4							1000
Actual 5							1000
Actual 6							1000
Actual 7							1000
Actual 8							1000
Actual 9							1000
Total				b			10000

Digit Recognition Table

Of the 1000 times the digit is a 3, the machine is expected to claim it is a 3 for (91%) 910 of them. Of the 1000 times the digit is an 8, the machine is expected to claim it is a 3 for (2%) 20 of them. Of the 1000 times the digits is a 0 (and all other digits), the machine is expected to claim it is a 3 for (0.8%) 8 of them.

	Claim 0	Claim 1	Claim 2	Claim 3	Claim 4	etc.	Total
Actual 0				8			1000
Actual 1				8			1000
Actual 2				8			1000
Actual 3				910			1000
Actual 4				8			1000
Actual 5				8			1000
Actual 6				8			1000
Actual 7				8			1000
Actual 8				20			1000
Actual 9				8			1000
Total				b			10000

Digit Recognition Table

We have our answer. The machine is expected to claim a digit is a 3 a total of 994 times, so this is the appropriate sample space to use in the probability calculation. $P(actual\, 3|claim\, 3) = 910/994 = 0.915493$.

	Claim 0	Claim 1	Claim 2	Claim 3	Claim 4	etc.	Total
Actual 0				8			1000
Actual 1				8			1000
Actual 2				8			1000
Actual 3				910			1000
Actual 4				8			1000
Actual 5				8			1000
Actual 6				8			1000
Actual 7				8			1000
Actual 8				20			1000
Actual 9				8			1000
Total				994			10000

Tree method for conditional probability calculations

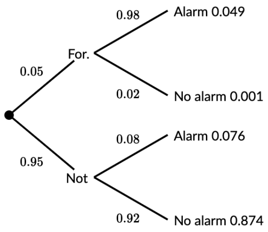
The Khan academy has an excellent collection of instructional material for both basic probability rules. One approach it introduces to help with conditional probability calculations is making a tree diagram.

`https://www.khanacademy.org/math/ap-statistics/probability-ap#stats-conditional-probability`

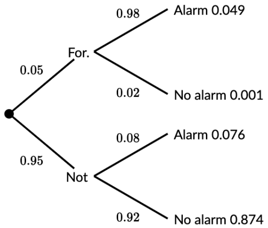
This approach is about half-way between the table approach and the pure math approach, and I recommend trying it out and seeing if that way "clicks" for you. You're more than welcome to use it on homework/exams!

Tree method for conditional probability calculations

An airport security alarm is supposed to detect if a bag contains a forbidden item. 98% of bags with forbidden items trigger the alarm, while 8% of bags with no forbidden items trigger the alarm. From previous experience, 5% of bags contain forbidden items. If the alarm goes off, what's the probability the bag has a forbidden item?



Tree method for conditional probability calculations



$$P(forbidden|alarm) = \frac{0.049}{0.049 + 0.076} = 0.392$$

Math Behind Conditional Probability Calculations

Notation

Let A and B be two events

- A = apples are in the shopping cart
- B = bananas are in the shopping cart

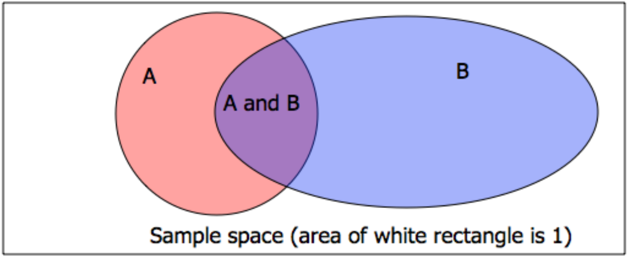
Then the conditional probability that A occurs, given that B has occurred, is denoted as $P(A|B)$ (probability that apples are in the cart given that bananas are there) can be written:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

- $P(A \text{ and } B)$ is the probability that *both* A and B occur (both apples and bananas are in the cart) and is called the **joint** probability of A and B.
- $P(B)$ is the probability that B occurs in general (without regard to A) and is called the **marginal** or **unconditional** probability of B.

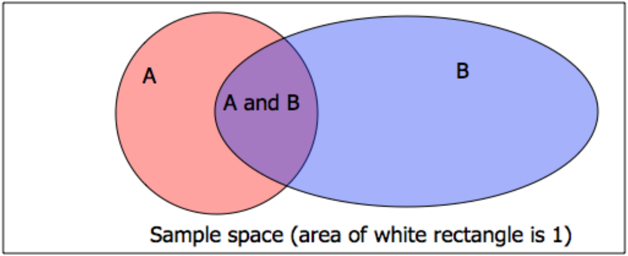
Visual derivation of formula

The formula is remarkably easy to derive visually with a Venn diagram.



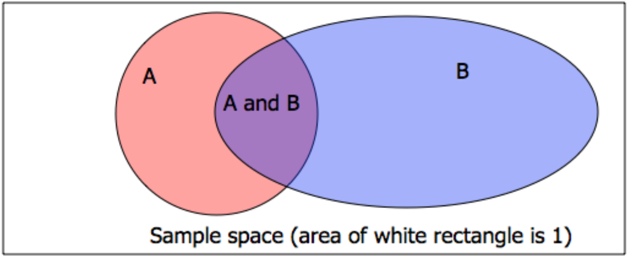
With no other information, all we know is that after a trial occurs we'll end up *somewhere* in the sample space. Thus, the prior probability that A occurs $P(A)$ is the fraction of the sample space contained in A. That's the area of the circle representing A.

Visual derivation of formula



However, when we know B has occurred, we know that the only places in the sample space that are possible to be are inside the blue oval. The posterior probability of A, given B has occurred, is the fraction of the blue oval that contains event A.

Visual derivation of formula



The area of the blue oval that contains event A corresponds to the probability "$P(A \text{ and } B)$". The area of the blue oval itself is $P(B)$. Thus:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Reduction of the sample space

As discussed in the section on the "Intuitive Approach for Calculating Conditional Probabilities", the key to understanding this formula (and all of conditional probability) is realizing that a conditional probability is just a probability with respect to a different sample space than the one we started out with.

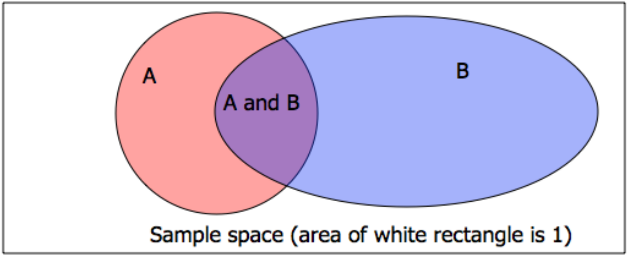
When we calculate the probability of A on the condition that B occurs, we're restricting ourselves to only parts of the original sample space where B occurs. The calculation proceeds as before, just the "reduced" sample space – the posterior probability of A is the fraction of this "reduced" sample space that contains A.

Reduction of the sample space

$$P(A) = \text{ fraction of sample space that contains event A}$$

$$P(A|B) = \text{ fraction of ``reduced'' sample space that contains event A}$$

The "reduced" sample space are all outcomes where B has occurred.



Multiplication Rule (A and B) rediscovered

If we start with our formula for conditional probability:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

and multiply both sides of the equation by $P(B)$, we rediscover the multiplication rule from Unit 1.

$$P(A \text{ and } B) = P(A|B)P(B)$$

Or, swapping the identity of events A and B

$$P(A \text{ and } B) = P(B|A)P(A)$$

Second (more useful) formula for calculating conditional probabilities

We have:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Often P(A and B) is not immediately obvious based on what we know, and it's often more straightforward to compute $P(A|B)$ using it's alternative definition. Since the multiplication rule says $P(A \, and \, B)$ can be re-written as $P(B|A)P(A)$, we have

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

Second (more useful) formula for calculating conditional probabilities

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

- $P(A)$ is called the *prior* probability of A . It is also called the *marginal* or *unconditional* probability of A, and represents the overall frequency that A occurs in general. For example, the fraction of shopping carts that contain apples.
- $P(A|B)$ is called the *posterior* probability of A (given the new information that B has occurred). Of the shopping carts with bananas, the fraction that also have apples.

Example-Numbers

A random integer 1-100 is picked. What's the probability it is even? Given the product of its digits is 36, what is the probability it is even?

- A = even number
- B = product of digits is 36

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

- $P(A) = 0.5$ (prior probability that the number is even)
- $P(B) = 0.03$ (picking either a 49, 66, or 94)
- $P(A \text{ and } B) = 0.02$ (picking either a 66 or 94)
- $P(A|B) = 0.02/0.03 = 2/3$

Intuitively this makes sense. There's only 3 numbers whose digits, when multiplied together, equal 36. These 3 numbers are our "reduced" sample space when conditioning on B. Two of those three numbers are even, so there's a 2 in 3 chance.

Example-Insurance

Example-Insurance

A random person from a database with 100 customers who have the following characteristics is to be picked (this matches up exactly to the Table approach from an earlier section).

	Buys Pet Insurance	Doesn't Buy Pet Insurance	Total
Has Pet	9	70	79
Does Not Have Pet	1	20	21
Total	10	90	100

Let A be "buys insurance" and B be "owns pet"

- $P(A) = 10/100$
- $P(B) = 79/100$
- $P(A \text{ and } B) = 9/100$
- $P(A|B) = P(A \text{ and } B)/P(B) = (9/100)/(79/100) = 9/79$ (of the 79 people with a pet, 9 buy insurance)
- $P(B|A) = P(A \text{ and } B)/P(A) = (9/100)/(10/100) = 9/10$ (of the 10 people who buy insurance, 9 have pets)

Example: librarian vs. waitress

We know a women is either a waitress or a librarian. We also know she typically dresses conservatively and wears glasses. She's quiet, intelligent, and likes to read. Which is more likely: she is a waitress or she is a librarian?

$$P(librarian|description) = \frac{P(description\ and\ librarian)}{P(description)}$$

I'm not sure how to track down data that gives the proportion of women who are both librarians *and* meet the description, so let's use the second formulation.

$$P(librarian|description) = \frac{P(description|librarian)P(librarian)}{P(description)}$$

Example: librarian vs. waitress

$$P(librarian|description) = \frac{P(description|librarian)P(librarian)}{P(description)}$$

$P(librarian)$ is the prior probability of her being a librarian, i.e., given no other information. According to `www.ala.org` and `www.infoplease.com`, there are about 120,000 female librarians and 1,400,000 female waitresses. We know she is one of these two. If we assume its equally likely this woman is randomly plucked from this pool, then the prior probability of her being a librarian is rather small: $P(librarian) = 120000/(120000 + 1400000) = 7.9\%$.

Example: librarian vs. waitress

- Let's assume that stereotypes are true and that 70% of female librarians meet the description. Thus $P(description|librarian) = 0.70$.
- Imagine that the description of the woman (smart, quiet, reads, conservatively dressed, glasses) applies to 15% of the pool of waitresses and librarians. Thus, $P(description) = 0.15$.
- $P(librarian|description) = 0.70 \times 0.079/0.15 = 0.37$.
- The posterior probability of her being a librarian given updated information is less than 50%. Still more likely she's a waitress!
- Full disclosure: my values for $P(description|librarian) = 0.70$ and $P(description) = 0.15$ are guesses (not based on evidence). If you disagree with these values, try out some of your own and see if the answer changes much.

Independence revisited

A and B are **independent** events if the posterior probability of A (given B has occurred) is the same as the prior probability of A (and vice versa). In other words, knowledge of whether B did or did not occur is irrelevant to the probability that A occurs.

$$P(A|B) = P(A) \qquad \text{when } A \text{ and } B \text{ independent}$$
$$P(B|A) = P(B) \qquad \text{when } A \text{ and } B \text{ independent}$$

Independence revisited

Examples of independent events:

- A is the event "11 is picked in the Powerball this week". B is the event "11 is picked in the Powerball next week".
- A is "the closing price of Apple stock is greater than \$200" and B is "the daily high temperature of Knoxville is 92".

Examples of nonindependent events:

- A is the event "buying gas next week" and B is the event "owns a car".
- A is the event "wears shorts" and B is the daily high temperature in Knoxville.

How the Independence Math Work Out

When A and B are **independent** events, then $P(A \text{ and } B) = P(A)P(B)$.

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

$$P(B|A) = \frac{P(B \text{ and } A)}{P(A)} = \frac{P(B)P(A)}{P(A)} = P(B)$$

Bayes Theorem and Partitioning

Bayes Theorem

The second formula for calculating conditional probability has a special name: **Bayes' Theorem**. It provides a mathematical method for updating the probability that an event A will occur when new information B becomes available.

$$P(A|B) = \frac{P(A \, and \, B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

Bayes Theorem

$$P(A|B) = \frac{P(A \, and \, B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

- $P(A)$ is called the *prior* probability of A. It is also called the *marginal* or *unconditional* probability of A, and represents the overall probability that A occurs in general.
- $P(A|B)$ is called the *posterior* probability of A (given the new information that B has occurred)
- $P(A \, and \, B)$ is called the *joint* probability of A and B while $P(B)$ is the *marginal* or *unconditional* probability of $P(B)$

In previous illustrations, finding $P(B)$ was straightforward. Often, it requires employing a "trick".

Partitioning

Often, the denominator in Bayes' Theorem $P(B)$ isn't straightforward to calculate.

- P(have disease | test is positive) requires P(test is positive) to be calculated. However, the probability this event depends on the accuracy of the test and whether the patient has the disease.
- P(car behind door A | goat revealed behind door B) requires P(goat revealed behind door B) to be calculated. But this probability depends on what the person has chosen and the location of the car.

Partitioning is a powerful trick for calculating these "but that depends on ..." probabilities that we find in the denominator of Bayes' Theorem.

Visual Partitioning - Winning a Game again Random Opponent

An algorithm randomly selects your next opponent in a video game. There are three possibilities: hard, average, easy. The probability you win against the opponent depends on its difficulty.

Opponent Type	Hard	Average	Easy
Probability of **selection**	0.20	0.30	0.50

Opponent Type	Hard	Average	Easy
Probability of **winning**	0.10	0.50	0.80

Visual Partitioning - Winning a Game again Random Opponent

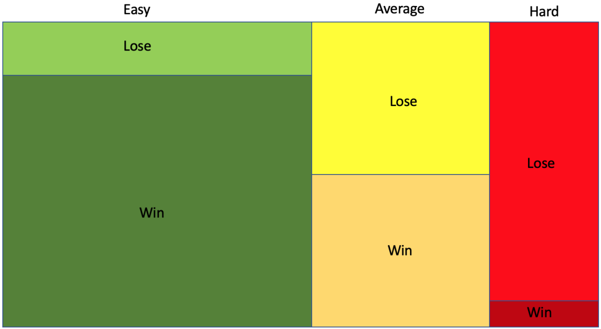
You are unaware of the difficult of your chosen opponent. Given that you beat your opponent, what is the probability that the difficulty of that opponent was "hard". In other words, what is the fraction of your wins that occur against "hard" opponents?

$$P(hard|win) = \frac{P(win|hard)P(hard)}{P(win)}$$

We see that to compute this probability, we need to determine the probability of winning against a randomly picked opponent. But the probability of winning depends on the difficulty of the opponent, so how do we proceed?

Visual Partitioning - Winning a Game again Random Opponent

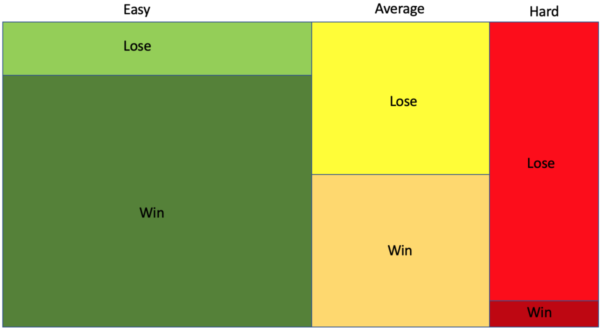
Let's work through it visually. The rectangle gives the sample space. 50% of the sample space is colored a shade of green (since there's a 50% chance of playing an easy opponent). 30% of the sample space is colored a shade of yellow, and 20% of the sample space is colored a shade of red. The darker shades of each color correspond to the event of winning.



Outer rectangle is Sample Space

Visual Partitioning - Winning a Game again Random Opponent

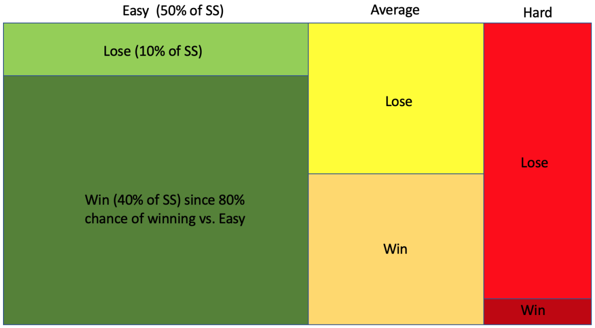
The probability of winning can be found by summing up the areas devoted to the "win" event, and there's three different darker areas to calculate, one for each opponent type. P(win) = P(win and easy) + P(win and average) + P(win and hard).



Outer rectangle is Sample Space

Visual Partitioning - Winning a Game again Random Opponent

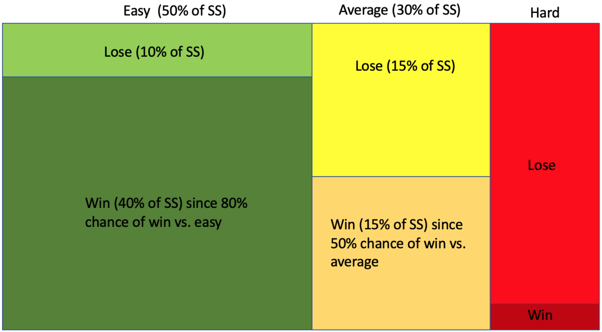
What's the area of the dark green (getting an easy opponent and winning)? According to the description, we end up with an easy opponent 50% of the time (the dual-green-shaded rectangle has an area of 0.5). Then, 80% of those times we win, so 80% of 50% of the sample space has us winning against an easy opponent. $0.8 \times 0.5 = 0.4$



Outer rectangle is Sample Space

Visual Partitioning - Winning a Game again Random Opponent

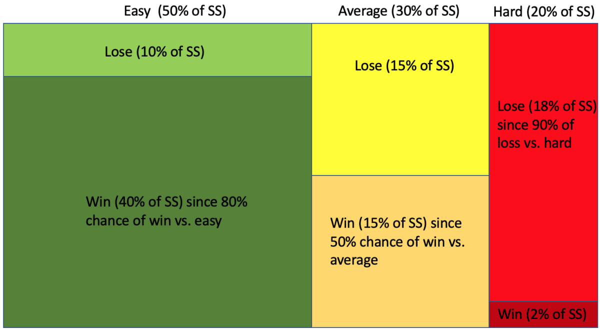
What's the area of the dark yellow (getting an average opponent and winning)? We end up with an average opponent 30% of the time (the dual-yellow-shaded rectangle has an area of 0.3). Then, 50% of those times we win, so 50% of 30% of the sample space has us winning against an average opponent.
$0.5 \times 0.3 = 0.15$



Outer rectangle is Sample Space

Visual Partitioning - Winning a Game again Random Opponent

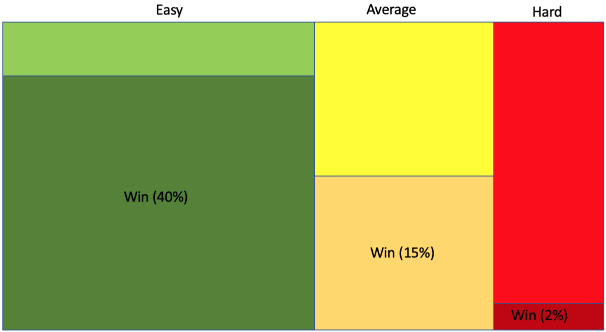
What's the area of the dark red (getting a hard opponent and winning)? We end up with a hard opponent 20% of the time (the dual-red-shaded rectangle has an area of 0.2). Then, 10% of those times we win, so 10% of 20% of the sample space has us winning against a hard opponent. $0.1 \times 0.2 = 0.02$



Outer rectangle is Sample Space

Visual Partitioning - Winning a Game again Random Opponent

So, the overall probability of winning is $0.4 + 0.15 + 0.02 = 0.57$



Outer rectangle is Sample Space

Mathematical Partitioning - Winning a Game again Random Opponent

Looking back, how did we get $P(win)$? First, we realized it depended on the difficulty of the opponent. The overall probability of winning is the sum of the probabilities of winning against each opponent:

$$P(win) = P(Win\,and\,Hard) + P(Win\,and\,Average) + P(Win\,and\,Easy)$$

Implicitly, we then used the multiplication rule. To find $P(Win\,and\,Easy)$, we multiplied two probabilities: $P(Win|Easy)P(Easy)$. 50% of the time we get an easy opponent, then we win 80% of those matches.

$$P(win) = P(Win|Hard)P(Hard) + P(Win|Average)P(Average) + P(Win|Easy)P(Easy)$$

Mathematical Partitioning - Winning a Game again Random Opponent

Opponent Type	Hard	Average	Easy
Probability of selection	0.20	0.30	0.50

Opponent Type	Hard	Average	Easy
Probability of winning	0.10	0.50	0.80

$$P(win) = P(Win|Hard)P(Hard) + P(Win|Average)P(Average) + P(Win|Easy)P(Easy)$$

$$P(win) = 0.10 \cdot 0.20 + 0.50 \cdot 0.30 + 0.80 \cdot 0.50 = 0.57$$

Calculating $P(win)$ by breaking it down by opponent is called **partitioning**.

Partitioning Basics

Let B be the event that we want the probability of, and let $E_1, E_2, \ldots, E_k$ be a set of *mutually exclusive* events, exactly one of which *must* occur. These events spell out exactly what events make up the "but that depends on ..." complication for finding $P(B)$.

$$P(B) = \text{sum over } i \text{ of } P(B|E_i)P(E_i)$$

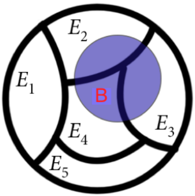
For example, if B is the event "win", then we can partition on "opponent difficulty" (easy/average/hard, we get exactly one of these).

$$P(win) = \text{sum over all difficulties of } P(win|difficulty)P(difficulty)$$

$$P(win) = P(win|easy)P(easy) + P(win|average)P(average) + P(win|hard)P(hard)$$

Visualization of Partitioning Formula

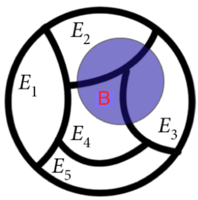
Let the sample space be represented with the circle below. B is the event of interest and E represents our partition (the set E_1, E_2, ..., E_5; exactly one of which will occur). Since $P(B)$ corresponds to the area of the purple shaded region, we can sum up the areas of its component purple regions where B and E overlap.



$$P(B) = P(B \text{ and } E_1) + P(B \text{ and } E_2) + \ldots + P(B \text{ and } E_5)$$

In this case, the purple region doesn't overlap E_1 or E_5, so those parts of the sum equal 0.

Justification



$$P(B) = P(B \text{ and } E_1) + P(B \text{ and } E_2) + \ldots + P(B \text{ and } E_5)$$

But the multiplication rule says $P(B \text{ and } E_1) = P(B|E_1)P(E_1)$, etc., so:

$$P(B) = P(B|E_1)P(E_1) + P(B|E_2)P(E_2) + \ldots + P(B|E_5)P(E_5)$$

$$P(B) = \text{sum over } i \text{ of } P(B|E_i)P(E_i)$$

Mathematical Partitioning - Winning a Game again Random Opponent

The probability of winning depends on what type of opponent you face. So to answer the question, we need to "partition on opponent type". Let's start with the partitioning formula and translate

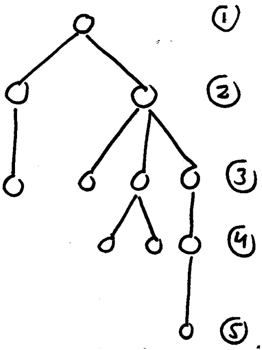
$$P(B) = \text{sum over } i \text{ of } P(B|E_i)P(E_i)$$

- Event B is winning
- E represents the partitioning. Since winning depends on opponent type, our partition is the set of all possible opponents.
- E_1 = hard opponent selected
- E_2 = average opponent selected
- E_3 = easy opponent selected

$$P(win) = P(win|easy)P(easy) + P(win|average)P(average) + P(win|hard)P(hard)$$

Mathematical Partitioning - Branching Process and Ultimate Extinction

Imagine that you possess the so-called "luck gene" and are the only person in the world to have it. It is a dominant gene so that all your offspring inherit it, and all of theirs, etc. What is the probability that the luck gene will persist forever or will eventually go extinct?



In this example the gene persisted through 5 generations, then died out.

Mathematical Partitioning - Ultimate Extinction

Surprisingly, you can find this probability by partitioning!

For simplicity, assume that the number of kids you produce (and the number of kids *those* will produce, etc.) has the distribution:

kids	$P(kids)$
0	0.10
1	0.50
2	0.20
3	0.20

Average number of kids: 1.5.

Mathematical Partitioning - Ultimate Extinction

Let's denote the probability of extinction as p. This probability of course depends on the number of kids you have. The number of kids you have forms a partition, so

$$P(extinct) = \sum_{i=0}^{3} P(extinct|kids = i)$$

$$P(extinct) = P(extinct|kids = 0)P(kids = 0) + P(extinct|kids = 1)P(kids = 1)$$
$$+P(extinct|kids = 2)P(kids = 2) + P(extinct|kids = 3)P(kids = 3)$$

- $P(extinct|kids = 0) = 1$, since no kids are produced
- $P(extinct|kids = 1) = p$, since if one kids is produced we are in the same position as we are now once you die out.
- $P(extinct|kids = 2) = p^2$, since *both* lines of kids must go extinct (note: we have to assume the lines are independent for this formula to work)
- $P(extinct|kids = 3) = p^3$ as above.

Mathematical Partitioning - Ultimate Extinction

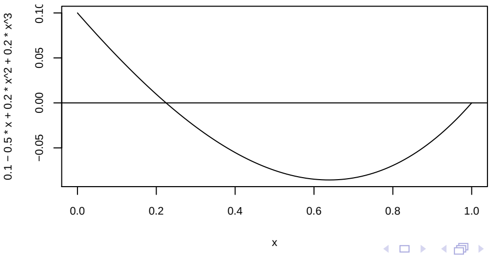
Remarkably, this problem is solved by using just a little bit of algebra!

$$p = 1 \times 0.1 + p \times 0.5 + p^2 \times 0.2 + p^3 \times 0.2$$

Thus, the probability of extinction is the smallest value of p that solves

$$0.1 - 0.5p + 0.2p^2 + 0.2p^3 = 0$$

```
curve(0.1-0.5*x+0.2*x^2+0.2*x^3);  abline(h=0)
#around 0.2?
```



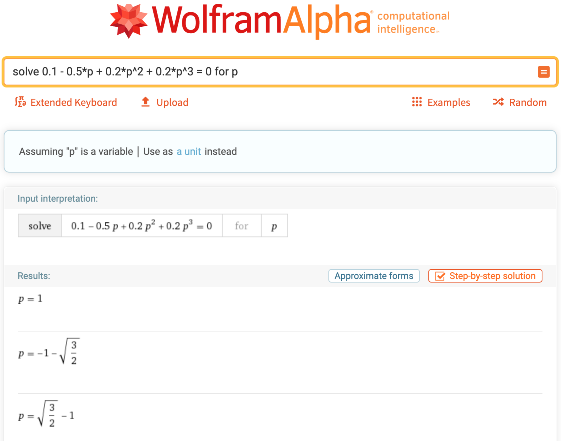
Mathematical Partitioning - Ultimate Extinction

An exact answer can be found using `uniroot()`, a function that determines at what value of x and function of x equals 0 (have to it a range to search).

```
f <- function(x) { 0.1-0.5*x+0.2*x^2+0.2*x^3 }
uniroot(f,lower=0,upper=0.5) #it's clear the answer is somewhere between 0 and 0.5
## $root
## [1] 0.2247514
##
## $f.root
## [1] -2.479747e-06
##
## $iter
## [1] 5
##
## $init.it
## [1] NA
##
## $estim.prec
## [1] 6.103516e-05
```

Mathematical Partitioning - Ultimate Extinction

Wolfram-Alpha (`https://www.wolframalpha.com/` solves equations like a champ.



Mathematical Partitioning - Ultimate Extinction

Wolfram-Alpha is sometimes "too" helpful, giving solutions that won't work for us. After some investigating, we find the exact probability that the luck gene becomes extinct as $p = \sqrt{3/2} - 1$.

```
1-sqrt(3/2)  #Not a valid probability
## [1] -0.2247449
sqrt(3/2) - 1   #That's the one!
## [1] 0.2247449
```

Disease

A rare disease affects 0.04% of the US population. Tests are extremely accurate. If you have the disease, the test comes back positive 99.9% of the time. If you do not have the disease, the test comes back negative 98% of the time. You just took the test and it came back positive. What is the probability that you have it?

We will use Bayes Theorem to update the prior probability of having the disease (0.04%) to a posterior probability taking into account information from the test (that it came back positive).

Disease

$$P(infected|test+) = \frac{P(test+|infected)P(infected)}{P(test+)}$$

- $P(infected)$ is the prior probability of having the disease (0.04% or 0.0004)
- $P(infected|test+)$ is the posterior probability of having the disease given the test comes back positive.
- $P(test+|infected)$ is probability the test comes back positive when the person is infected: 99.9% or 0.999
- $P(test+) =?$ Well that depends on your disease status. Partition!

$P(test+)$

Since testing positive depends on disease status, we need to partition on its possible values. There are two options for disease status: you either are infected or you are not.

$$P(test+) = P(test+|infected)P(infected) + P(test+|not\ infected)P(not\ infected)$$

- $P(test+|infected) = 0.999$
- $P(infected) = 0.0004$
- $P(test+|not\ infected) = 0.02$ (complement rule, since 98% of coming back negative given not infected)
- $P(not\ infected) = 0.9996$ (complement rule, since 0.04% chance of being infected)

$$P(test+) = 0.999 \times 0.0004 + 0.02 \times 0.9996 = 0.0203916$$

Disease solution

$$P(infected|test+) = \frac{P(test+|infected)P(infected)}{P(test+)}$$

$$P(infected|test+) = \frac{0.999 \times 0.0004}{0.999 \times 0.0004 + 0.02 \times 0.9996} = 0.0195963$$

We recover the same value as when we did this using the Table Approach.

While the prior probability of being infected is 0.04%, the additional information of the test coming back positive has updated this to a posterior probability of only about 2%. This still sounds very small, but the probability of having the disease did increase by a factor of 50 from 0.0004 to 0.02.

Digit Recognition

Post offices have machines that automatically pick out the numbers in zip codes. Imagine that all digits are equally likely to appear and the classifier claims the digit is a 3. What is the probability the digit is actually a 3? Imagine the classifier is 91% accurate, and when a misclassification occurs it is more likely to classify it as an 8 than as any other digit.

$$P(actual\,3|claim\,3) = \frac{P(claim\,3|actual\,3)P(actual\,3)}{P(claim\,3)}$$

- $P(actual\,3) = 0.10$, since digits 0-9 are equally likely.
- $P(claim\,3|actual\,3) = 0.91$ (91% accuracy)
- $P(claim\,3)$ is elusive, since that will depend on what the digit actually is.

Digit Recognition

Since the actual digit forms a partition (a digit *must* be 0-9 and the options are mutually exclusive), we have

$$P(claim\,3) = P(claim\,3|actual\,0)P(actual\,0) + \ldots + P(claim\,3|actual\,9)P(actual\,9)$$

We see the answer depends on how often a digit gets misclassified as a 3. Imagine $P(claim\,3|actual\,i) = 0.008$ for all i except 3 and 8, and that $P(claim\,3|actual\,8) = 0.02$. In other words, the classifier misclassifies a 4 as a 3 0.8% of the time and the classifier misclassifies an 8 as a 3 2% of the time.

Also, $P(claim\,3|actual 3) = 0.91$.

Digit Recognition

Rearranging terms:

$$P(claim\,3) = 8 \times (0.008 \times 0.10) + (0.02 \times 0.10) + (0.91 \times 0.10) = 0.0994$$

Therefore:

$$P(actual\,3|claim\,3) = \frac{P(claim\,3|actual\,3)P(actual\,3)}{P(claim\,3)} = \frac{0.91 \times 0.10}{0.0994} = 0.9155$$

We recover the same value as when we did this using the Table Approach.

Why are we terrible at conditional probability?

Why are we terrible at conditional probability?

Our brains are amazing at some things it's extremely difficult for computers to do (e.g., pattern recognition, part of the reason why CAPTCHAs exist), but we don't intuitively process conditional probability correctly.

Prof Petrie's theory: if we are interested in the probability that A occurs, once we are given new information and *should* be calculating $P(A|B)$ we instead intuitively calculate $P(B|A)$ instead and feel it's correct. We outright ignore the prior probability that A occurs, or have trouble incorporating it into our calculation. It seems like our brain is wired backwards for the calculation!

Why we are terrible at conditional probability

Disease example:

- We want P(have disease | test positive).
- Our brain seems to think P(test positive | have disease) is a reasonable answer.
- Our brain didn't take into account how rare the disease was (i.e., the prior probability). Even though the test is very accurate, the disease is so rare that a positive result doesn't mean much.

Why we are terrible at conditional probability

There's nothing *inherently* tricky about calculating a conditional probability. As we saw in the visual and table solution to the disease example, $P(A|B)$ is just the fraction of *some* sample space that contains A, specifically the part of the original sample space where B occurs.

And there's the problem. Our brains are not good at "reducing" the sample space when conditioning on an event. We have trouble seeing what the sample space where B occurs looks like, and thus cannot adequately anticipate what fraction of that sample space contains A.