BAYES THEOREM

A Visual Introduction For Beginners

DAN MORRIS

Bayes' Theorem:

A Visual Introduction for Beginners

Dan Morris

***Bayes' Theorem: A Visual Introduction for Beginners* by Dan Morris**

Published by Blue Windmill Media

For permissions contact:
matthew@bluewindmill.co
Cover by Blue Windmill Media

Contents

Bayes Everyday

If you've recently searched for something on Google, Bayes' Theorem was used to display your search results.

The same is true for those recommendations on Netflix.

Hedge funds? Self-driving cars? Search and rescue?

Bayes' Theorem is used in all of the above and more.

At its core, Bayes' Theorem is a simple mathematical formula that has revolutionized how we understand and deal with uncertainty. If life is seen as black and white, Bayes' Theorem helps us think about the gray areas. When new evidence comes our way, it helps us update our beliefs and create a *new belief*.

Ready to dig in and visually explore the basics? Let's go.

Don't Waste Your Time

A few points to help you make the most of this booklet:

1. CHANGE YOUR SETTINGS. If you are reading on a Kindle E-reader, change your settings! It will make reading this booklet MUCH EASIER. For spacing, select the mid-range option (on Kindle Paperwhite, the middle option). For margins, select the widest option. For orientation, select portrait. For font, select Bookerly, and font size select the third smallest option (at least on the Kindle Paperwhite).

2. This booklet is designed as a visual introduction to Bayes' Theorem. It is for BEGINNERS. If you already have a general understanding of the Theorem, you might not get much out of this book.

3. You don't need to read front to back! Skip around to what you find the most helpful.

Welcome To This Guide

Welcome to *Bayes' Theorem: A Visual Introduction for Beginners*. This booklet is packed with examples and visual aids to help clarify *what* Bayes' Theorem is and *how it* works. At its core, Bayes' Theorem is very simple and built on elementary mathematics.

Before we dig into different definitions, it needs to be stated that Bayes' Theorem is often called Bayes' Rule or Bayes' Formula. So, don't be confused - they are the same, and we will be using both theorem and formula throughout this booklet. Second, we need to make it clear that Bayes' Theorem is a law of probability theory. It helps us work with, revise, and understand probabilities when we are presented with new evidence.

Practically speaking, the theorem helps us quantify or *put a number* on our skepticism and make more informed, rational choices. It helps us answer the following:

When we encounter new evidence, how much should it change our confidence in a belief?

Here are a few examples:

- You just had a test for cancer and it came back positive. What is the probability that you have cancer if the test is positive?
- Your friend has a new dog and when you visit she slobbers all over you, but does that mean the dog likes you? What is the probability that the dog likes you given that she licks you?
- Your friend claims that stock prices will decrease if interest rates increase. What is the probability stock prices will decrease if interest rates increase?
- You were just pulled over by the police and given a breathalyzer test. It came back positive. What is the probability you are truly drunk, given that the test is positive?

Bayes' Theorem Explained: 4 Ways

Here are a few ways Bayes' Theorem can be explained:

1. Bayes' Theorem helps us update a belief based on new evidence by creating a *new belief*.
2. Bayes' Theorem helps us revise a probability when given new evidence.
3. Bayes' Theorem helps us change our beliefs about a probability based on new evidence.
4. Bayes' Theorem helps us update a hypothesis based on new evidence.

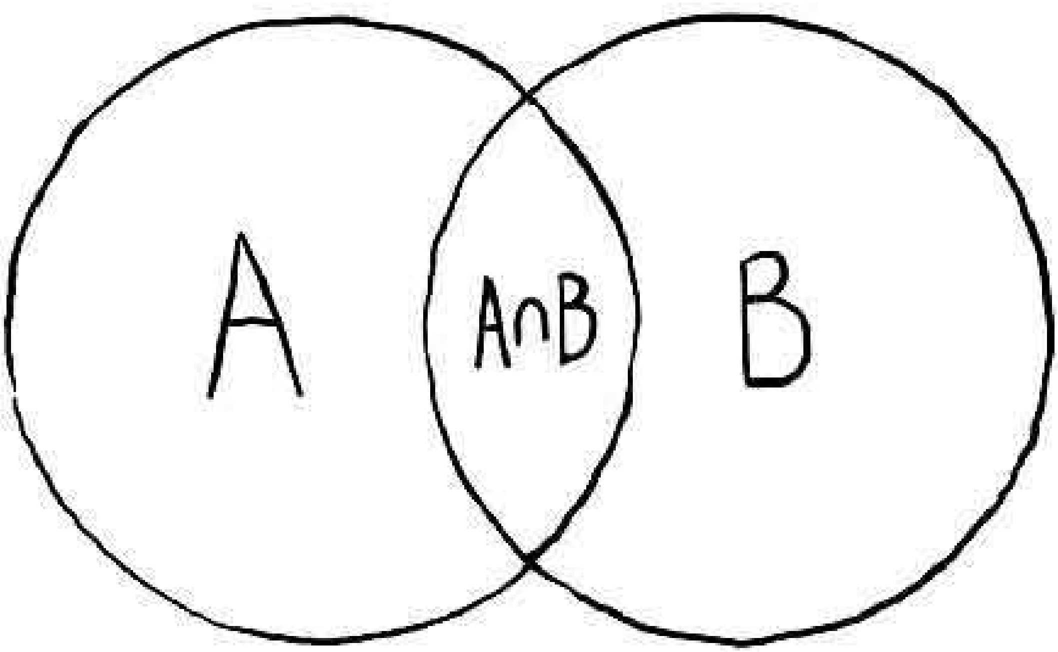
The only problem? Applying the theorem is not intuitive, at least not for most people. This is where visualizing a problem that entails using Bayes' Theorem can be a BIG HELP.

Visual Aids

When working with small amounts of data there are a few different visual aids you can use:

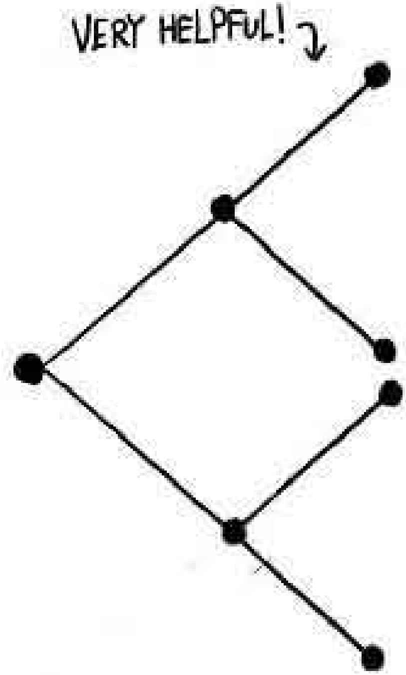
- Venn diagrams
- Decision trees
- Letters (e.g., H, T, T, T for head/tail coin flips)
- Physical Objects (e.g., real coins)

In this booklet, we'll be using Venn diagrams and decision trees. Venn diagrams are an excellent way to help us visually understand problems.



On the other hand, decision trees are a great tool that can help us solve problems where probabilities are not provided and must be discovered.

They are a fantastic visual aid that, for many students of probability, can help unblock areas of frustration and confusion.



Booklet Structure

Here is how we have structured this booklet:

1. **There are two Problem and Answer sections and a bonus section.** Each section deals with a popular way Bayes' Theorem is typically presented in introductory textbooks. **All numbers provided in each problem are hypothetical. Some are pegged close to realistic statistics, but for the sake of clarity and ease we have adjusted each number.*

2. **Each Problem and Answer section has three scenarios.** To help you fully understand each question type, we will work through 3 scenarios. The idea behind this to help drive the problem-solving process home. The scenarios include the flu, breathalyzer test, and peacekeeping.

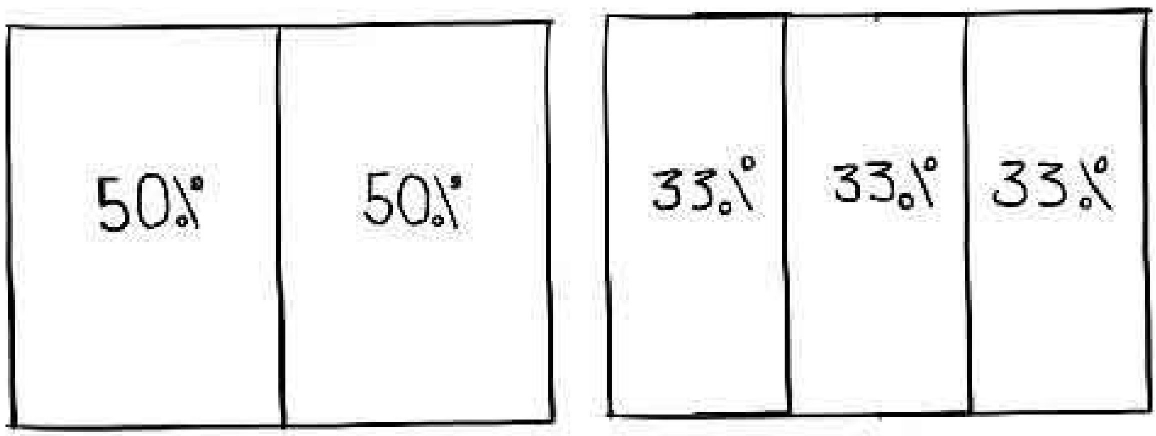
3. **And for those who are interested,** we have included a few additional sections on understanding scenarios, the history of Bayes' Theorem, fascinating stories and uses, and even a few tips on how to think like a Bayesian in real life.

4. **We have also included** an extended definitions section, which includes a Proof for Bayes' Theorem.

5. **To round out this booklet,** we have included a list of recommended readings for those looking to dive deeper into the Theorem and learn more about its fascinating uses and history.

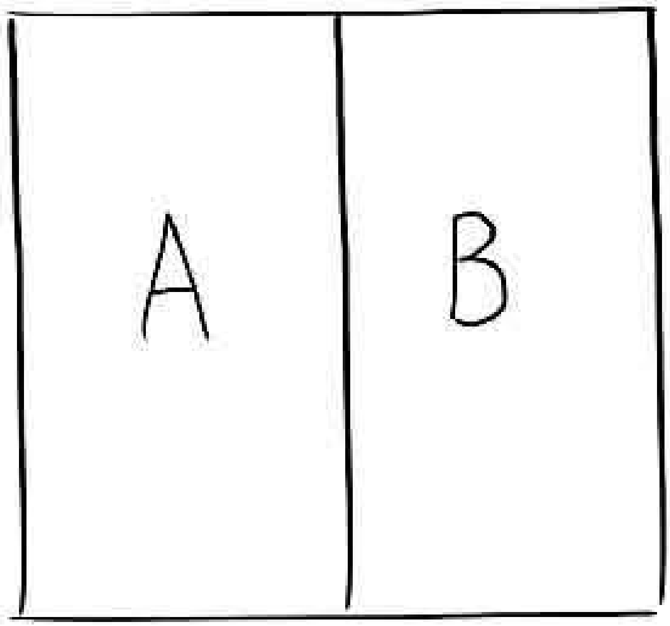
A Visual Intro: Part 1

If you are confused with the *concept* of Bayes' Theorem, this is a fantastic place to start. Before we dive into Example Section 1, let's take a look at Bayes' Theorem *without* using the formula. This visual introduction is an expanded version of this excellent video.



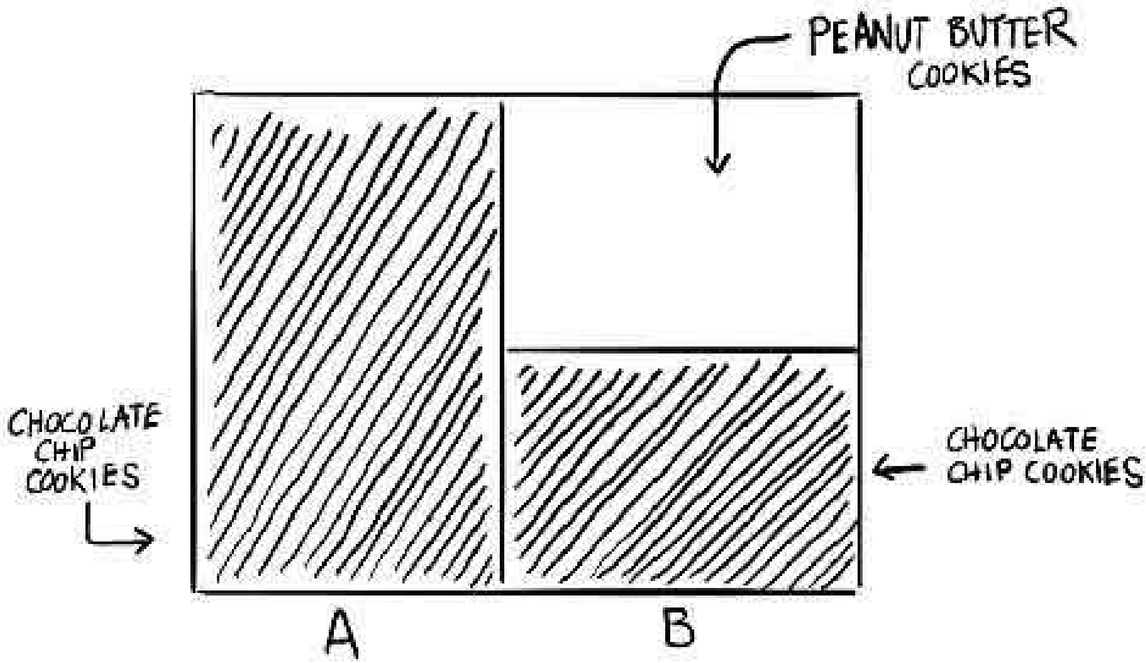
To begin, let's draw a rectangle. Don't get hung up on the shape - it could be any shape, but a rectangle is easy to work with. The area *inside* the rectangle represents all possible outcomes for our experiment. For example, if there are two possible outcomes that are equally probable, we would divide the rectangle into two equal halves. This means that each outcome has a 50% likelihood of occurring. Or, if there are three possible outcomes that are all equally probable, we would divide the rectangle into equal thirds. This would mean that each outcome has a 33.33% chance of occurring.

For this example, we are going to stick with two equal outcomes and title each outcome A and B, respectively.


A
B

Now, imagine that each probability represents a small cardboard box. Box A is filled with 10 chocolate chip cookies. There is nothing else in Box A except home baked, warm, mouth-watering chocolate chip cookies. To demonstrate this, we will shade in Box A. In Box B there are also cookies, but there are two different types. There is an even mix of 5 peanut butter cookies and 5 chocolate chip cookies.

To demonstrate this, we will draw a line and cut the box in half, and then shade in the chocolate chip cookies in the bottom half of the rectangle. We will leave the top half blank.



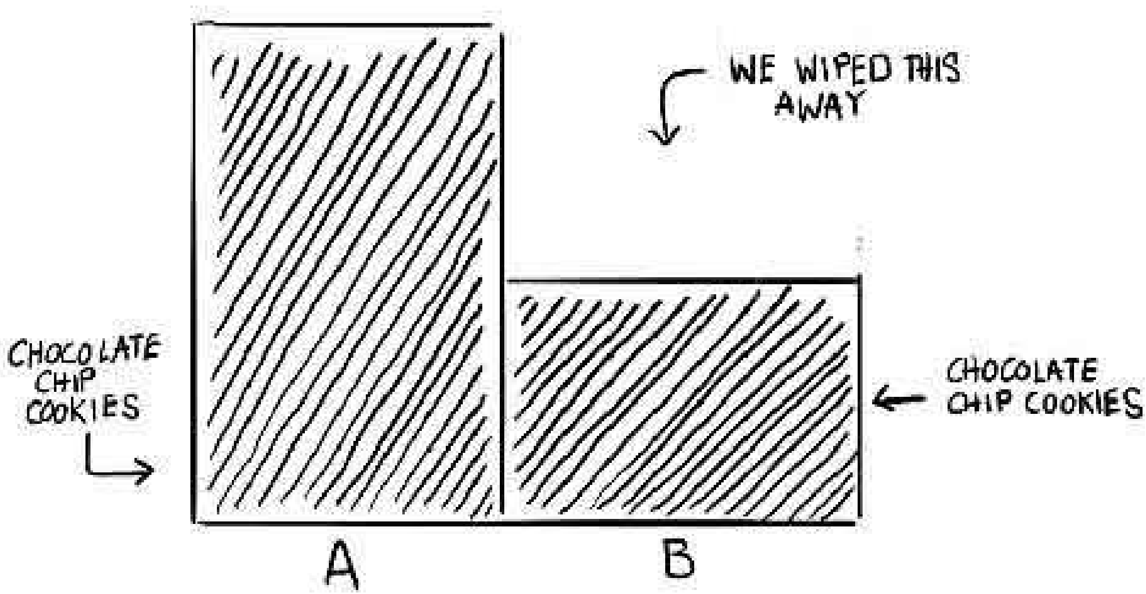
Let's step back and look at the rectangle. Can you see the chocolate chip cookies in the shape of an L? Those areas represent all of our chocolate chip cookies in both boxes, while the white area represents the peanut butter cookies.

Now, what if you were to close your eyes and have both boxes shuffled, and then reach and select a cookie from one of them - and it was chocolate chip?

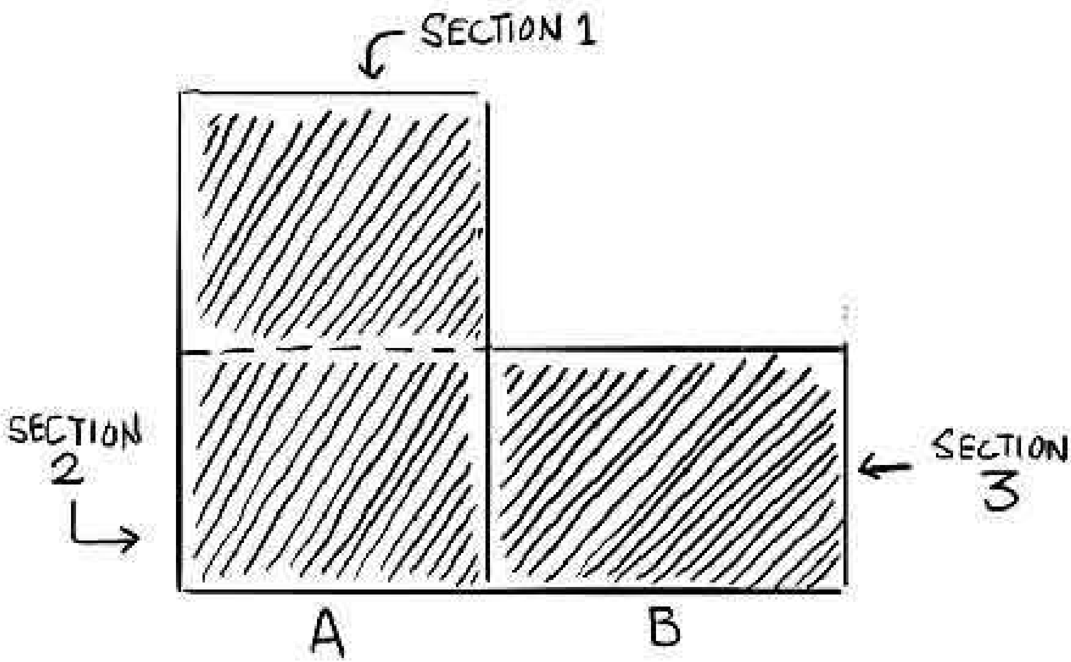
If you had to guess what box the cookie came from, what box would you select? Many people would select Box A, and we'll take a hunch that you are one of them. But let's take a closer look at why this is. Both Box A and B have chocolate chip cookies, but Box A has exactly *double* the amount of chocolate chip cookies than Box B. Within a split second your brain assessed this and came away with the conclusion that Box A has a greater probability of being selected than Box B.

Here's the magic! This calculation is a *very basic*, natural use of Bayes' Theorem. Given evidence (the amount and type of cookies in each box), you were able to quickly come to the conclusion that Box A has a greater probability of being selected than Box B.

Now, let's step back once more. When your hand selected a chocolate chip cookie, something disappeared: the probability of selecting a peanut butter cookie is now gone. So, to visualize this, let's wipe away the portion of Box B that represents the peanut butter cookies.



Our boxes are now in the shape of an L, and we can also see that there are *double* the amount of cookies in Box A than Box B. In fact, if we were to break the Boxes apart into equal sections, we would have 3 areas: 2 sections in Box A, and 1 section in Box B.



By looking at this, we can see that Box B has a probability of ⅓ , or ~33% of being selected. Box A has a probability of ⅔, ~66% of being selected. This difference in probability is what your brain *roughly* calculated before and the whole reason why you selected Box A. Your Brain looked at ~33% vs ~66% and selected the highest percentage, which comes from Box A. * The ~ symbol means approximately.

What we have just done is demonstrate the *concept* of Bayes' Theorem and *solve a problem* all without using the formula. Now, before we solve this same problem with the formula, it might be helpful to define the formula and its components.

The Bayes' Theorem Formula: A Basic Overview

The formula for Bayes' Theorem is shown below. As you can see, there are three components to it. We find it helpful to call these components *ingredients* and think of the answer as all of the ingredients combined.

For every question you come across, you'll need to find each ingredient and plug it into the formula.

$$P(A|B) = \frac{\overbrace{P(B|A)}^{\text{INGREDIENT\#1}}\ \overbrace{P(A)}^{\text{INGREDIENT\#2}}}{\underbrace{P(B)}_{\text{INGREDIENT\#3}}}$$

Basic Definitions To Get You Started

For a more thorough explanation and additional terms, take a look at the back of this booklet.

- The vertical bar | stands for *given that*.
- P stands for *Probability*.
- A & B are *Events*.
- P(A) and P(B) are the probabilities of events A and B. Each event is separate from the other and *does not impact the other.*
- P(A|B) is the probability of event A being true *given that* event B is true.
- P(B|A) is the probability of event B being true *given that* event A is true.

Using the definitions from the previous page, the entire formula can be read as follows:

THE PROBABILITY OF "B" BEING TRUE GIVEN THAT "A" IS TRUE

THE PROBABILITY OF "A" BEING TRUE

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

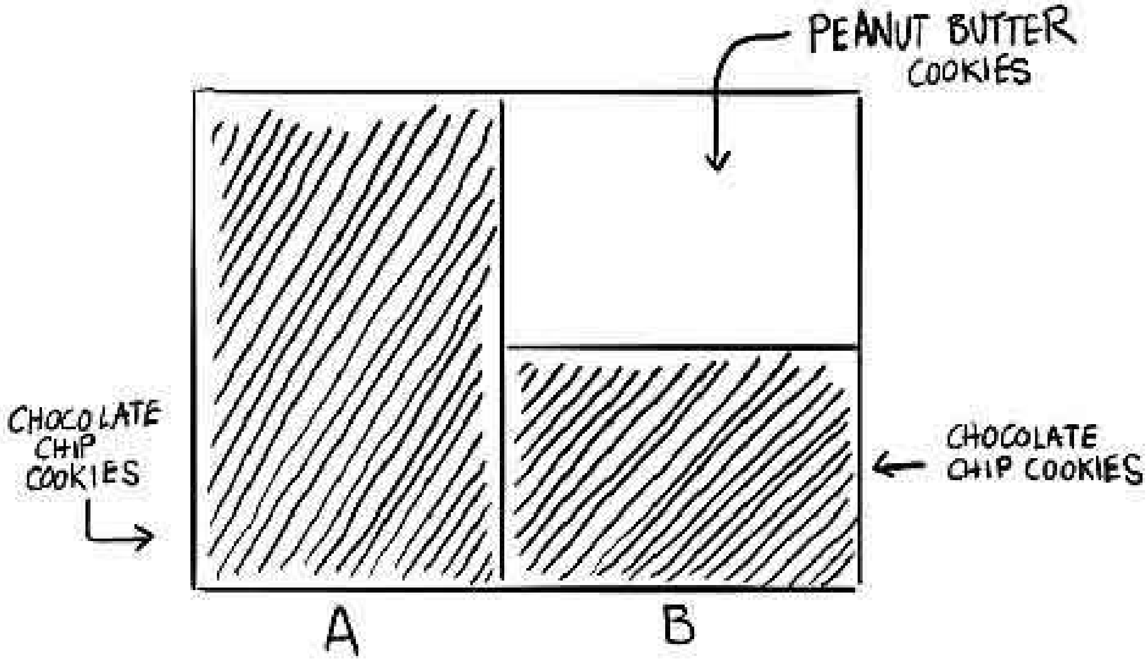
THE PROBABILITY OF "A" BEING TRUE GIVEN THAT "B" IS TRUE

THE PROBABILITY OF "B" BEING TRUE

The formula as a whole is built using basic algebra. It might look complicated but it is actually quite user-friendly. Every time you use the formula all you need to do is remember the three ingredients, find them, and then plug them into the formula to solve. You will then have an updated probability based on new information; you'll have P(A|B), which is technically called the *Posterior Probability* and is a **normalized weighted average.**

A Visual Intro: Part 2

In Part 1 of the Visual Introduction, Bayes' Theorem was demonstrated visually without using its formula. Now, in Part 2 we'll see how we can derive the same numbers using the formula. To refresh your memory, we had two boxes of cookies in front of us. One box was filled with 10 chocolate chip cookies. The other box had 5 chocolate chip and 5 peanut butter cookies. We then closed our eyes and picked a cookie out of a box, and when we opened them back up we had selected a chocolate chip cookie.



After doing this we discovered the following:

- There is a ~66% probability we chose the cookie from Box A.
- There is a ~33% probability we chose the cookie from Box B.

This time, let's follow 4 steps to find the answer using Bayes' formula. We will find the answer for Box A first, and then deduce from this to find the answer for Box B.

Step 1: To start, we always need to determine what we are wanting to find.

We want to know the probability of Box A *given that* we selected a chocolate chip cookie.

Step 2: Write what you want to find as a formula.

$$P(\text{BOX A} \mid \text{CC COOKIE}) = \frac{P(\text{CC COOKIE} \mid \text{BOX A})\, P(\text{BOX A})}{P(\text{CC COOKIE})}$$

Step 3: Find each ingredient. Then, plug it in.

- P(Box A) = .5 * To answer this we ask the following: What is the probability of drawing from Box A? Remember, this probability is independent of all other events. Since there are only two boxes and the probability of selecting from either is equal, the answer is .5
- P(CC Cookie) = .75 * To answer this we ask the following: What is the probability that we will select a chocolate chip cookie? Remember, this probability is independent of all other events. There are 20 cookies total in both boxes, and 15 of them are chocolate chip. So, 15/20 is .75
- P(CC Cookie | Box A) = 1* To answer this question, we ask the following: What is the probability of selecting a chocolate chip cookie given that we have selected from Box A? Since there are only chocolate chip cookies in Box A, the probability is 1. * A probability of 1 represents a 100% probability of something occurring.

Now, we can plug each ingredient into the formula:

$$P(A|B) = \frac{1 \times .5}{.75}$$

$$= \frac{.5}{.75}$$

$$= 0.6\overline{6}$$

$$= 66\%$$

Answer: We now know that there is a ~66% probability that we selected from Box A given that we have a chocolate chip cookie. To find the probability of selecting Box B, we can follow steps 1-3 *again* by replacing the term *Box A* with *Box B*. Or, we can simply deduce from our answer that if there is a ~66% probability Box A was selected, there must be a ~33% chance Box B was selected. Since all probability adds up to 1, we can discover this by doing the following: 1-.66 = .33, or ~33%.

Example Section 1: Solving For One Possible Outcome With All Data Provided

In this section, we use Venn diagrams to visualize our problems so that they are easier to understand and solve. If reading the term Venn diagram makes you shudder, Wikipedia provides a good overview to get you up to speed. This article at Stanford University is also helpful, but you really don't need to worry. We explain everything step-by-step.

Introduction

When searching for a probability we are sometimes given all of the components or ingredients, and simply need to A) Identify each one B) Label each one, and C) Plug each one into Bayes' formula. Hands down these are the easiest questions to apply Bayes' formula to. In this section we will be dealing with these types of probabilities using Venn diagrams. Venn diagrams work great for these by helping us visually understand the question.

Here are a few tips as you approach each question:

- Try not to get overwhelmed or lost in the question. Always begin by writing down what you want to discover.
- Try not to confuse P(A|B) with P(B|A). Always double check your numbers! This is a common error (technically called a Base Rate Fallacy).
- Remember to take your time. There is no need to rush.

Scenario 1: The Flu

Let's say that you are at work one day and have just finished lunch. You suddenly feel horrible and find yourself lying down. Wasn't your friend at work recently sick with the flu? What if you have it? Will you have to cancel your big trip next week?

You have a headache and sore throat, and you know that people with the flu have the same symptoms roughly 90% of the time. In other words, 90% of people with the flu have the same symptoms you currently have. Does this mean you have the flu?

Wanting to gain a little more information you roll over, grab your phone and search Google. You find a reputable article that says that only 5% of the population will get the flu in a given year. Ok. So, the probability of having the flu, in general, is only 5%.

You then spot one more statistic that says 20% of the population in a given year will have a headache and sore throat at any given time.

Do you have the flu? What should you do?

Pulling The Scenario Apart

First, let's remember what Bayes' Theorem does: it helps us update a hypothesis based on new evidence.

In this scenario, your hypothesis is that you have the flu and your headache and sore throat are your evidence. Now, after seeing that 90% of people with the flu have your symptoms, many of us would stop and conclude that we have the flu. We would look at the 90% statistic and sigh, resolved to the fact that we likely have the flu. This reaction is very common and called Base Rate Fallacy or Base Rate Neglect. The CIA has a nifty article on this, and it explains how people often gravitate towards the easiest information available when making decisions.

So, we are left wondering. Is our assumption based on the 90% statistic right?

This is where Bayes' Theorem comes in and helps us have a clearer picture. By using the theorem, we are forced to look at *all* data and update our hypothesis with new evidence. In the scenario we are given two *additional* pieces of information that can help us come to a more precise probability of having the flu *given* our symptoms.

Let's review all the information we do have before moving on.

1. We know people with the flu have a headache and sore throat roughly 90% of the time.
2. We know the probability of having the flu, in general, is only 5%.
3. We know that 20% of the population in a given year will have a headache and sore throat at any given time.

To begin solving this problem, we always need to determine what we are wanting to find.

We want to know what the probability is of having the flu *given* our current symptoms.

Now that we know what we are solving for, we are going to tackle this problem two ways. Depending on how you learn you may prefer one over the other, and that is ok. People learn differently, and that is why we included both options.

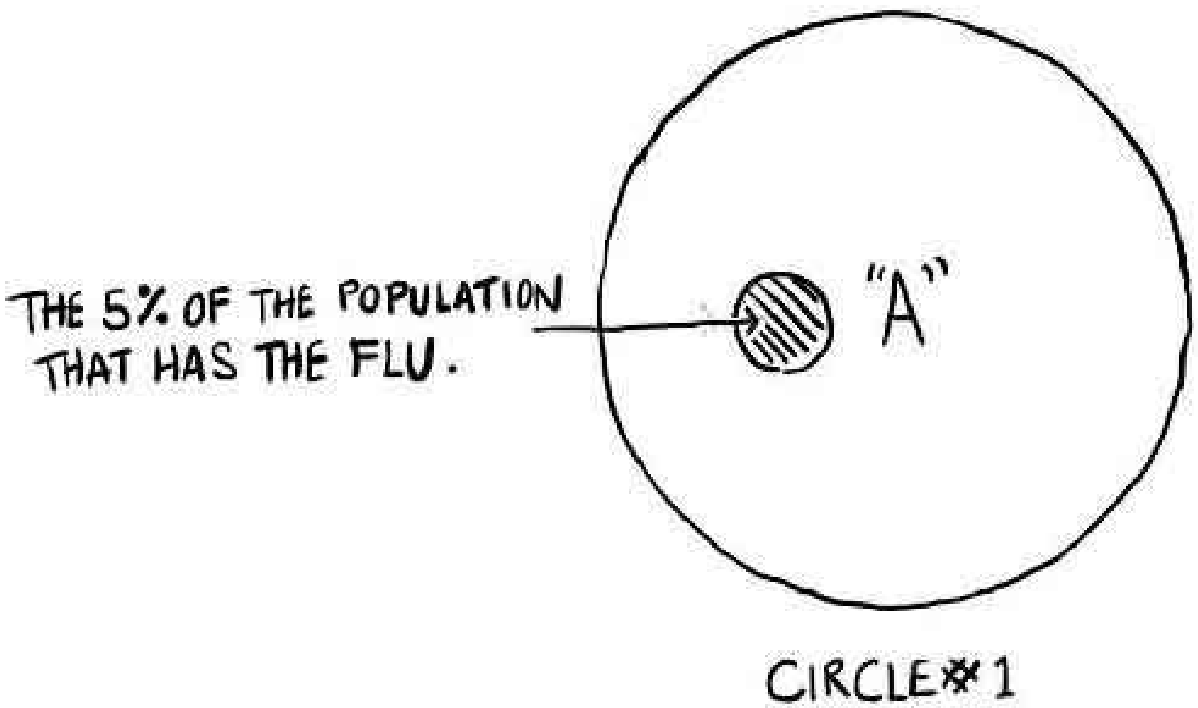
Feel free to read through each or skip back and forth.

1. Visualize the Problem: we will visualize the above flu problem by using a Venn diagram.
2. Plugging Into Bayes' Formula and Solving: we will solve the above flu problem by plugging our numbers into Bayes' formula.

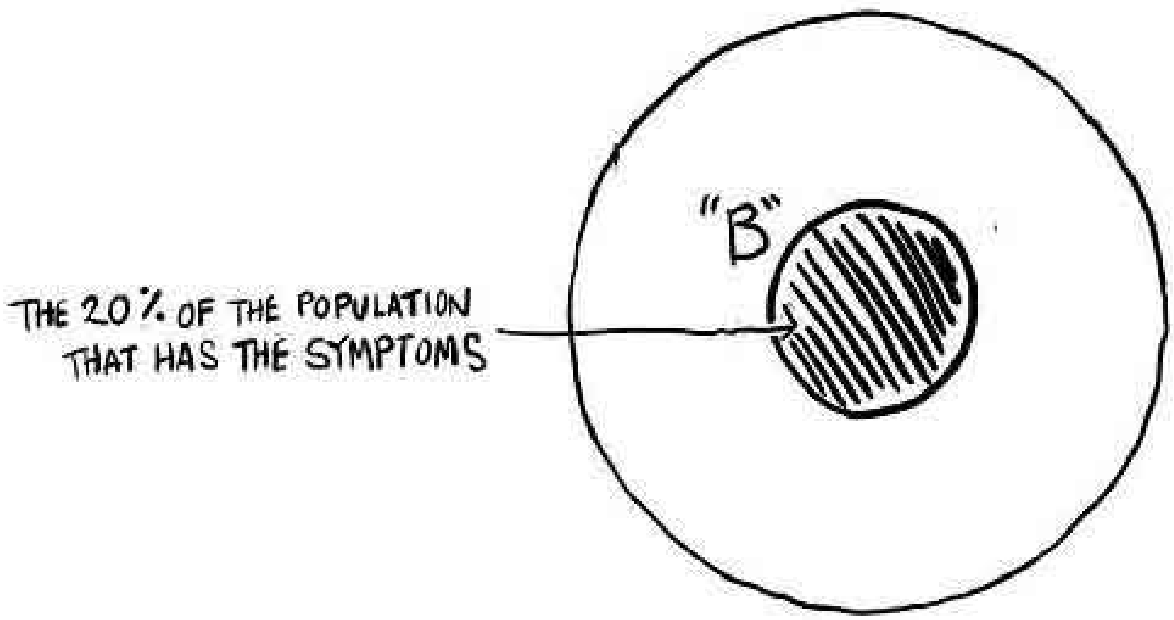
Scenario 1.1 Visualize The Problem

To visualize the problem, we'll draw two circles and merge them into a Venn diagram.

Circle #1: The area inside this circle represents all possible outcomes. In this example, the area represents all people who could get sick with the flu. The shaded circle labeled "A" represents the 5% of the population who have the flu. Within the circle is the entire population, and there are two possible outcomes for the population: people can have the flu, or not have the flu."A" is an event, and its probability is 5%. This probability is represented in our formula as P(A).



Circle #2: The area inside this circle also represents all possible outcomes. In this instance it represents all people who could have your symptoms (headache and sore throat). The shaded circle labeled "B" represents the 20% of the population that does have the symptoms. What this means is that within the entire circle there are two possible outcomes: people have the symptoms or do not have the symptoms. "B" is an event, and its probability is 20%. This probability is represented in our formula as P(B).

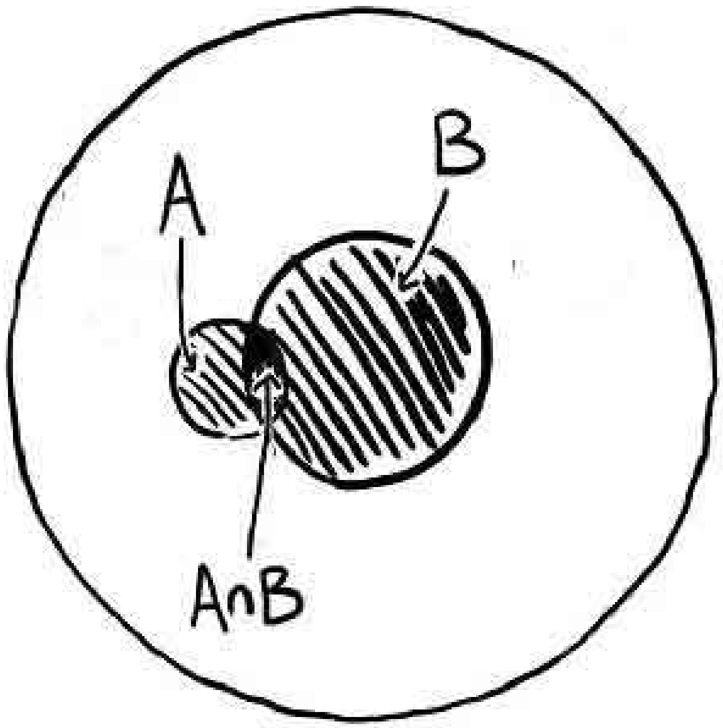


CIRCLE #2

Circle #3: In this circle, we have combined both events “A” and “B” - and this is where the magic happens!

Here is a quick breakdown of how you can read this:

- The white area inside this circle represents people who do not have either the flu or the symptoms.
- The area where only Circle A covers shows us people who only have the flu.
- The area where only Circle B covers shows us people who only have the symptoms.
- Now, take a look at Circle B and see where it overlaps with Circle A. This is what we are really interested in! This is our question from Step 1 in visual form. We want to know the probability $P(A|B)$ of having the flu given our symptoms. This probability is found where both events occur together and is called an intersection. Another way to look at it is like this: if we are in area B, what is the probability we are also in area AB (where A and B overlap)?



With both circles now merged, we can visually see our question and what we are trying to solve for. Although we won't be solving the question with a Venn diagram, the diagram does help us visualize what we are trying to understand.

If P(A) is the probability of you having the flu, and P(B) is the probability of you having your symptoms, what is the probability of you having both? While we don't yet know the actual answer, we can clearly see what we are trying to solve for.

Scenario 1.2 Plugging Into Bayes' Formula And Solving

Now let's solve the problem by using Bayes' formula. For the sake of ease, we'll begin by re-stating what we want to find.

Step 1: Determine what you want to find.

Again, we are solving for the same thing we did above with the Venn diagram but are restating this for clarity. We want to know what the probability is of having the flu *given* our current symptoms.

Step 2: Write the above as a formula.

Let's translate what we are solving for into the formula. In other words, we'll bring the language of Step #1 above into the formula.

Here is Bayes' formula:

$$P(A|B) = \frac{\overbrace{P(B|A)}^{\text{INGREDIENT\#1}}\ \overbrace{P(A)}^{\text{INGREDIENT\#2}}}{\underbrace{P(B)}_{\text{INGREDIENT\#3}}}$$

Now, let’s translate with what we are solving for.

$$P(\text{FLU} \mid \text{SYMPTOMS}) = \frac{P(\text{SYMPTOMS} \mid \text{FLU})\, P(\text{FLU})}{P(\text{SYMPTOMS})}$$

Step 3: Find each ingredient and label it.

From the scenario we know the following:

*We have changed the ingredients provided in the scenario from percents into decimals. We will do this every time before we begin to plug the ingredients into the formula.

- P(A) - In our formula, this ingredient is represented as P(Flu) and answers the question: What is the probability of you having the flu? This number is .05
- P(B|A) - In our formula, this ingredient is represented as P(Symptoms | Flu). This number is .9
- P(B) - In our formula, this ingredient is represented as P(Symptoms) and answers the question: What is the probability of you having the symptoms? This number is .2.

Step 4: Plug each ingredient into the formula and solve.

$$P(A|B) = \frac{.9 \times .05}{.2}$$

$$= .225$$

$$= 22.5\%$$

Conclusion:

So, after plugging each ingredient into the formula our answer is 22.5%. We can conclude from this that if you have a sore throat and headache you only have a 22.5% probability of having the flu. Wow! Now, remember what Bayes' Theorem does: it helps us update a hypothesis based on new evidence.

Originally, we thought that the probability of having the flu was as high as 90%! This belief was based on our latching on to P(B|A). However, our answer P(A|B) is very different! The 22.5% we ended with is more accurate than the 90% probability we started with.

This problem is a fantastic illustration of the *power* that Bayes' Theorem can give us when facing tough uncertainties. It is also a tweaked example of a questionnaire given to 1000 gynecologists. In the study, only 21% of gynecologists chose the correct answer while almost 50% chose the equivalent of our 90%! If you'd like to read more on this, Cornell University has a fantastic article.

Next time you are sick, remember Bayes!

Scenario 2: Breathalyzer

You are a police officer in Baltimore and it's New Year's eve. As usual, roadblocks are set-up at various points throughout the city to combat drunk driving.

Throughout the evening you and your fellow officer are giving breathalyzer tests to random drivers. Around 2 am you randomly pull over a vehicle and have the driver take a breathalyzer test, and the result is positive. You assume the test is accurate and think nothing of it as you process the driver.

After your shift ends early that morning you are talking with your partner. She doesn't believe that the breathalyzer tests are anywhere near accurate. You ask her why and she tells you some stats: it is true that the breathalyzer *always* detects a truly drunk person, but only 1 in 1000 drivers are typically drunk. What's more, the probability of the test being positive is only 8%.

You shake your head while you try to put together what she has said. You've never questioned the test before but now you are. What should you believe?

Pulling The Scenario Apart

In this breathalyzer scenario, your hypothesis is that the person is drunk and your evidence is a positive breathalyzer test. Originally, you never questioned the accuracy of the test. You thought it was accurate since it has a 100% success rate of *always* detecting a truly drunk person. But after listening to your partner you are now questioning this belief.

As in our flu example above, this is where Bayes' Theorem comes in and helps us have a better understanding of probability. In this scenario we are given two *additional* pieces of information that can help us come to a more precise probability. Let's review those quickly before we move on.

1. We know that <u>100% of the time the breathalyzer test will give a positive result for a truly drunk driver</u>. By truly drunk we mean that a blood test would confirm that the person is over the blood alcohol limit.
2. We know that 1 in 1000 drivers drives drunk<u>, so the probability of any driver being drunk is 0.1%* </u>*(we calculated this by dividing 1/1000 and then multiplying by 100)*.
3. We also know <u>that the breathalyzer test will give a positive result 8% of the time</u> regardless if it is accurate or not.

To begin solving the question, we always need to determine what we are wanting to find.

We want to know the probability of someone actually being drunk *given that* the breathalyzer test is positive.

Perfect. We now know what we are solving for, so let's move on and tackle it in a few ways.

We are going to look at this problem through two different lenses.

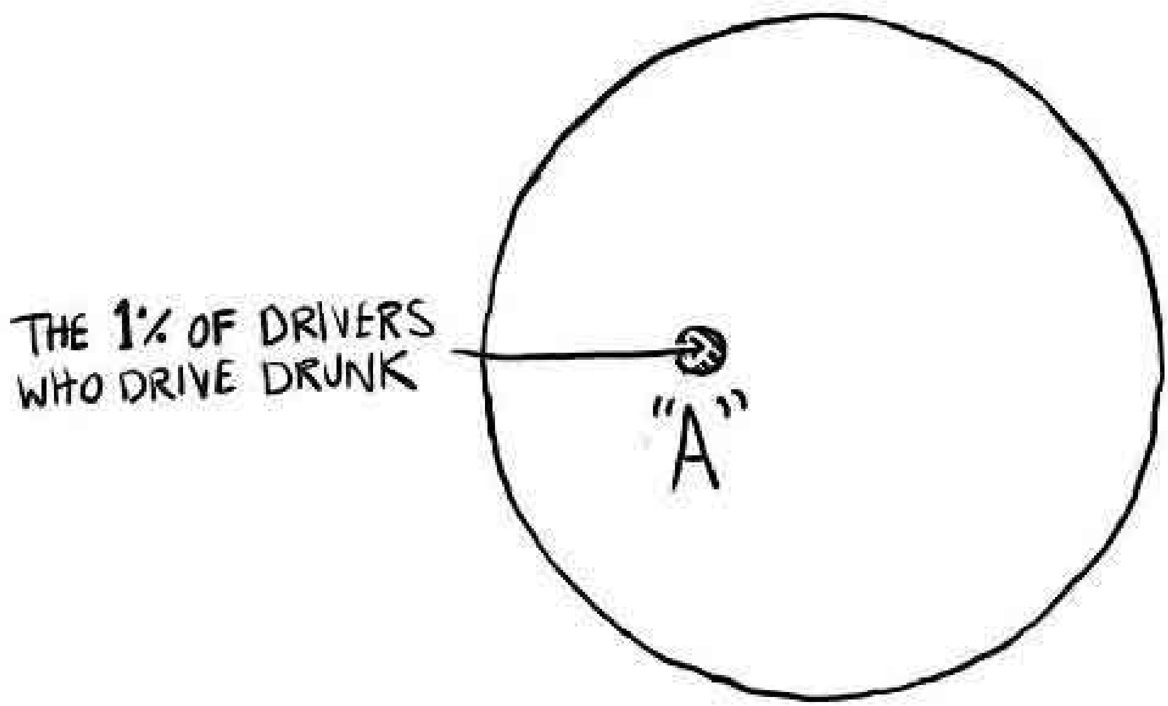
1. Visualize the Problem: we will visualize the problem by using a Venn diagram.

1. Plugging Into Bayes' formula and Solving: we will solve the problem by plugging our numbers into Bayes' formula.

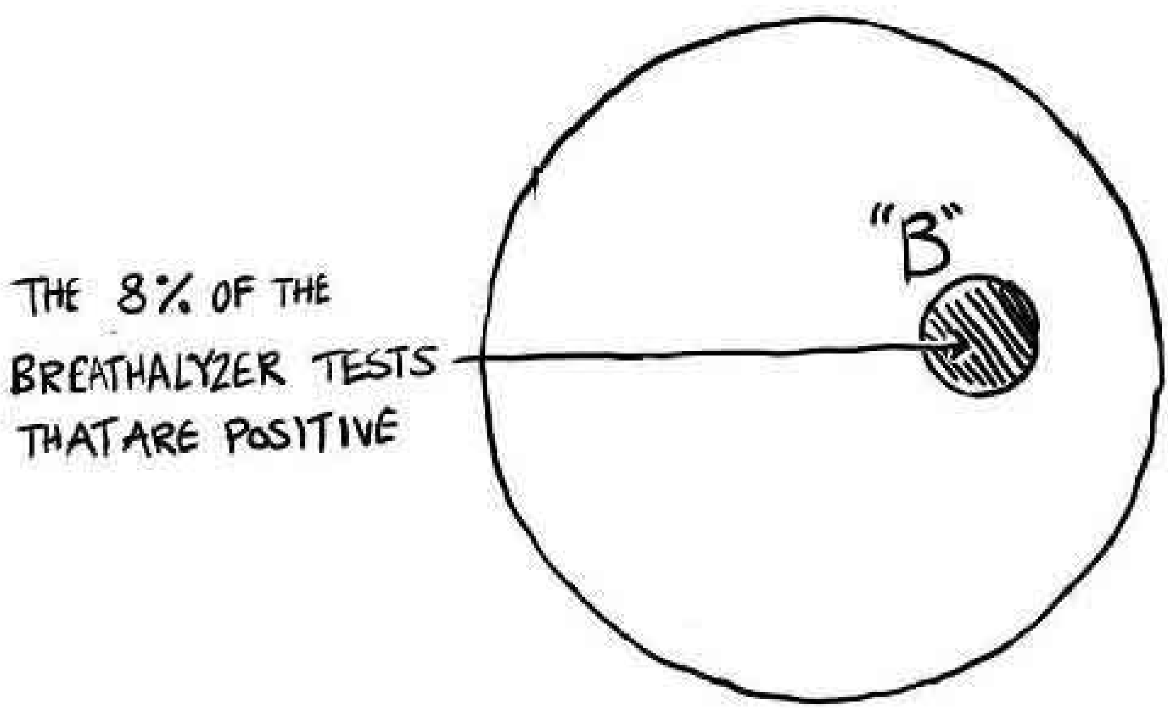
Scenario 2.1 Visualize the Problem

Let's visualize with a Venn diagram.

Circle #1: The area inside this circle represents all possible outcomes. In this scenario, it represents all people who *could be drunk while driving*. The small circle labeled "A" represents the .1% of drivers who actually are drunk. "A" is an event, and its probability is .1%. This probability is represented in our formula as P(A).



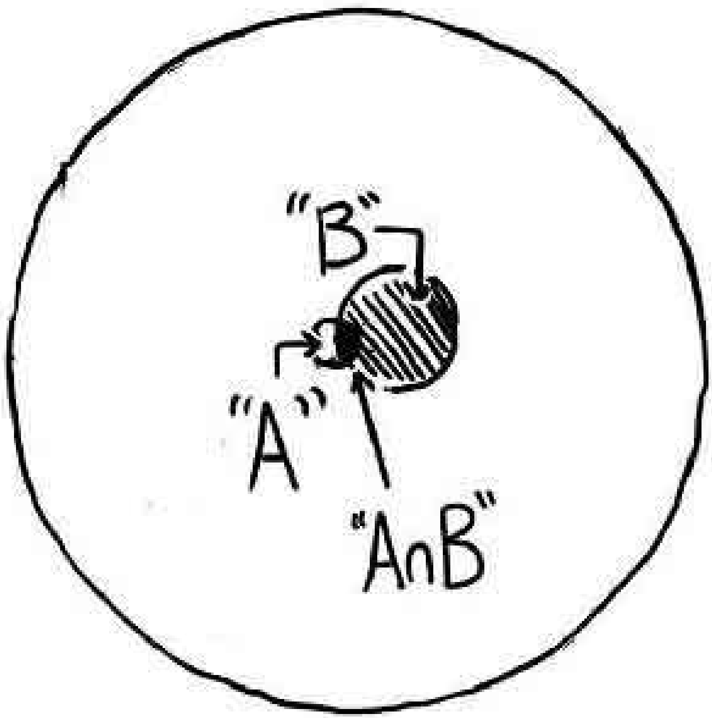
Circle #2: The area inside this circle represents all possible outcomes. In this scenario, it represents *all possibilities for the breathalyzer test*. The small circle labeled "B" represents the 8% of the tests that are positive."B" is an event, and its probability is 8%. This probability is represented in our formula as P(B).



Circle #3: This is where all the pieces come together. In this circle we have combined both events “A” and “B”.

Here is how the entire visual can be understood:

- The white area inside this circle represents people who are not drunk drivers and breathalyzer tests that are negative.
- The area where only Circle A covers shows us people who *are drunk while driving*.
- The area where only Circle B covers shows us the *total amount of breathalyzer tests that are positive*.
- Boom! Take a look at the dark area where the two circles overlap. This is what we are *really* interested in! *This is our question that we want to be answered, but in visual form*. We want to know the probability $P(A|B)$ of a driver being drunk given that the breathalyzer test is positive. This probability is found where both events occur together and is called an intersection.



With both circles now merged, we can visually see our question and what we are trying to solve for. Although we won't be solving the question with a Venn diagram, the diagram does help us visualize what we are trying to understand.

If P(A) is the probability a driver driving drunk, and P(B) is the probability a breathalyzer test being positive, what is the probability of both?

Scenario 2.2 Plugging Into Bayes' Formula And Solving

Let's follow our four steps again. To make things clear, we'll clarify what we want to find.

Step 1: Determine what you want to find.

We want to know the probability of someone being truly drunk *given that* the breathalyzer test is positive.

Step 2: Write the above as a formula.

Let's translate what we are solving for into the formula. In other words, we'll bring the language of Step #1 above into the formula.

Here is Bayes' formula:

$$P(A|B) = \frac{\overbrace{P(B|A)}^{\text{INGREDIENT\#1}}\ \overbrace{P(A)}^{\text{INGREDIENT\#2}}}{\underbrace{P(B)}_{\text{INGREDIENT\#3}}}$$

Now, let’s translate with what we are solving for.

$$P(\text{DRUNK}|\text{POSITIVE}) = \frac{P(\text{POSITIVE}|\text{DRUNK})\ P(\text{DRUNK})}{P(\text{POSITIVE})}$$

Step 3: Find each ingredient and label it.

From the scenario we know the following:

- $P(A)$ - In our formula, this ingredient is represented as $P(Drunk)$ and answers the question: What is the probability of a driver being drunk? This number is .001
- $P(B|A)$ - In our formula, this ingredient is represented as $P(Positive|Drunk)$. This number is 1, not .1, but 1 as in 100%.
- $P(B)$ - In our formula, this ingredient is represented as $P(Positive)$ and answers the question: What is the probability of a breathalyzer test being positive? This number is .08

Step 4: Plug each ingredient into the formula and solve.

$$P(A|B) = \frac{1 \times .001}{.08}$$

$$= \frac{.001}{.08}$$

$$= 0.0125$$

$$= 1.25\%$$

Conclusion:

So, after all the work and plugging each ingredient into the formula our answer is 1.25%. We can conclude from this that the probability of a driver having a positive test *and actually being drunk* is 1.25%. In other words, for every 1000 drivers being tested we have the following:

- We have 1 truly drunk driver who *tested positive* with the breathalyzer.
- We have roughly 79 other drivers who also tested positive with the breathalyzer but *are not drunk*.
- In total we have 80 drivers testing positive, and the probability of one of them testing positive *and truly being drunk* is around 1.25%.

That's an eye opener! Remember what Bayes' Theorem does: it can help us quantify skepticism and enable us to have a clearer understanding. Originally, we thought the probability of the driver being drunk was quite high, but now we see it is only around 1.25%.

In this scenario, the police officer would now take a very different view of breathalyzer tests and would likely be much more skeptical of their accuracy.

*As a side note, there are other complexities to situations like this, such as the test *actually* being random, etc. Did the police officer pull the vehicle over because of how the driver was driving? Did the officer provide a breathalyzer test because of how they were acting? Responding? New information such as this would impact the entire equation. For sake of ease and teaching we have kept the problem very basic.

Scenario 3: Peacekeeping - A Surprise Attack

You are a soldier and have recently shipped out across the Atlantic on your fourth peacekeeping tour. A few weeks into your mission you are on patrol and see an injured family across the road from you.

You are about to go to them when suddenly there is a surprise attack and you find yourself pigeonholed against a burnt out vehicle. You stop to listen and are suddenly filled with horror as you see a truck turning the corner. There is no doubt it is an enemy vehicle, but you didn’t have time to see if the truck was rigged with a gunner on the back - and if there is you don’t want to be caught in the open.

You quickly do some mental calculations and recall what you learned in your debrief. The rebels have roughly 54 dilapidated trucks and 22 of them are rigged with guns in the back. Rebels in a truck are one thing, but rebels in a truck rigged with a gun on the back? You don’t want to be caught in the open with that.

You pop your head out to get a better look and a wave of bullets hits the vehicle in front of you. The rebel truck is now about 150 yards away, but you are still uncertain if the shots came from the truck or somewhere else. If the truck is rigged with a gun, the chance of it having fired at you is pretty high, maybe at 80%.

You continue to think. Considering how heavy the firepower was and the environment you are in, you peg the possibility of being shot at 50%. What should you do? Should you risk crossing the street to help the family?

Pulling The Scenario Apart

As in our other scenarios, let's break this scenario apart to see exactly what we are dealing with.

In this scenario, your hypothesis is that the truck is rigged with a gun and your evidence is the wave of bullets. Your initial assumption is that the truck is likely rigged with a gun since there is an 80% chance it would have fired at you *if it was rigged*. But is that right? Are you making an accurate conclusion?

Again, this is where Bayes' Theorem can help us *better understand* the situation and make a *more informed* decision. In the scenario we are given two *additional* pieces of information that can help us come to a more precise probability of the truck being rigged with a gun *given* the intense wave of bullets. Let's take a moment to review:

1. We know that the probability of the truck firing at you if it is rigged with a gun is 80%.
2. We know the probability of the truck being rigged with a gun is 40%* *(we calculated this by dividing 22 rigged trucks by 54 total trucks. 22/54 = ~40%.)*
3. We know that the probability of a rebel having heavy firepower is 50%.

To start, we always need to determine what we are wanting to find.

We want to know the probability of the truck being rigged with a gun *given that* we were just fired at with heavy firepower.

Let’s move on and solve this problem.

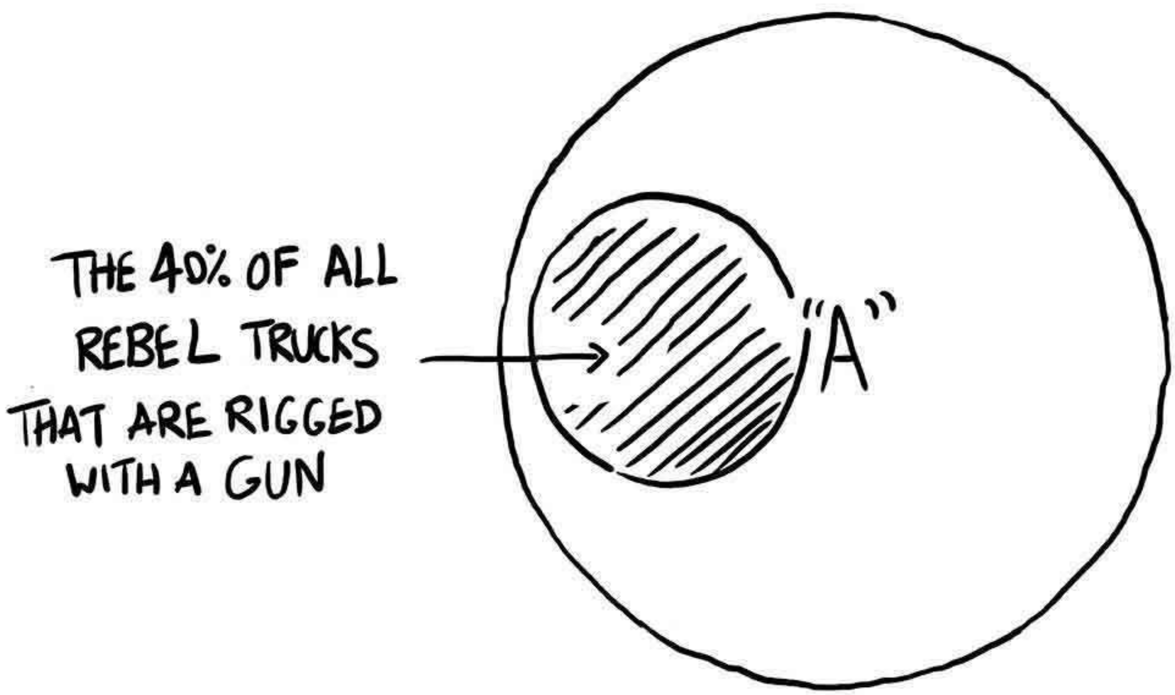
1. Visualize the Problem: we will visualize the problem by using a Venn diagram.

2. Plugging Into Bayes' formula and Solving: we will solve the problem by plugging our numbers into Bayes' formula.

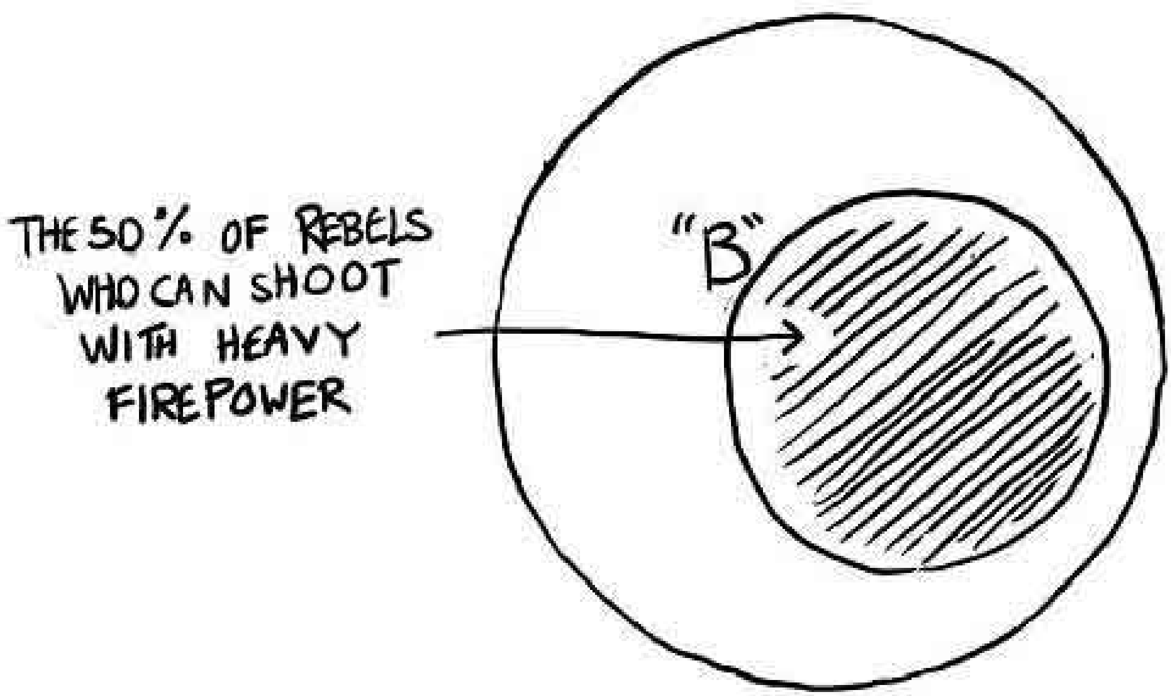
Scenario 3.1 Visualize The Problem

Let’s visualize with a Venn diagram.

Circle #1: The area inside this circle represents all possible outcomes. In this example, the area represents all rebel trucks. The shaded circle labeled “A” represents the 40% of all rebel trucks that are rigged with a gun. * To get this number we simply divide the number of rigged trucks (22) by the number of total trucks (54). The answer is ~40%. “A” is an event, and its probability is 40%. This probability is represented in our formula as P(A).



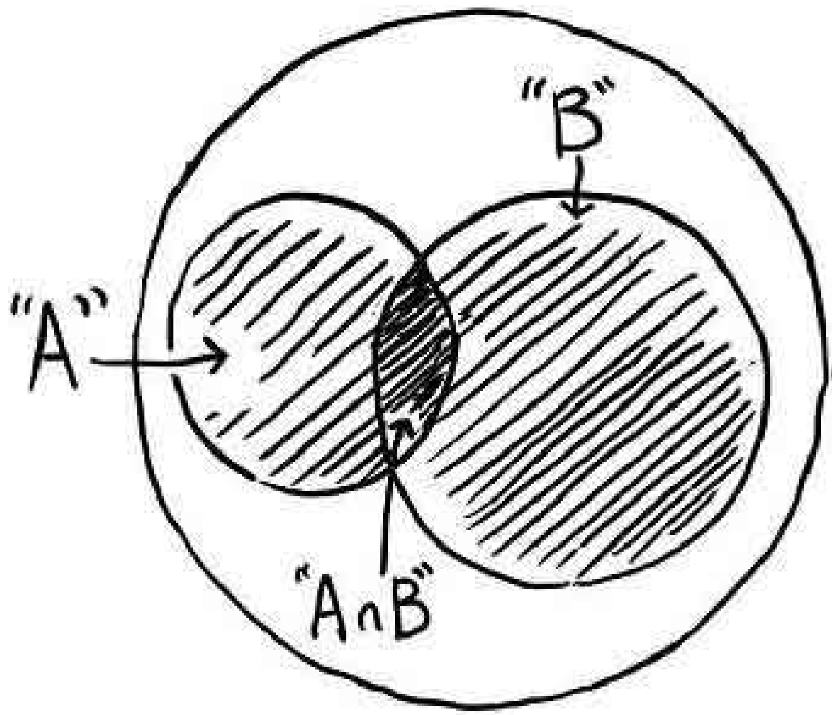
Circle #2: The area inside this circle also represents all possible outcomes. In this instance it represents firepower. The shaded circle labeled “B” represents the 50% of rebels who could shoot you with *very heavy* firepower. What this means is that within the entire circle there are two possible outcomes: rebels either have heavy firepower or do not have heavy firepower. “B” is an event, and its probability is 50%. This probability is represented in our formula as P(B).



Circle #3: Alright. Let's take a look at how these two *events* combine. This is where the magic happens!

Here is a quick breakdown of how you can understand the entire diagram:

- The white area inside this circle represents rebel trucks that are not rigged with a gun and rebels who do not have heavy firepower.
- The area covered by Circle A shows us the total amount of rebel trucks that are rigged with a gun.
- The area covered by Circle B shows us rebels who have heavy firepower capability.
- Now, move your eyes to the dark area where the two circles overlap. This is what we are really interested in! This is our original question in visual form. We want to know the probability $P(A|B)$ of a rebel truck being rigged with a gun given heavy firepower. In other words, if we are in Circle B, what is the probability of being in A as well? This probability is found where both events occur together and is called an intersection.



With both circles now merged, we can visually see our question and what we are trying to solve for. If P(A) is the probability of a rebel truck being rigged with a gun, and P(B) is the probability of heavy firepower, what is the probability of you being where they overlap - and both events occurring at the same time?

Scenario 3.2 Plugging Into Bayes' Formula And Solving

To solve by using Bayes' Theorem we'll follow four steps. For sake of ease, we'll begin by re-stating what we are wanting to find.

Step 1: Determine what you want to find.

We want to know the probability of the truck being rigged with a gun *given that* we were just fired at with heavy firepower.

Step 2: Write the above as a formula.

Let's translate what we are solving for into the formula. In other words, we'll bring the language of Step #1 above into the formula.

Here is Bayes' formula:

$$P(A|B) = \frac{\overbrace{P(B|A)}^{\text{INGREDIENT\#1}}\ \overbrace{P(A)}^{\text{INGREDIENT\#2}}}{\underbrace{P(B)}_{\text{INGREDIENT\#3}}}$$

Now, let’s translate with what we are solving for.

$$P(\text{RIGGED}|\text{FIREPOWER}) = \frac{P(\text{FIREPOWER}|\text{RIGGED})\ P(\text{RIGGED})}{P(\text{FIREPOWER})}$$

Step 3: Find each ingredient and label it.

From the scenario we know the following:

- P(A) - In our formula, this ingredient is represented as P(Rigged) and answers the question: What is the probability of a rebel truck being rigged with a gun? This number is .4
- P(B|A) - In our formula, this ingredient is represented as P(Firepower|Rigged). This number is .8
- P(B) - In our formula, this ingredient is represented as P(Firepower) and answers the question: What is the probability of a rebel having heavy firepower? This number is .5

Step 4: Plug each ingredient into the formula and solve.

$$P(A|B) = \frac{.8 \times .4}{.5}$$

$$= \frac{.32}{.5}$$

$$= 0.64$$

$$= 64\%$$

Conclusion:

Once we've plugged all of our ingredients into the formula we arrive at 64%. Based on this, we can conclude that the probability of the truck being rigged with a gun *given that* heavy firepower came our way is 64%. We originally thought the probability was higher around the 80% mark, but now we can see it hovers about 15 % lower.

If your original instinct was to stay behind the burnt out vehicle, knowing this probably won't change that. A difference of 15% is not that great. But, if the result (posterior probability) had been lower around 20% or 30% instead of 64%, that might have changed your actions.

We chose this last scenario to contrast the first two and demonstrate that using Bayes Theorem does not always provide a clear answer. Sometimes a probability only slightly changes.

How To Understand Scenarios

Before we jump into Example Section 2 and are forced to discover P(B), it might be helpful to know how to read and understand scenarios. If you feel confident with this, please skip this section and move on to Example Section 2. However, basic skills in this area are often overlooked and not taught properly. So, if you are keen to brush up on how to understand scenarios, keep reading!

Discovering What You Are Looking For: P(A|B)

This is the first and most important step you'll take when solving a question. Before you can solve a question, you need to know *what you are solving for.* Here are three steps that can help you do this:

Step 1: Ignore all the extra filler provided in the question and always begin by writing down each statistic given. Doing this will help you focus on what matters most and *disregard* the rest. Be sure to give each statistic a label. For example:

- 100% - positive breathalyzer tests for truly drunk drivers.
- .1% - number of drivers who drive drunk.
- 8% - positive results given by the breathalyzer.

Step 2: Look at what you've collected in the list above. For the majority of scenarios, you will have two or three numbers that are provided. Use these to help you brainstorm questions. Some great questions to ask are:

- What probability am I wanting to know?
- What belief is being questioned?

Again, ignore all the extra filler in the question (by filler we mean the backstory and other details that don't have anything to do with the numbers above).

Step 3: Working with Step 2, begin to write down what you think you are trying to solve for. Remember, you can write your question out rough and then refine it (as in our example below). And always remember to include the terms *probability* and *given that*. Doing this will help you define each ingredient with confidence.

- Version 1: If a breathalyzer is positive, what are the chances someone is drunk?
- Version 2: What is the probability of someone being drunk if the breathalyzer is positive?
- Final Version: What is the probability of a driver being drunk *given that* their breathalyzer test is positive?

Defining P(B), P(A) P(B|A)

Once we know what we are looking for, defining the various ingredients is straightforward. It is also much easier if you have used the terms *probability* and *given that* in Step 3 above. Here's how to do it:

Step 1: Write out what you are searching for. For example:

- What is the probability of a driver being drunk *given that* their breathalyzer test is positive?

Step 2: Label each ingredient:

- Event "A" is always found between *probability* and *given*. In this case, it is *Driver Being Drunk*. The probability of this is P(A).
- Event "B" is always found after *given that*. In this case, it is *Breathalyzer Test is Positive*. The probability of this is P(B).
- P(B|A) is a combination of the above. In this case, P(B|A) is: The probability of a breathalyzer test being positive, *given that* a driver is drunk.
- P(A|B) is also a combination of the above. In this case, P(A|B) is: The probability of a driver being drunk *given that* their breathalyzer test is positive.

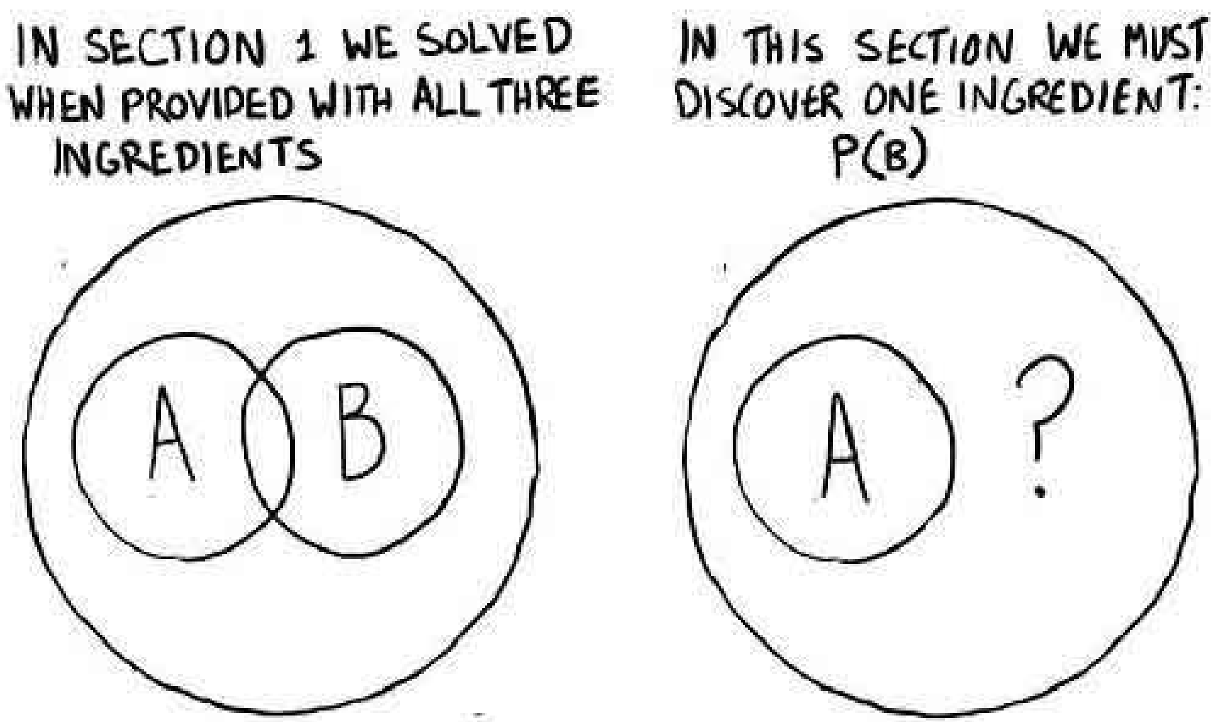
Once you understand what you are looking for and have defined your ingredients, you are on your way to solving the question. That's all there is too it!

Example Section 2: Solving for One Possible Outcome With No P(B) Provided

In this section we use decision trees to visualize and solve our problems. Decision trees are quite straightforward, but if you need a refresher [Wikipedia] provides a great overview. [This short lesson from Penn State] is also helpful. However, just like Venn diagrams we approach decision trees with small steps. Plus, we are not going to be using any complex terminology. You don't need to worry!

Introduction

In Example Section 1 we used Venn diagrams to visually understand our questions and were provided with all the ingredients to solve. In this section, we are only provided with two of the three ingredients and must discover the third. Here is what this looks like using Venn diagrams:



In real life this is typically the case and often requires us to discover one or more of the ingredients. P(B) is often the culprit, and so in this section we will focus on discovering this ingredient. We will be using the same scenarios as in Section 1, but we have tweaked them because we are now solving for P(B).

This time around we will be using two techniques to solve each scenario:

- Decision trees
- A classic approach using Bayes' formula

Decision trees are a fantastic and powerful tool that can be used to quickly find P(B) and give you a clear understanding of *how it is discovered*. They are a great visual aid for helping you grapple with and comprehend probability questions.

For each scenario in this section we will also solve the question by using a traditional Bayes' formula approach. Let’s dig in!

Scenario 1: The Flu

The Scenario: (Expanded from Example Section 1)
Let's say that you are at work one day and have just finished lunch. You suddenly feel horrible and find yourself lying down. The first thing that passes through your mind is food poisoning, but you don't think that is the case. You then remember that your co-worker was recently off for a few days with the flu. Could you have the flu? Will you have to cancel your big trip next week?

You grab your phone and search for some answers. Google tells you that 5% of the population will get the flu each year. A few minutes pass by and you remember that you just downloaded a new app that predicts illness. Why not see what it says?

You open it and input your symptoms, and within a few seconds the app predicts that you have the flu. It also displays the following:

- It correctly predicts people having the flu 75% of the time.
- 20% of the time it predicts that people have the flu when *they do not* have the flu.

You throw the phone onto the seat beside you. What do you make of this?

To begin solving this question, we always need to determine what we are wanting to find.

We want to know the probability of having the flu *given that* the app predicted that you have the flu.

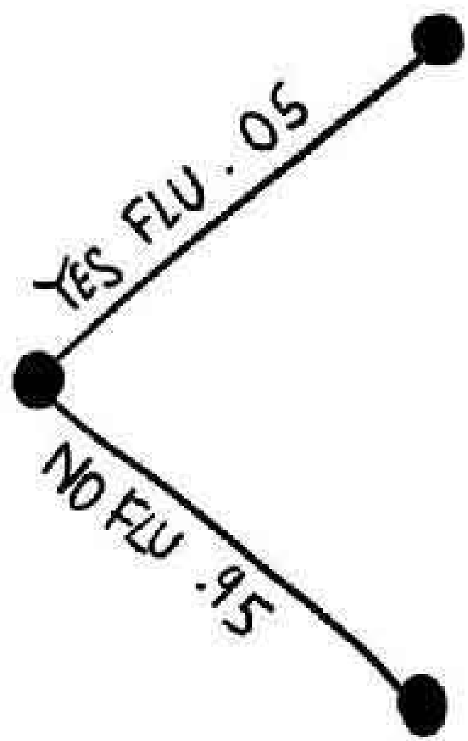
Great. Now that we know our goal, we can move forward and begin to solve by using a decision tree.

Scenario 1.1 Solve Using A Decision Tree

To start, we are going to solve this problem by using a decision tree. We will point out connections to Bayes' formula throughout this process.

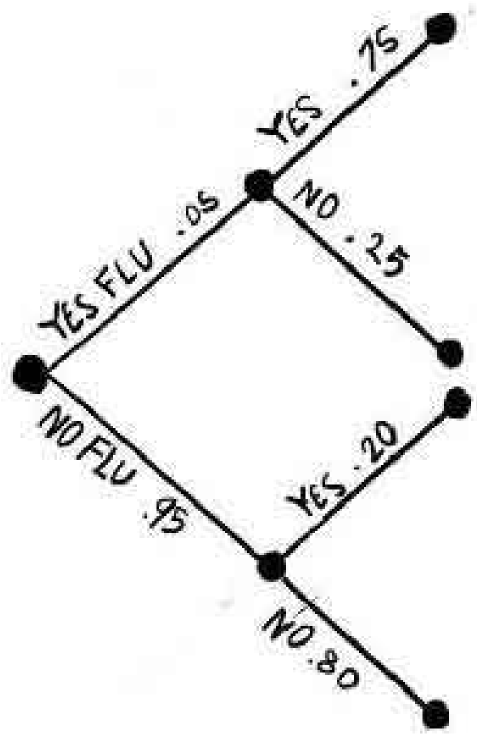
Step 1.

Begin by drawing two lines that connect at a single dot. These lines represent the possibility of you having the flu or *not* having the flu. We'll add in the number we know (.05), and then subtract from 1 to get our second number (.95). Notice that both branches add up to 1.



Step 2.

Next, we will add two more branches to each initial branch. Each pair of branches represents the possibility of the app predicting *yes* or *no*. In other words, if it will predict that you have the flu or do not have the flu. We'll add in our numbers from the scenario. Notice that each pair of branches adds up to 1.



Step 3.

Now we will label each pathway and compute its value. Each pathway begins with either having or not having the flu, and ends with either a yes or no prediction.

There are a total of 4 pathways:

Path 1: Yes Flu *to* Predict Yes

- Label YY
- Compute: .05 x .75 = .0375

Path 2: Yes Flu *to* Predict No

- Label YN
- Compute: .05 x .25 = .0125
- How did we find .25 ? Since the probability of both branches *together* is 100%, or 1, we subtracted .75 from 1. 1-.75 = .25.

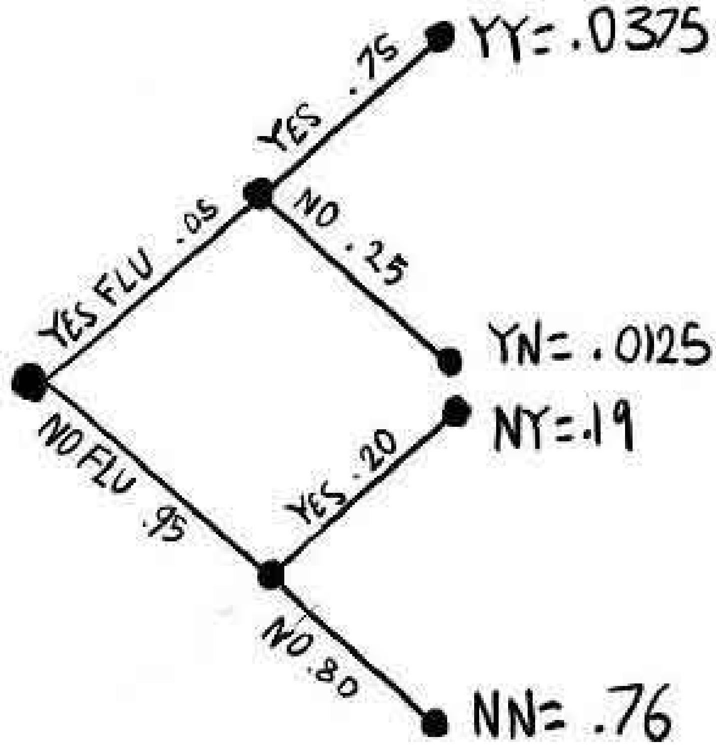
Path 3: No Flu *to* Predict Yes

- Label NY
- Compute: .95 x .20 = .19

Path 4: No Flu *to* Predict No

- Label NN
- Compute: .95 x .80 = .76

- How did we find .80 ? Since the probability of both branches *together* is 100%, or 1, we subtracted .80 from 1. 1-.20 = .80.



Step 4.

Once we have our numbers, we need to sum Path 1 and Path 3. Here is why we are doing this:

- The sum of Path 1 and Path 3 gives us P(B). In other words, it tells us the probability of the app predicting yes. This is our missing ingredient.
- When we sum Path 1 and Path 3 we are adding both ways the app can predict yes. Path 1 represents the app predicting yes *when you do have the flu*. Path 3 represents the app predicting yes *when you do not have the flu*. When we add these paths together, we find P(B).

$$\text{PATH } 1 = .0375$$

$$\text{PATH } 3 = .19$$

$$= .0375 + .19$$

$$= .2275$$

Step 5.

Now, we will divide Path 1 by the number we arrived at in Step 4 above. The end result will be our updated probability, which is P(A|B).

Path 1 represents the *probability of the app predicting yes when you do have the flu*. It is the equivalent of when we multiply P(B|A) with P(A).

THIS IS P(B|A) P(A)

$$= \frac{.0375}{.2275}$$

THIS IS P(B)

$$= 0.1648$$

$$= 16.5\%$$

Conclusion:

The probability that you have the flu *given that* the app predicted you do is only 16.5%. Now, let's move on and solve this same problem using Bayes' formula. We'll expand on our conclusion once we've done that.

Scenario 1.2 Solve Using Bayes' Formula

Let’s solve the same problem but now we’ll use Bayes' formula.

Step 1: Determine what you want to find.

We want to know the probability of having the flu *given that* the app predicted that you have the flu.

Step 2: Write the above as a formula.

$$P(\text{FLU} \mid \text{YES PREDICT}) = \frac{P(\text{YES PREDICT} \mid \text{FLU})\, P(\text{FLU})}{P(\text{YES PREDICT})}$$

Step 3: Find each ingredient and label it.

From the scenario we know the following:

- P(A) - In our formula, this ingredient is represented as P(Flu) and answers the question: What is the probability of you having the flu? This number is .05
- P(B|A) - In our formula, this ingredient is represented as P(Yes | Flu) This number is .75

The only ingredient we are missing is:

- P(B)

Now, what is P(B)? How do we define what event "B" is so we can try to find it? To find our answers, let's go back to what we are trying to figure out, which we defined in Step 1. Step 1 can be broken into two parts, and P(B) is tucked into the second part.

The two parts of Step 1 are:

1. Probability of having the flu. This is P(A).
2. Probability of the app predicting yes "B" *given that* you have the flu "A" is .75. This is P(B|A), which contains both events "A" and "B". This tells us that the definition of event "B" is *the app predicting yes*.

Excellent. Now we know the definition of event "B". But what about its probability P(B)? We were given a third number in the scenario (20%), but this is not P(B). It is a *part* of P(B). To figure out where it fits and how to solve for P(B) we need to do the following:

Let's think for a moment. How many ways can the app arrive at a yes (positive) prediction? There are only 2 ways:

1. It can predict a positive prediction that is correct.
2. It can predict positive prediction that is false.

All we need to do is multiply the numbers of each *yes* path and then add the answers together.

PATH FY $= .05 \times .75$

$= .0375$

PATH NY $= .95 \times .20$

$= .19$

NOW, ADD THE PATHS.

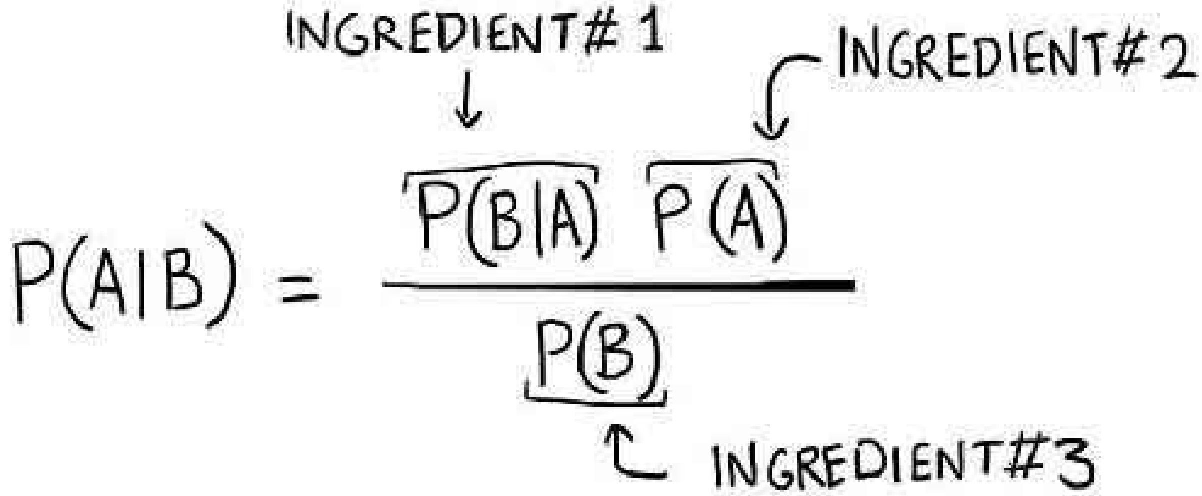
$= .0375 + .19$

$= .2275$

Now, the 20% from the scenario comes into play when the app predicts *yes* but the individual is not sick. If you look back at the paths we traced above (FY and NY), you can see the 20% in Path NY.

Step 4: Plug each ingredient into the formula and solve.

Now we have all three ingredients and can plug them into Bayes' formula!



- P(A) = P(Having Flu) = .05
- P(B|A) = P(Positive prediction *given that* you have the Flu) = .75
- P(B) = P(Positive Prediction) = .2275

Let's plug them in and solve.

$$P(A|B) = \frac{.75 \times .05}{.2275}$$

$$= \frac{.0375}{.2275}$$

$$= .1648$$

$$= 16.5\%$$

Conclusion:

We arrived at the same answer as before. The probability of having the flu *given that* the app predicted you do is 16.5%. Practically speaking what does this mean? How does it impact you?

Well, 16.5% is not very high, which means the probability of you having the flu is quite low. Since it is not anywhere near 50%, you decide to remain at work and not take a half sick day. At the end of the day you'll reassess your symptoms again.

Scenario 2: Breathalyzer

The Scenario: (Expanded from Example Section 1)

You are a police officer in Baltimore and it's the night after new year's eve. Even though new year's has passed, there are still a few drunk driving roadblocks across the city. You are on duty with the same partner as the night before, and you can't get your previous conversation with her out of your head. The fact is, you are now much more skeptical about the accuracy of breathalyzer tests. In fact, you almost outright do not trust them at all.

About half way through your shift your partner brings the subject of the tests up. She says that she did some more research last night and you might be interested in what she found. You nod your head as she continues. This is what she discovered:

- Approximately 3 out of every 1000 drivers will drive while drunk. This is .3%.
- The breathalyzer test *does not always* detect a drunk person. This is not 100% like you both previously thought, but 98%.
- 4% of the time breathalyzer tests *give a positive result for someone who is not drunk*. This is called a *false positive*.

You shake your head and take a deep breath. You are now really confused. What should you believe?

To start, we always need to determine what we are wanting to find.

We want to know the probability of someone actually being drunk *given that* the breathalyzer test is positive.

Perfect. Now let’s move on and tackle the question with a decision tree.

Scenario 2.1 Solve Using A Decision Tree.

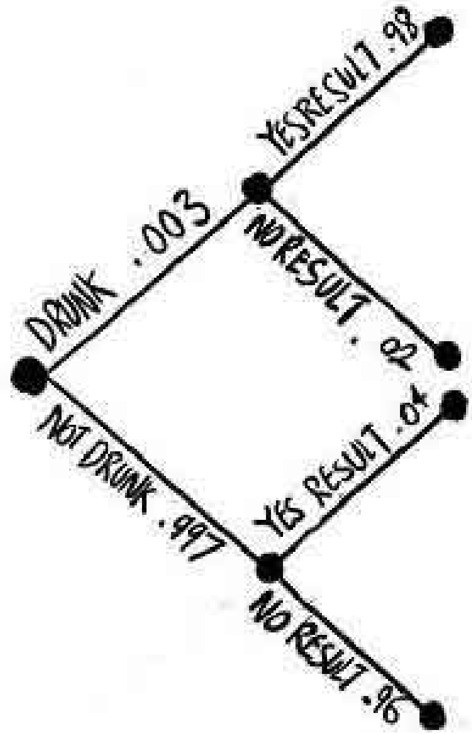
Step 1.

Begin by drawing two lines that connect at a single dot. These lines represent the possibility a driver being drunk and not being drunk. We'll add in the number we know (.003), and then subtract from 1 to get our second number (.997). Both branches add up to 1.

Step 2.

Next, we will add two more branches to each initial branch. Each pair of branches represents the possibility of the breathalyzer test indicating a positive or negative result. Notice each pair of branches adds up to 1.

How often does the test give a positive result that someone is drunk? Or a negative result stating that they are not drunk? We'll add in our numbers from the scenario.



Step 3.

Now we will label each pathway and compute its value. Each pathway begins with either being drunk or not being drunk, and ends at either a positive or negative (yes or no) breathalyzer result. There are a total of 4 pathways:

- Path 1: Drunk - Yes Result

- Label DY
- Compute: .003 x .98 = .00294

- Path 2: Drunk - No Result

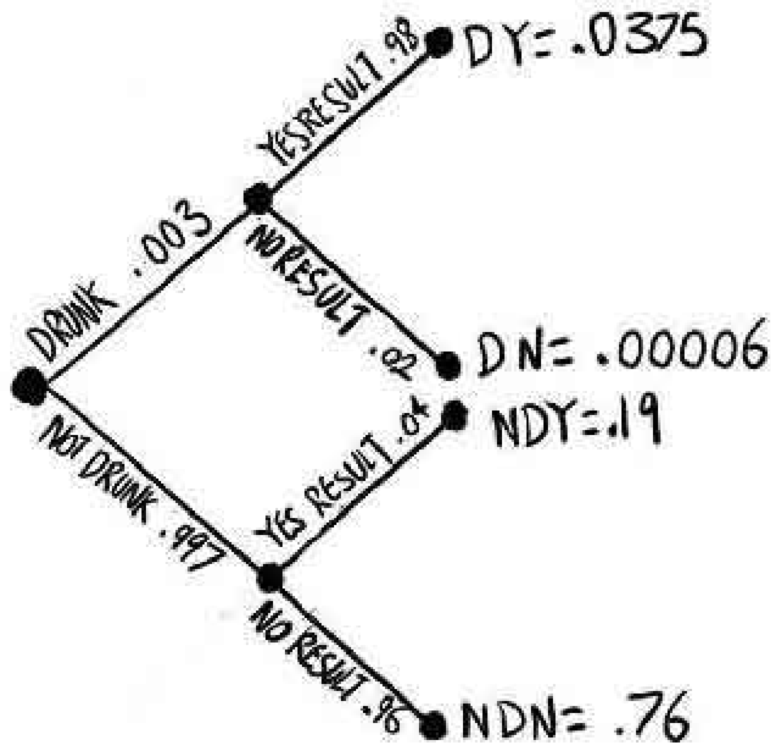
- Label DN
- Compute: .003 x .02 = .00006
- * How did we find .02 ? Since the probability of both branches together is 100%, or 1, we subtracted .98 from 1. 1-.98 = .02

- Path 3: Not Drunk - Yes Result

- Label NDY
- Compute: .997 x .04 = .03988

- Path 4: Not Drunk - No Result

- Label NDN
- Compute: .997 x .96 = .95712
- * How did we find .96 ? Since the probability of both branches together is 100%, or 1, we subtracted .04 from 1. 1-.04 = .96



Step 4.

Once we have our numbers, we need to sum Path 1 and Path 3. Here is why we are doing this:

- The sum of Path 1 and Path 3 tells us the *probability of the breathalyzer test having a positive (yes) result*. In other words it tells us P(B), which is our missing ingredient.
- When we sum Path 1 and Path 3 we are adding both ways the breathalyzer can come to a positive (yes) result. Path 1 represents the breathalyzer having a positive result *when the driver is truly drunk*. Path 3 represents the breathalyzer having a positive result *when the driver is not drunk*. When we add these paths together, we find P(B).

$$\text{PATH } 1 = .00294$$

$$\text{PATH } 3 = .03988$$

$$= .00294 + .03988$$

$$= .0428$$

Step 5.

Now, we will divide Path 1 by the number we arrived at in Step 4 above. The result will be our updated probability, which is P(A|B).

- Path 1 represents the probability of the breathalyzer having a positive result when the driver is truly drunk. It is the equivalent of when we multiply P(B|A) with P(A).

THIS IS P(B|A) P(A)

$$= \frac{.00294}{.0428}$$

THIS IS P(B)

$$= .0686$$

$$= \sim 7\%$$

Conclusion:

The probability of the driver truly being drunk *given that* the breathalyzer tests positive is roughly 7%. That’s low, but it is almost 6x higher than before! We’ll look at this a bit more after we solve using Bayes' formula.

Scenario 2.2 Solve Using Bayes' Formula

Let's solve the same problem but now we'll use Bayes' formula.

Step 1: Determine what you want to find.

We want to know the probability of someone actually being drunk *given that* the breathalyzer test is positive.

Step 2: Write the above as a formula.

$$P(\text{DRUNK} \mid \text{POSITIVE}) = \frac{P(\text{POSITIVE} \mid \text{DRUNK})\, P(\text{DRUNK})}{P(\text{POSITIVE})}$$

Step 3: Find each ingredient and label it.

From the scenario we know the following:

- P(A) - In our formula, this ingredient is represented as P(Drunk) and answers the question: What is the probability of a driver being drunk? This number is .003
- P(B|A) - In our formula, this ingredient is represented as P(Positive | Drunk). This number is .98

The only ingredient we are missing is:

- P(B)

Now, what is P(B)? How do we define what event "B" is so we can try to find it? To find our answers, let's go back to what we are trying to figure out, which we defined in Step 1. Step 1 can be broken into two parts, and P(B) is tucked into the second part.

The two parts of Step 1 are:

1. Probability of a driver being drunk is .003. This is P(A).
2. Probability of the breathalyzer predicting yes P(B) given that the driver is truly drunk P(A) is .98. This is P(B|A), which contains both events "A" and "B". Therefore, event "B" is *the breathalyzer test predicting yes*.

Fantastic. Now we know the definition of event "B", but what about its probability P(B)? We were given a third number in the scenario (4%), but this is not P(B). It is a *part* of P(B). To figure out where it fits and how to solve for P(B) we need to do the following:

Let's think for a moment. How many ways can the breathalyzer arrive at a yes (positive) result? There are only 2 ways:

1. It can produce a positive result that is correct.
2. It can produce a positive result that is false.

PATH DY $= .003 \times .98$

$= .00294$

PATH NDY $= .997 \times .04$

$= .03988$

NOW, ADD THE PATHS

$= .00294 + .03988$

$= .04282$

$= .0428$

All we need to do is multiply the numbers of each *yes* path and then add the answers together. When we do that, we'll have P(B).

Now, the 4% from the scenario comes into play when the breathalyzer has a *yes* result but the driver is not truly drunk. If you look back at the paths we traced above (DY and NDY), you can see the 4% in Path NDY.

Step 4: Plug each ingredient into the formula and solve.

We now have all three ingredients and we can plug them into Bayes' formula!

Here’s the formula:

$$P(A|B) = \frac{\overbrace{P(B|A)}^{\text{INGREDIENT\#1}}\ \overbrace{P(A)}^{\text{INGREDIENT\#2}}}{\underbrace{P(B)}_{\text{INGREDIENT\#3}}}$$

Now, let's plug our numbers into the formula.

$$P(A|B) = \frac{.98 \times .003}{.0428}$$

$$= \frac{.00294}{.0428}$$

$$= .0686$$

$$= \sim 7\%$$

Conclusion:

Based on our calculations, the probability that a driver is truly drunk *given that* the breathalyzer has a positive result is roughly 7%. That is extremely low, but it is higher than the 1.25% we came to in the previous breathalyzer scenario. Knowing this, you will still likely be extremely cautious and skeptical of the test, but not *as cautious* as you were before.

Scenario 3: Peacekeeping

The Scenario: (Expanded from Example Section 1)
Your are a soldier on a peacekeeping mission overseas and were recently caught in a surprise attack, pigeonholing you on a narrow street. A rebel truck had appeared and was making its way towards you, but after a few tense seconds it suddenly turned and disappeared.

Now with the threat gone, you begin to walk down the dusty street and turn a corner, only to quickly turn back. Five men are 100 yards down the street and holding rifles. You peek your head around the corner once more to get a better look.

Are they rebels, or are they part of the coalition? You radio HQ and are told based off of local intelligence (Intel) that the men are likely rebels.

You crouch against the wall as you think about what to do.

- There are roughly 100 rebels in the city and 75 coalition troops.
- Local Intel is not always reliable. In your experience it correctly predicts rebels 65% of the time.
- Intel has been sketchy lately and has incorrectly predicted men as rebels *when they are not rebels* 15% of the time.

To start, we always need to determine what we are wanting to find.

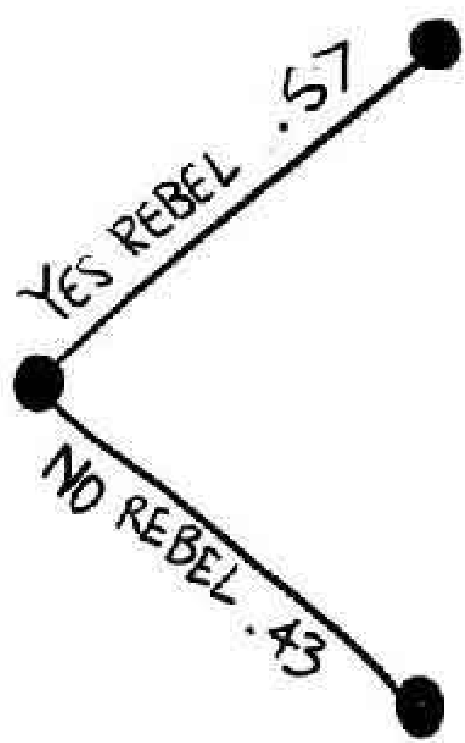
We want to know the probability that the soldiers are rebels *given that* Intel says they are.

Excellent. Now, we can move on to solving the problem by using a decision tree.

Scenario 3.1 Solve Using a Decision Tree

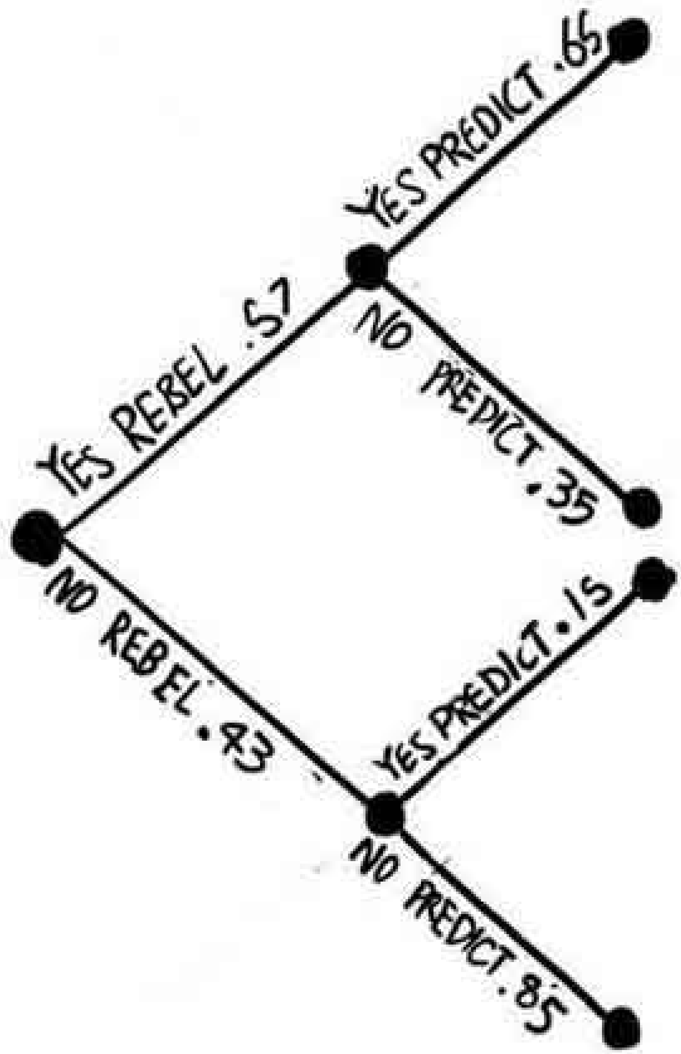
Step 1.

Begin by drawing two lines that connect at a single dot. These lines represent the possibility of someone *being a rebel* and *not being a rebel*. We'll add in the number we know (.57), and then subtract from 1 to get our second number (.43). Both branches add up to 1.



Step 2.

Next, we will add two more branches to each initial branch. Each pair of branches represents the possibility of Intel being *correct* or *incorrect*.We'll add in our numbers from the scenario. Notice each pair of branches adds up to 1.



Step 3.

Now we will label each pathway and compute its value. Each pathway begins with someone being a *rebel* or *not a rebel*, and ends with either a yes or no Intel prediction. There are a total of 4 pathways:

Path 1: Yes Rebel - Predict Yes

- Label YY
- Compute: .57 x .65 = 0.3705
- How did we find .57 ? There are 100 rebel troops and 75 coalition troops, for a total of 175 troops. If we divide the total rebel troops by all troops, we get our answer. 100/175 = .57.

Path 2: Yes Rebel - Predict No

- Label YN
- Compute: .57 x .35 = 0.1995
- How did we find .35 ? Since the probability of both branches together is 100%, or 1, we subtracted .65 from 1. 1-.65 = .35, or 35%.

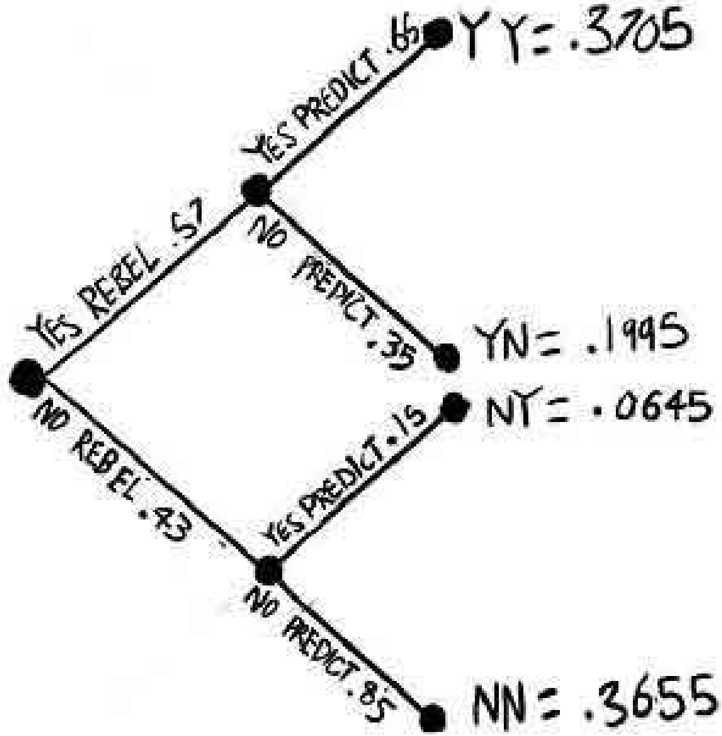
Path 3: No Rebel - Predict Yes

- Label NY
- Compute: .43 x .15 = 0.0645
- How did we find .43 ? There are 100 rebel troops and 75 coalition troops, for a total of 175 troops. If we divide the total coalition troops by all troops, we get our answer. 75/175 = .43

Path 4: No Rebel - Predict No

- Label NN
- Compute: .43 x .85 = 0.3655

- ** How did we find .85 ? Since the probability of both branches together is 100%, or 1, we subtracted .15 from 1. 1-.15 = .85



Step 4.

Once we have our numbers, we need to sum Path 1 and Path 3. Here is why we are doing this:

- The sum of Path 1 and Path 3 tells us the *probability of Intel having a positive (yes) prediction*. In other words it tells us P(B), which is our missing ingredient.
- When we sum Path 1 and Path 3 we are adding both ways Intel can come to a positive (yes) prediction. Path 1 represents Intel predicting "yes" *when the soldier is a rebel*. Path 3 represents Intel predicting "yes" *when the soldier is not a rebel*.When we add these paths together, we find P(B)

PATH 1 = .3705

PATH 3 = .0645

$$= .3705 + .0645$$

$$= .435$$

Step 5.

Now, we will divide Path 1 by the number we arrived at in Step 4 above. The end result will be our updated probability, which is P(A|B).

Path 1 represents the *probability of Intel predicting a yes result when the soldier is a rebel*. It is the equivalent of when we multiply P(B|A) with P(A).

THIS IS P(B|A) P(A)

$$= \frac{.3705}{.435}$$

THIS IS P(B)

$$= .8517$$

$$= 85\ \%$$

Conclusion:

Once we calculate the numbers, we see that the probability the soldiers are rebels *given that* Intel says they are is 85%. Let’s now solve this problem using Bayes', and we will expand on the conclusion once we have done that.

Scenario 3.2 Solve Using Bayes' Formula

Let’s solve the same problem but now we’ll use Bayes' formula.

Step 1: Determine what you want to find.

We want to know the probability that the soldiers are rebels *given that* Intel says they are.

Step 2: Write the above as a formula.

$$P(\text{REBEL} \mid \text{YES PREDICT}) = \frac{P(\text{YES PREDICT} \mid \text{REBEL})\, P(\text{REBEL})}{P(\text{YES PREDICT})}$$

Step 3: Find each ingredient and label it.

From the scenario we know the following:

- P(A) - In our formula, this ingredient is represented as P(Rebel) and answers the question: What is the probability of the soldiers being rebels? This number is .57
- P(B|A) - In our formula, this ingredient is represented as P(Yes Predict | Rebel). This number is .65

The only ingredient we are missing is:

- P(B)

Now, what is P(B)? How do we define what event "B" is so we can try to find it? To find our answers, let's go back to what we are trying to figure out, which we defined in Step 1. Step 1 can be broken into two parts, and P(B) is tucked into the second part.

The two parts of Step 1 are:

1. Probability of a soldier being a rebel is .57. This is P(A).
2. Probability of Intel predicting that a soldier is a rebel P(B) *given that* the soldier is a rebel P(A) is .65. This is P(B|A), which contains both event "A" and "B". Therefore, event "B" is *Intel predicting yes*.

Fantastic. Now we know the definition of "B", but what about its probability P(B)? We were given a third number in the scenario (15%), but this is not P(B). It is a *part* of P(B). To figure out where it fits and how to solve for P(B) we need to do the following:

Let's think for a moment. How many ways can Intel arrive at a yes (positive) prediction? There are only 2 ways:

1. It can produce a positive result that is correct.
2. It can produce a positive result that is false.

PATH YY $= .57 \times .65$

$= .3705$

PATH NY $= .43 \times .15$

$= .0645$

NOW, ADD THE PATHS.

$= .3705 + .0645$

$= .435$

All we need to do is multiply the numbers of each *yes* path and then add the answers together. When we do that, we'll have P(B).

The 15% from the scenario comes into play when Intel predicts a *yes* result but the soldier is not a rebel. It is a false prediction. If you look back at the paths we traced above (YY and NY), you can see the 15% in Path NY.

Step 4: Plug each ingredient into the formula and solve.

Now we have all three ingredients and we can plug them into Bayes' formula.

Here’s the formula:

$$P(A|B) = \frac{\overbrace{P(B|A)}^{\text{INGREDIENT\#1}}\ \overbrace{P(A)}^{\text{INGREDIENT\#2}}}{\underbrace{P(B)}_{\text{INGREDIENT\#3}}}$$

Now, let's plug our numbers into the formula.

$$P(A|B) = \frac{.65 \times .57}{.435}$$

$$= \frac{.3705}{.435}$$

$$= .8517$$

$$= 85\%$$

Conclusion:

At first you had minimal confidence in what Intel told you, but now you see that the probability of the soldiers being rebels *given that* Intel says they are is 85%. This completely changes your outlook and forces you to turn back and wait for the remainder of your team to arrive. If the group are rebels, you don't want to approach them alone.

Example Section 3: Solving For Two Possible Outcomes With All Probability Data Provided

Sometimes we are presented with two possibilities and we want know which is greater. For example, say you are sneezing and coughing. Is it an allergy or cold? Given the information you have (your symptoms), which has the greater probability?

This section will be the same as Example Section 1, except that for each problem we will be using Bayes' Formula twice, and then comparing each posterior probability to find the answer. Basically, we'll be applying Bayes' Formula to both possibilities, and then comparing our answers to find which is the greatest.

Scenario 1: The Flu or Food Poisoning

The Scenario (Continued from Example Section 1)

It's just after lunch and you are feeling sick. You need some space, so you quickly make your way to the parking lot and lie down in your car. Your mind begins to race with what you could have.

Your friend at work recently had the flu, so that's a possibility. You also just had fish at a new restaurant for lunch. Could you have food poisoning? You grab your phone and search for some information. Your hunch is that it's just the flu, but you are not sure. This is what you find:

- You have a slight headache and sore throat, and you see that people with the flu have the same symptoms as you 90% of the time. People with food poisoning have the same symptoms 75% of the time.
- You see that the probability of having the flu is 5%, while the probability of having food poisoning is 16%.
- You then spot one more statistic that says 20% of the population in a given year will have a headache and sore throat at any given time.

After reading this you are just more confused. What probability is higher? Do you have a greater probability of having the flu or food poisoning?

To solve this problem we need to apply Bayes' Theorem to each possibility. Then, we will compare the two to find which is higher, and therefore has a greater probability.

Let's start with the flu. And since we've already pulled this scenario apart in Section 1, we'll skip right to visualizing with a Venn diagram.

Possibility #1: The Flu

The first possibility is that you have the flu. To figure out the probability, we will begin by defining what we are wanting to find. Then, we will visualize the problem before solving it using Bayes' formula.

Step 1: Determine what you are wanting to find.

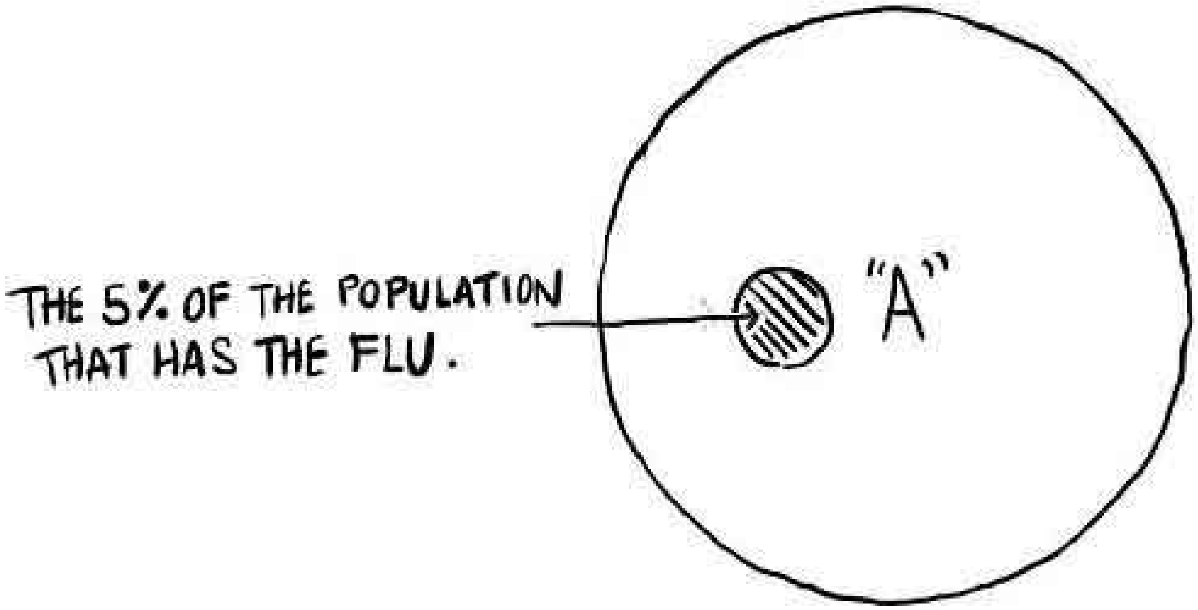
We want to know what the probability is of having the flu given our current symptoms.

Scenario 1.1 Visualize The Problem

To visualize the problem, we'll draw two circles and merge them into a Venn diagram.

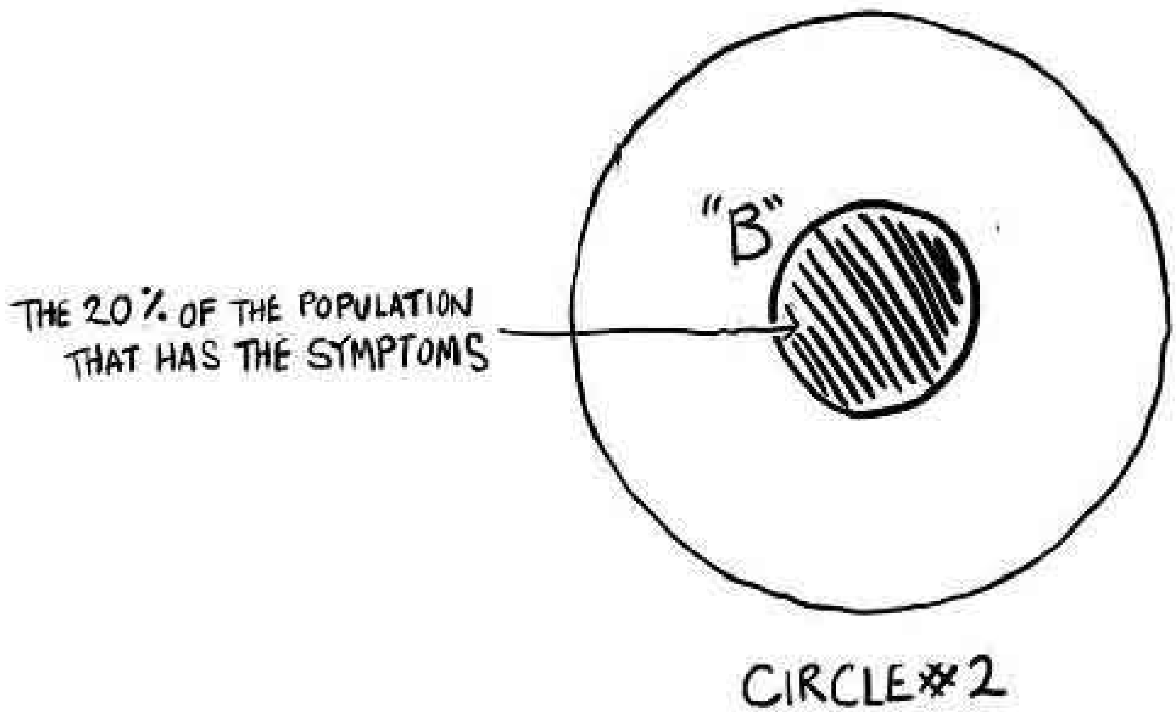
Circle #1: The area inside this circle represents all possible outcomes. In this example, the area represents all people *who could get sick with the flu* - in other words, the entire population. The shaded circle labeled "A" represents the 5% of the population who *have* the flu.

Let's step back now. What does this exactly mean? Within the circle is the entire population, and there are two possible outcomes for the population: people can have the flu, or not have the flu. "A" is an event, and its probability is 5%. This probability is represented in our formula as P(A).



CIRCLE #1

Circle #2: The area inside this circle also represents all possible outcomes. In this instance it represents all people *who could have the symptoms* - this is the entire population. The shaded circle labeled "B" represents the 20% of the population that does have the symptoms. What this means is that within the entire circle there are two possible outcomes: people have the symptoms or do not have the symptoms. "B" is an event, and its probability is 20%. This probability is represented in our formula as P(B).

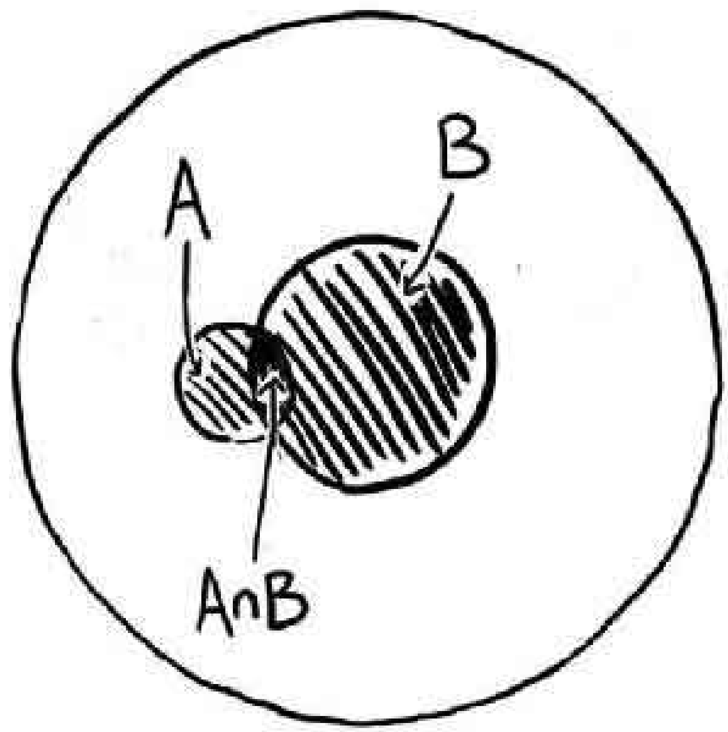


Circle #3: In this circle we have combined both events "A" and "B", and this is where the magic happens!
Here is a quick breakdown of how you can read this:

- The white area inside this circle represents people who do not have either the flu or the symptoms.
- The area where only Circle A covers shows us people who only have the flu.
- The area where only Circle B covers shows us people who only have the symptoms.

Now, take a look at Circle B and see where it overlaps with Circle A. This is what we are *really* interested in! *This is our question from Step 1 in visual form*. We want to know the probability P(A|B) of having the flu given our symptoms.

This probability is found where both events occur together and is called an intersection. Another way to look at it is like this: if we are in area B, what is the probability we are also in area AB (where A and B overlap)?



With both circles now merged, we can visually see our question and what we are trying to solve for. Although we won't be solving the question with a Venn diagram, the diagram does help us visualize what we are trying to understand.

If P(A) is the probability of you having the flu, and P(B) is the probability of you having your symptoms, what is the probability of you having both? While we don't yet know the actual answer we can clearly visualize what we are trying to solve for.

Scenario 1.2 Plugging Into Bayes' Formula And Solving

Now let’s solve the problem by using Bayes' formula. For the sake of ease, we’ll begin by re-stating what we are wanting to find.

Step 1: Determine what you want to find.

We want to know what the probability is of having the flu *given* our current symptoms.

Step 2: Write the above as a formula.

Let’s translate what we are solving for into the formula. In other words, we’ll bring the language of Step #1 above into the formula.

Here is Bayes' formula:

$$P(A|B) = \frac{\overbrace{P(B|A)}^{\text{INGREDIENT\#1}}\ \overbrace{P(A)}^{\text{INGREDIENT\#2}}}{\underbrace{P(B)}_{\text{INGREDIENT\#3}}}$$

Now, let’s translate with what we are solving for.

$$P(\text{FLU}|\text{SYMPTOMS}) = \frac{P(\text{SYMPTOMS}|\text{FLU})\,P(\text{FLU})}{P(\text{SYMPTOMS})}$$

Step 3: Find each ingredient and label it.

From the scenario we know the following:

*We have changed the ingredients provided in the scenario from percents into decimals. We will do this every time before we begin to plug the ingredients into the formula.

- P(A) - In our formula, this ingredient is represented as P(Flu) and answers the question: What is the probability of you having the flu? This number is .05
- P(B|A) - In our formula, this ingredient is represented as P(Symptoms|Flu). This number is .9
- P(B) - In our formula, this ingredient is represented as P(Symptoms) and answers the question: What is the probability of you having the symptoms? This number is .2.

Step 4: Plug each ingredient into the formula and solve.

$$P(A|B) = \frac{.9 \times .05}{.2}$$

$$= .225$$

$$= 22.5\%$$

Conclusion:

So, after all the work and plugging each ingredient into the formula, our answer is 22.5%. Now, let’s move on and look at food poisoning. Once we have calculated the probability of food poisoning, we'll be able to stack each probability side by side to see which is greatest.

Possibility #2: Food Poisoning

The second possibility is that you have food poisoning. To figure out the probability, we will begin by defining what we are wanting to find. Then, we will visualize the problem before solving it using Bayes' Formula.

Step 1: Determine what you are wanting to find.

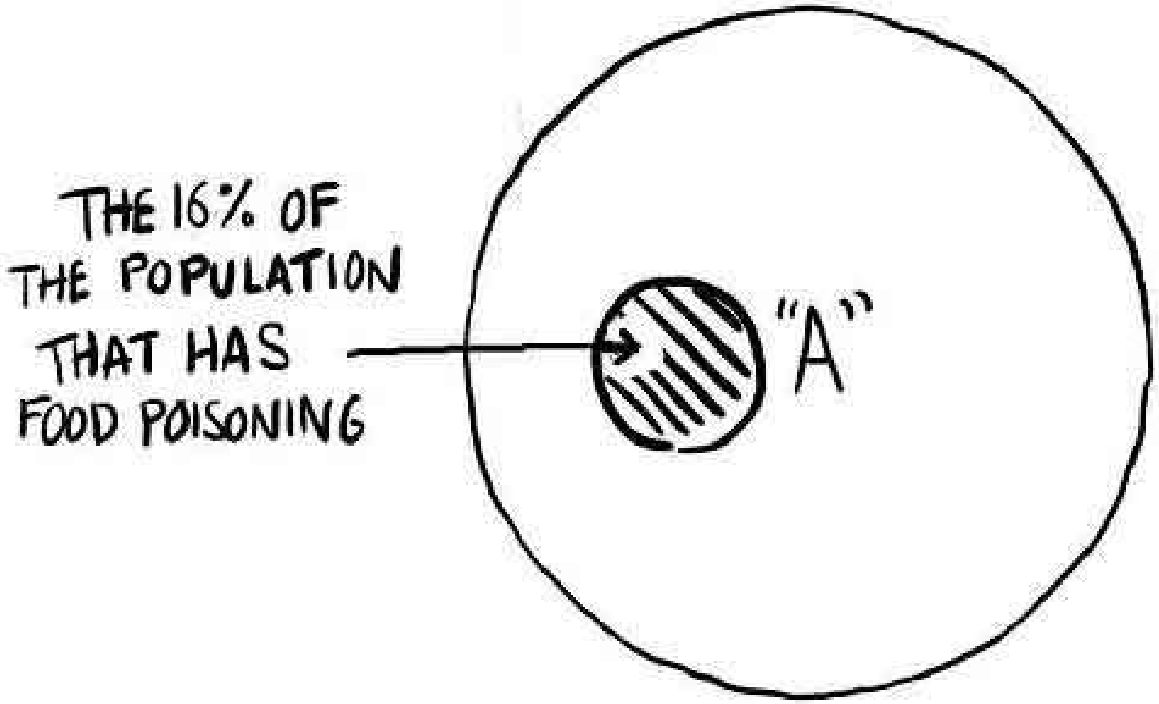
We want to know the probability of having food poisoning *given* our current symptoms.

Scenario 2.1 Visualize The Problem

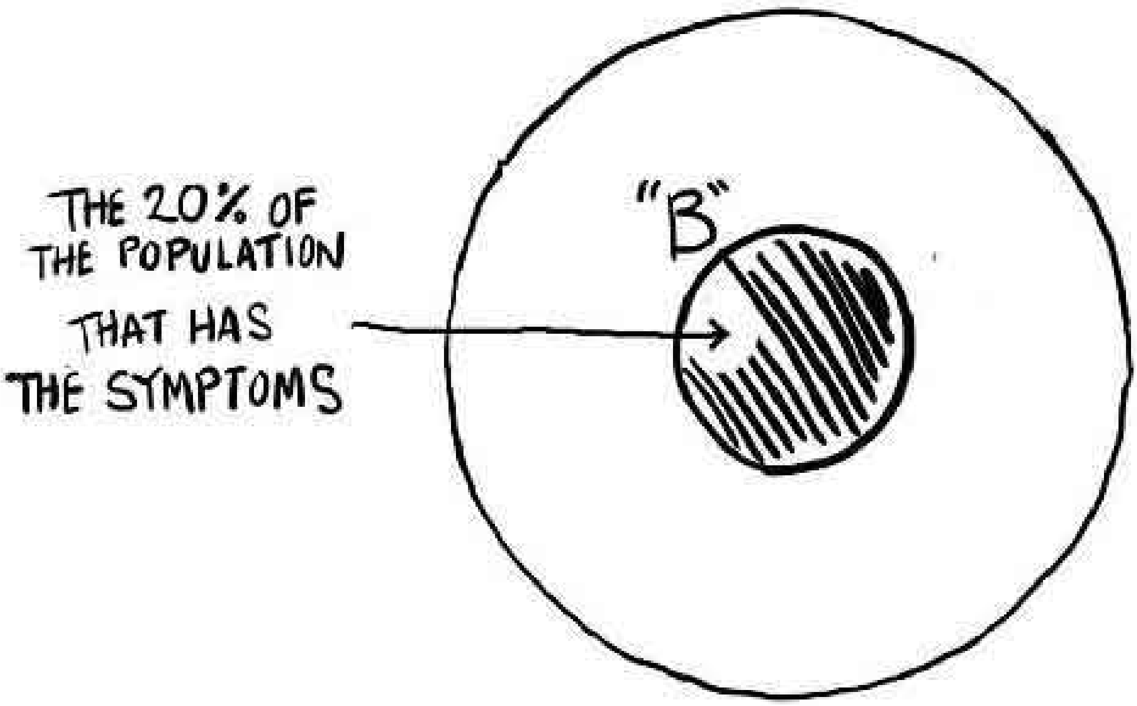
To visualize the problem, we'll draw two circles and merge them into a Venn diagram.

Circle #1: The area inside this circle represents all possible outcomes. In this example, the area represents all people *who could have food poisoning* - in other words, the entire population. The shaded circle labeled "A" represents the 16% of the population who *do have food poisoning*.

Let's step back now. What does this exactly mean? Within the circle is the entire population, and there are two possible outcomes for the population: people can have food poisoning, or not have food poisoning. "A" is an event, and its probability is 16%. This probability is represented in our formula as P(A).


THE 16% OF
THE POPULATION
THAT HAS
FOOD POISONING
"A"

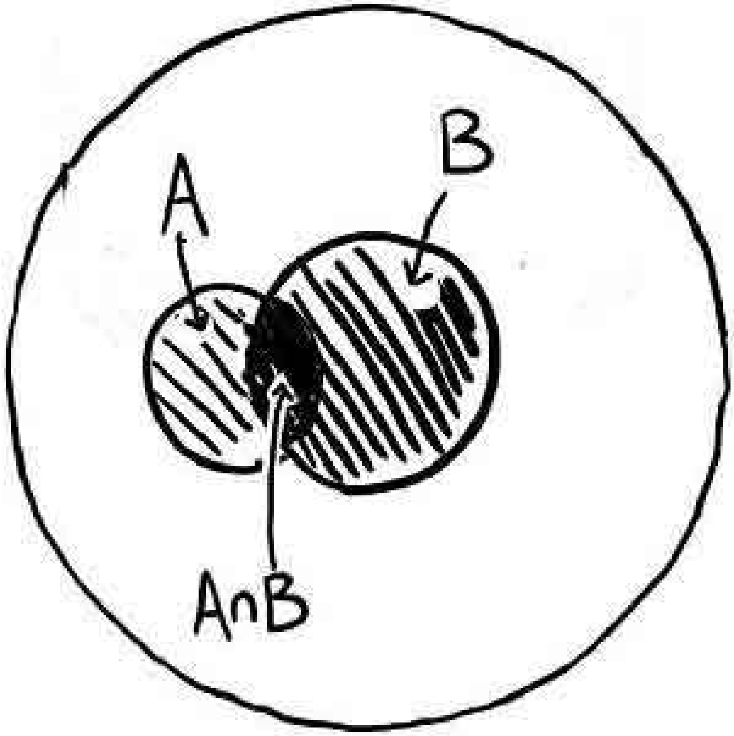
Circle #2: The area inside this circle also represents all possible outcomes. In this instance it represents all people *who could have the symptoms* - this is the entire population. The shaded circle labeled "B" represents the 20% of the population that does have the symptoms. What this means is that within the entire circle there are two possible outcomes: people have the symptoms or do not have the symptoms. "B" is an event, and its probability is 20%. This probability is represented in our formula as P(B).



Circle #3: In this circle we have combined both events “A” and “B”, and this is where we want to look!

Here is a quick breakdown of how you can read this:

- The white area inside this circle represents people who do not have either food poisoning or the symptoms.
- The area where only Circle A covers shows us people who only have food poisoning.
- The area where only Circle B covers shows us people who only have the symptoms.
- Now, take a look at Circle B and see where it overlaps with Circle A. This is what we are really interested in! This is our question from Step 1 in visual form. We want to know the probability of having the flu P(A) given our symptoms P(B). This probability is found where both events occur together and is called an intersection.



With both circles now merged, we can visually see our question and what we are trying to solve for. If P(A) is the probability of you having food poisoning, and P(B) is the probability of you having your symptoms, what is the probability of you being where they overlap - and having both?

Scenario 2.2 Solve Traditionally With Bayes' Formula

To solve by using Bayes' Formula we'll follow four steps. To make things easy, we'll begin by re-stating what we are wanting to find.

Step 1: Determine what you want to find.

We want to know the probability of having food poisoning *given* our symptoms.

Step 2: Write the above as a formula.

Let's translate what we are solving for into the formula. In other words, we'll bring the language of Step #1 above into the formula.

Here is Bayes' formula:

$$P(A|B) = \frac{\overbrace{P(B|A)}^{\text{INGREDIENT\#1}}\ \overbrace{P(A)}^{\text{INGREDIENT\#2}}}{\underbrace{P(B)}_{\text{INGREDIENT\#3}}}$$

Now, let's translate with what we are solving for.

$$P(\text{POISONING}|\text{SYMPTOMS}) = \frac{P(\text{SYMPTOMS}|\text{POISONING})\ P(\text{POISONING})}{P(\text{SYMPTOMS})}$$

Step 3: Find each ingredient and label it.

From the scenario we know the following:
*We have changed the ingredients provided in the scenario from percents into decimals. We will do this every time before we begin to plug the ingredients into the formula.

- P(A) - In our formula, this ingredient is represented as P(Poisoning) and answers the question: What is the probability of you having food poisoning? This number is .16

- P(B|A) - In our formula, this ingredient is represented as P(Symptoms / Poisoning). This number is .75

- P(B) - In our formula, this ingredient is represented as P(Symptoms) and answers the question: What is the probability of you having the symptoms? This number is .2

Step 4: Plug each ingredient into the formula and solve.

$$P(A|B) = \frac{.75 \times .16}{.2}$$

$$= \frac{.12}{.2}$$

$$= .6$$

$$= 60\%$$

Conclusion:

The probability that you have food poisoning is ~60%. This is almost 3x the probability of you having the flu at 22.5%! Based on this, you probably want to head to a clinic to have a doctor check you over. If it was the flu, you would likely head home and rest. But if you have food poisoning you might need antibiotics to recover.

This problem is a great example of how Bayes' formula can help us in everyday life. We all face sickness at some point, and when we do many of us typically jump to the easiest conclusions. Using Bayes' formula' can help us look past the easy answer and assess all the data.

Once we've done that, we can have a better understanding of the uncertainties we face. In our scenario, doing this helped us realize that we should leave work and head to the nearest doctor.

History and Stories

Bayes' Theorem has a spectacular 200-year history, and there are a number of fantastic books that delve deeply into it. There are also many interesting stories about how Bayes is used in real life. In this section, we'll take a brief look at both, beginning with the history of Bayes' Theorem.

The Discovery Of Bayes' Theorem

1740's, Edinburgh, Scotland

Thomas Bayes was born in London, England in 1702. Bayes studied logic and theology at the University of Edinburgh and served as a minister at a Presbyterian church in Tunbridge, Wells (35 miles from London). He retired in 1752 and passed away in 1762.

Apart from his faith, Bayes had a deep love and interest in mathematics and was considered an amateur mathematician. In his later years, he became fascinated with probability, specifically inverse probability. No one knows why for sure, but one thing is clear: Bayes became consumed with figuring out *...the approximate probability of a future event he knew nothing about except its past, that is, the number of times it had occurred or failed to occur*[1].

Sometime in the 1740's Bayes figured out a solution that can be summed up as follows: *An Initial Belief + New Evidence = A New, Updated Belief*. But, for reasons unknown he never did anything with it and instead it sat on the solution until his death in 1762. After he passed away his friend Richard Price discovered the theorem, saw its importance, and for two years worked on it. Price then submitted it to The Royal Society and it was published a year later.

Across the English Channel in 1774, a French mathematician named Pierre-Simon Laplace recreated Bayes' solution. At the time he was completely unaware of Bayes' discovery, and had come to it on his own merit. For almost 40 years he worked on the theorem, and sometime between 1810 and 1814 he put the final touches on it, eventually (more or less) creating the formula that is known today as *Bayes' Theorem*. Laplace did most of the work, but Bayes name stuck with theorem for some reason.

However, Bayes' Theorem was generally not accepted and had few advocates until recently in modern history. From Laplace's final touches on the theorem until the mid 1960's, Bayes' Theorem was criticized as being subjective due to the formula's *prior*. Most mathematicians considered the formula taboo and would not touch it.

During World War 2 it was used by Alan Turing (a brilliant mathematician and played by Benedict Cumberbatch in *The Imitation Game)* to break the German's Enigma Code, but that information was classified for a long time after the war. It wasn't until the 1980's and the introduction of personal computing that Bayes' Theorem became widely accepted. It is now used across many sectors and likely touches your life on a daily basis.

Real Life: Search and Rescue

What if you were lost at sea? How could search and rescue find you?

In 2014 a missing fisherman was discovered using Bayes' Theorem. It is a fascinating story that *The New York Times* reported on here and here, but we'll sum it up for you below. Plus, we'll also include an idea of *how* the theorem could have been used to find the missing individual.

Here's the story:

John Aldridge is a fisherman who has fished for almost two decades. In 2006 Aldridge and his business partner Anthony Sosinski purchased their own boat and by 2014 had a successful crab and lobster operation. On July 24th of 2014 the partners were 40 miles off the coast of Long Island catching lobster.

During the night while Sosinksi was sleeping, Aldridge was working and fell into the Atlantic. Using his rubber boots as pontoons, he managed to stay afloat in the 72 degree water until he was found almost 12 hours later by the Coast Guard.

To find Aldridge, the Coast Guard used a piece of software called SAROPS (Search and Rescue Optimal Planning System). SAROPS creates a probability map that displays potential areas *where survivors could be located* based on input information.

This map is then continually updated to reflect changes, additional clues, and areas where no survivors have been found. In Aldridge's case, the Coast Guard only had two pieces of initial information to work with: a rough location and timeframe (he fell off his boat sometime during the evening of July 24th).

Once this data had been uploaded into SAROPS, it was continually updated with ocean currents, new data from Sosinksi, clues and additional wind patterns.

A *very simple form* of how Bayes' Theorem could be used in this situation is as follows. Please note that this is greatly simplified to demonstrate the possibility of using the theorem. In this example we have abbreviated the text:

P. Found = Person Found
Un. Search = Unsuccessful Search.

$$P(\text{P FOUND} \mid \text{UN. SEARCH}) = \frac{P(\text{UN. SEARCH} \mid \text{P FOUND})\,P(\text{P FOUND})}{P(\text{UN. SEARCH})}$$

Again, the above is only a guess of how Bayesian Inference could be used in the search and rescue process. For more information, you can read more on SAROPS here. Wikipedia's article on Bayesian Search Theory and this blog post are also helpful.

Bayes' Theorem has been used in many search and rescue operations, including finding the USS Scorpion in 1968 and Air France Flight 447 in 2009.

Real Life: Spam Filtering

If you hate spam, you love Bayes' Theorem.
Yep. It's true, even if you don't have a clue about what Bayes' Theorem is.

Spam filtering has *really improved* over the last decade to the point where most of us don't think about spam anymore. This is a fantastic and welcomed change from not too long ago. But how did spam filtering progress and become much more effective? After all, it wasn't too long ago when we were regularly inundated with lots of junk mail.

In 1998 Microsoft applied for a spam filter patent that used a Bayesian filter, and in doing so ignited a new war on spam - Bayesian style. Other competitors soon joined in and Bayes' Theorem quickly became the backbone of spam filtering.

Bayesian filters determine if an email is spam or not based on the email's content. When an email is received, each word is read and the filter determines the probability of it being spam or legitimate (often defined as spam or ham). What sets Bayesian filters apart from other email filters though is that they *learn and adapt to each individual email user.* And that, in a nutshell, is why they are so effective.

Here's how it works:
Spam filters built on a Bayesian network are typically pre-populated with a list of potential words and characteristics that spam contains. This list is usually derived from feeding the filter *copious amounts of both spam and legitimate emails,* which the filter uses to *learn* what constitutes spam. Take a moment and think of spam you've seen. What words pop up?

- Usually anything to do with sex
- Ink
- Deals
- Secret

- This list could on and on...

Where does the filter look for these words and phrases? The filter analyzes:

- Words in the body of the message.
- Words in the header.
- Words in the meta data.

This list is continually updated as each email is received and the filter *learns more and more* about what to look for. The filter learns in two different ways:

- Based off its own decisions.
- Based off the user's decisions (For example, you check an email as *spam*).

For example, if the word *sports* often appears in your email, the filter might conclude that the word *sports* has a very low probability of being spam (on a scale of 0-1, maybe a .1). However, if you never include *deals* with the word *sports*, the filter might conclude that an email with this phrase has a high probability of being spam (on a scale of 0-1, .65).

So, how exactly could Bayes' Theorem be used in detecting spam? Here is a *simplified* version. *Again, please note this is *very simplified* to demonstrate the concept of using Bayes' Theorem with spam filtering. There are many more complexities that are involved, e.g., calculating priors, etc.

In this example we have abbreviated the text:
C. Word = Certain Word

$$P(\text{SPAM} \mid \text{C WORD}) = \frac{(\text{C. WORD} \mid \text{SPAM})\, P(\text{SPAM})}{P(\text{C. WORD})}$$

If you are interested in learning more, we would recommend Wikipedia’s article, and Microsoft's paper on Junk Filters.

Real Life: Driverless Cars

Many people believe that not too far into the future our driving experience will change for the better. Proponents of driverless vehicles believe that cars will eventually drive themselves and offer a much safer experience.

From Tesla to Google and Ford, driverless cars are being developed and tested. They are even being deployed, with Uber recently launching its first self-driving fleet in Pittsburgh this past August, 2016.

Google is an early pioneer in the driverless car field, and in 2009 began its first tests with a Toyota Prius on the highways of California. Google's cars have currently (Sep 2016) driven more than 1.5 million miles all over the nation from Arizona to Washington, Texas and California.

Sebastian Thrun is a former VP at Google who helped launch Google's driverless car project, Google Glass, and the Loon project. He currently is a research professor at Stanford and the founder/CEO of Udacity. According to an article on Stanford's website, Google's driverless cars use a form of Bayesian modeling to help the vehicle know its location. It's all quite complicated, but here is a quote from the Stanford article.

For localization, the underlying model is generally some form of Bayesian model similar to a hidden markov model where the state space of the the unknown "location" variables are continuous. At each time point t there are two variables: an unknown variable which is the location of the car, x(t), and observations about the car's location based on the sensor inputs at that given time, y(t). The model assumes that x(t) is generated from x(t - 1) with some unknown distribution and that y(t) is generated from x(t) with some unknown distribution.

For more in-depth information, be sure to visit Google's self-driving page and read Time Magazine's fascinating article.

Thinking Like a Bayesian

Let's Get Real.

If you are like many people, you are probably excited about Bayes' Theorem but deep down you doubt it's useful in real life. Bayes' Theorem might seem interesting in a textbook, but what about beyond that? The fact is, it's often taught with examples that include rolling dice, gambling, cancer screenings, and *game shows*. Really? What can we do with that?

Most people don't have cancer screenings or go on game shows on a regular basis. Some people might gamble often, but the average person probably doesn't. So, after learning Bayes' Theorem we are left with the question of what to do with it.

Is Bayes' Theorem useful in everyday life? Is it practical? Can you take what you've learned in this book and *actually* use it? And if even if you can, will you benefit from it?

The answer to these questions is a resounding yes, and here is why:

- First, thinking like a Bayesian can be applied to almost any area of life.
- Second, thinking like a Bayesian can help you make better decisions *and* feel more confident about the decisions you've made.

In this section we'll offer some practical advice on how to use Bayesian thinking as you work, play, and live life. The ideas we present in this section are based off a phenomenal video by Julia Galef which you can find on YouTube.

Don’t Panic:

If you are panicking and thinking *yeah right*, *I’m not going to write the formula out*...well, you are in luck. By using Bayes' Theorem we *don’t mean* writing out the formula all the time on whatever scrap of paper you can find.

That isn’t practical. And it isn’t enjoyable either. The key to using Bayes' Theorem in *everyday decision making* is to make rough calculations quickly in your mind that will help you discover the *odds* of an event, not the *probability*.

That is really the *only way* the theorem can be used quickly on a daily basis.

If you want to discover *probability*, you’ll need to make use of the formula and pull out a pen and paper. BUT if you are okay with finding the *odds* of something, then you can likely do everything in your head. No pen or paper required. And you know what? Most people intuitively understand odds better than percentage. It’s true! You can read more about this via an article in the CIA Library.

To find the posterior odds, you’ll need to become familiar enough with Bayes' formula and its ingredients so that you know *what to ask and look for*. Then you can literally use this *any time* to help you make better decisions. And by anytime, we *literally* mean anytime. You could be in a meeting, in a classroom, exercising, or at the mechanics.

If it helps, you can visualize problems as you work through them in your mind. We’ve included visuals for each example in this section to help you get a start.

Scenario 1: Dating

Maybe you are currently dating someone. Maybe you wish you were. Or maybe you are living a happy life with your partner. Regardless, dating is an activity most of us can relate to.

So, here's one way of how Bayes' Theorem can be applied to dating.

You are dressed in your finest and on a date one Friday night. You've just finished having dinner at a restaurant and your date has excused herself to use the restroom. Now that you've got some time alone your heart begins to race. *Does she like you? Is she into you?* You then think through what happened during the meal and want to curl up into a ball. You almost choked on your pasta and kept on mumbling.

As your date makes her way back to the table you can feel your pulse quickening. *There's no way she likes you*. *You could barely talk and almost choked*. Your face begins to turn red as you force a smile.

Ten minutes later your cheeks are no longer burning, and you're in the middle of a *really good* conversation. She's even laughed and seemed to be flirting with you.

Fifteen minutes later you walk out the restaurant door and your head is spinning. You are more confused than ever. *She's been laughing and flirting with you*. *Does she like you?*

We'll approach this two ways:

1. Non-Visual Approach

2. Visual Approach

Non-Visual Approach

What we are going to do is find *four numbers and quickly compute them*. Remember, all numbers are rough *beliefs* that you are quickly putting together in your mind. We'll keep everything super simple.

Step 1:

Let's assign some probability numbers.

Number 1: Guess the probability that *she likes you* regardless of anything else. Based on your past dating success, you'd say this is 20%. You think at least 2/10 girls will like you.

- Number 1 = 2
- **to keep it simple all our numbers will be whole between 1 and 10 and have no decimal place.

Number 2: Guess the probability that *she doesn't like you* regardless of anything else. Based on our answer above, if your past history tells you that your dates like you 2/10 times, that means your date will *not like you* 8/10 times, or 80%.

- Number 2 = 8

Number 3: Guess the probability of your date laughing and flirting *if she likes you*. Considering that she is flirting, you guess that there is a 90% chance of this.

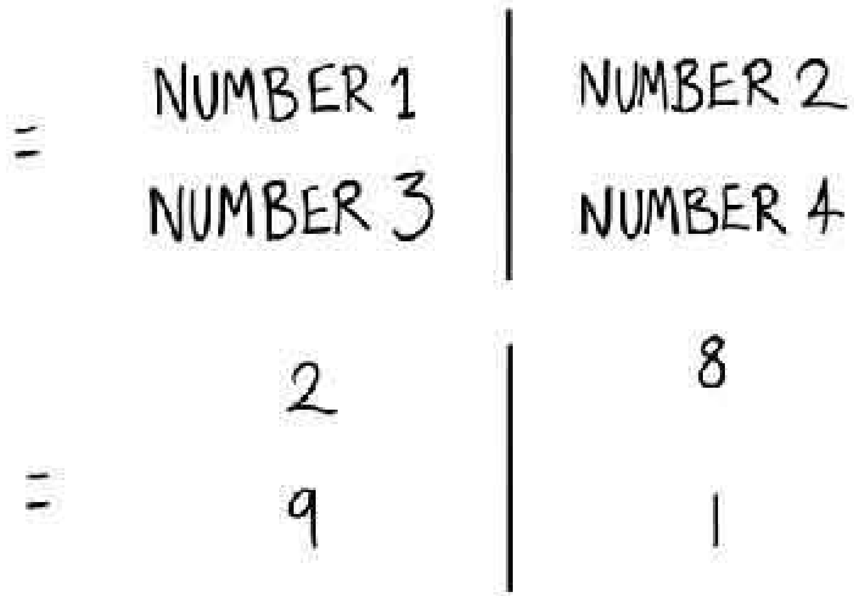
- Number 3 = 9

Number 4: Guess the probability of your date laughing and flirting if she *doesn't like you*. You think this is low, so you guess 10%.

- Number 4 = 1

Step 2:

Line your numbers up as ratios.



Step 3: Multiply down and simplify.

$$= \begin{array}{c|c} 2 & 8 \\ 9 & 1 \end{array}$$

$$= \begin{array}{c|c} 18 & 8 \end{array}$$

$$= \begin{array}{c|c} 9 & 4 \end{array}$$

= POSTERIOR ODDS ARE ROUGHLY 2:1

Conclusion:

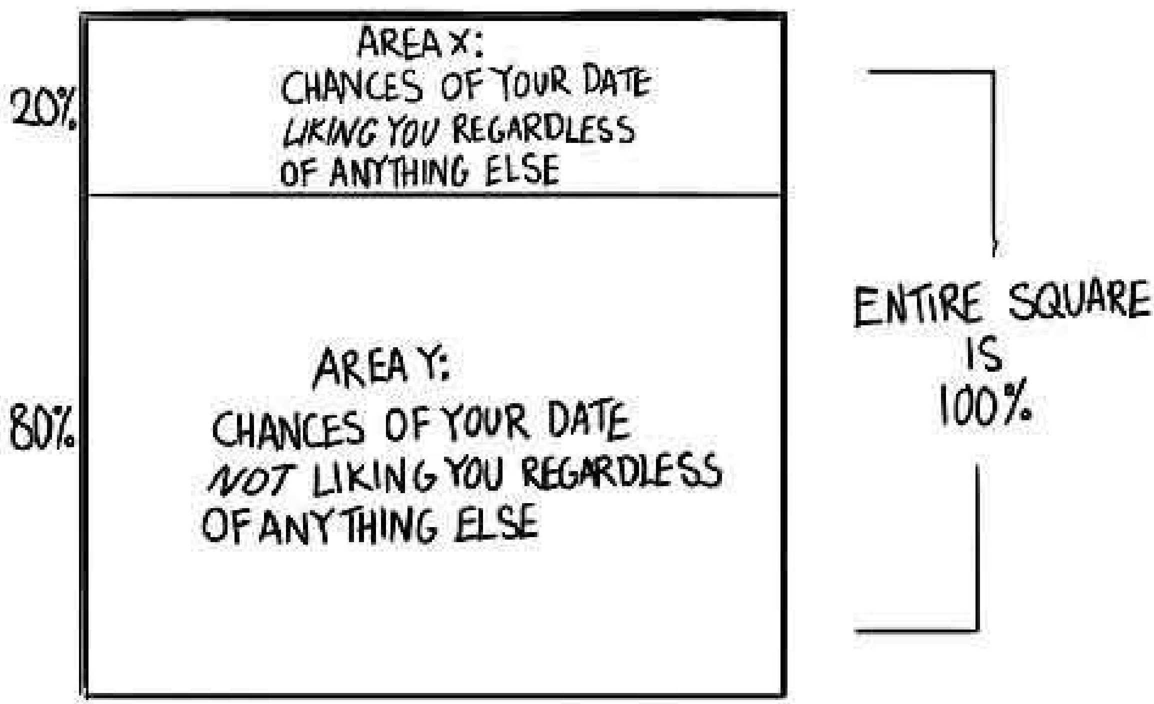
Based on a quick calculation, the odds that she *likes you* are 9 to 4, or slightly greater than 2 to 1. For simplicity, we'll stay with the 2 to 1 ratio. What does this mean? This means it is roughly twice as likely that she likes you as opposed to *not liking you*. And that's a good thing, right?!

Visual Approach

Here is the same problem but with a visual explanation.

Step 1:

Draw a square. Imagine the inside of the square represents 100%. Now, draw a horizontal line through the square to create an area that represents the chances of your date *liking you* regardless of anything else.



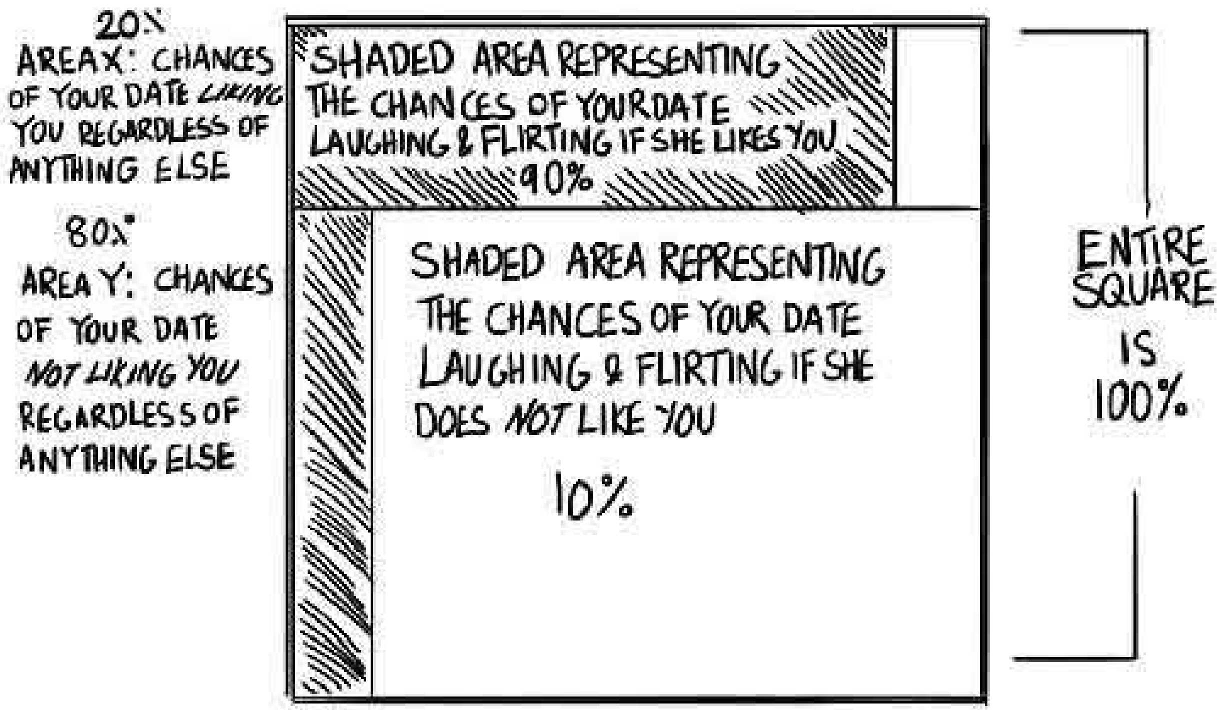
In Step 1 from the non-visual approach we guess 20%, so we will draw a line that creates a space of roughly 20% of the entire square. We'll call this Area X. Got it? Okay, now the remaining 80% of the square represents the possibility of her *not liking you* regardless of anything else. We'll call this Area Y.

We now have two areas with numbers.**to keep it simple all our numbers will be whole between 1 and 10 and have *no decimal place*.

- Area X = 2
- Area Y = 8

Step 2:

Now, if your date likes you, how likely is it that she would flirt and laugh? *Shade in* Area X with your answer. In our example, we chose 90%. What about if your date doesn't like you? How likely is it that she would flirt and laugh with you then? *Shade in* Area Y to show this. In our example, we chose 10%.



Step 3:

We now have four numbers. Let's line them up as an ratios, then multiply down and simplify to find our answer! Area X is on the left, and Y is on the right.

$$= \quad \begin{array}{c|c} 2 & 8 \\ 9 & 1 \end{array}$$

$$= \quad \begin{array}{c|c} 18 & 8 \end{array}$$

$$= \quad \begin{array}{c|c} 9 & 4 \end{array}$$

= POSTERIOR ODDS ARE ROUGHLY 2:1

Conclusion:

We arrive at the same answer as before. Based on a quick calculation, the odds that she likes you are 9 to 4, or slightly greater than 2 to 1. For simplicity, we'll stay with the 2 to 1 ratio. The left (2) represents the chance of her liking you and the right (1) represents the chance of her not liking you. This means your date is twice as likely to like you as opposed to not liking you. Hopefully knowing this can help you take a deep breath and enjoy the rest of the date!

Scenario 2: Can You Trust Your Mechanic ?

You've recently moved to a new city with your family. One morning you are stuck in traffic when your car starts to make odd noises. Once you arrive at work the noises have increased and you begin to worry.

During lunch you ask someone if they know a trustworthy mechanic, and they recommend Joe's Garage. You Google Joe's and see that there are no reviews online, but since your friend recommended it and you are in a pinch, you drop your car off at Joe's after work.

Now, once at home you research Joe's a bit more and discover a few reviews that you had missed before. There are good reviews, but the bad ones are *really bad* and you begin to sweat. Have you made a good choice? Is Joe an honest mechanic?

We'll approach this two ways:

1. Non-Visual Approach
2. Visual Approach

Non-Visual Approach

We'll follow the same steps as in our previous scenario. What we are going to do is find *four numbers and quickly compute them*. And don't forget, all numbers are rough *beliefs* that you are quickly putting together in your mind. We'll keep everything simple.

Step 1:

Let's assign some probability numbers.

Number 1: Guess the probability that *the mechanic is honest* regardless of anything else. Based on your history with mechanics, you'd say that the majority are honest. You think at least 7/10 mechanics are honest, or 70%.

- Number 1 = 7
- **to keep it simple all our numbers will be whole between 1 and 10 and have *no decimal place*.

Number 2: Guess the probability that he *is not honest* regardless of anything else. Based on our answer above, if your past history tells you that 7/10 mechanics are honest, that means 3/10 are not honest, or 30%.

- Number 2 = 3

Number 3: Guess the probability of your mechanic having bad reviews *if he is honest*. People will always complain even if service is fantastic, but not as often or regularly. So, you guess that there is a 30% chance of this.

- Number 3 = 3

Number 4: Guess the probability of your mechanic having bad reviews *if he is dishonest*. People complain loudly if someone rips them off, so you think this would be pretty high. So, you guess that there is a 90% chance of this.

- Number 4 = 9

Step 2:

Line your numbers up as ratios.

$$= \begin{array}{c|c} \text{NUMBER 1} & \text{NUMBER 2} \\ \text{NUMBER 3} & \text{NUMBER 4} \end{array}$$

$$= \begin{array}{c|c} 7 & 3 \\ 3 & 9 \end{array}$$

Step 3:

Multiply down and simplify.

$$= \begin{array}{c|c} 7 & 3 \\ 3 & 9 \end{array}$$

$$= \begin{array}{c|c} 21 & 27 \end{array}$$

$$= \begin{array}{c|c} 7 & 9 \end{array}$$

= POSTERIOR ODDS ARE 7:9

Conclusion:

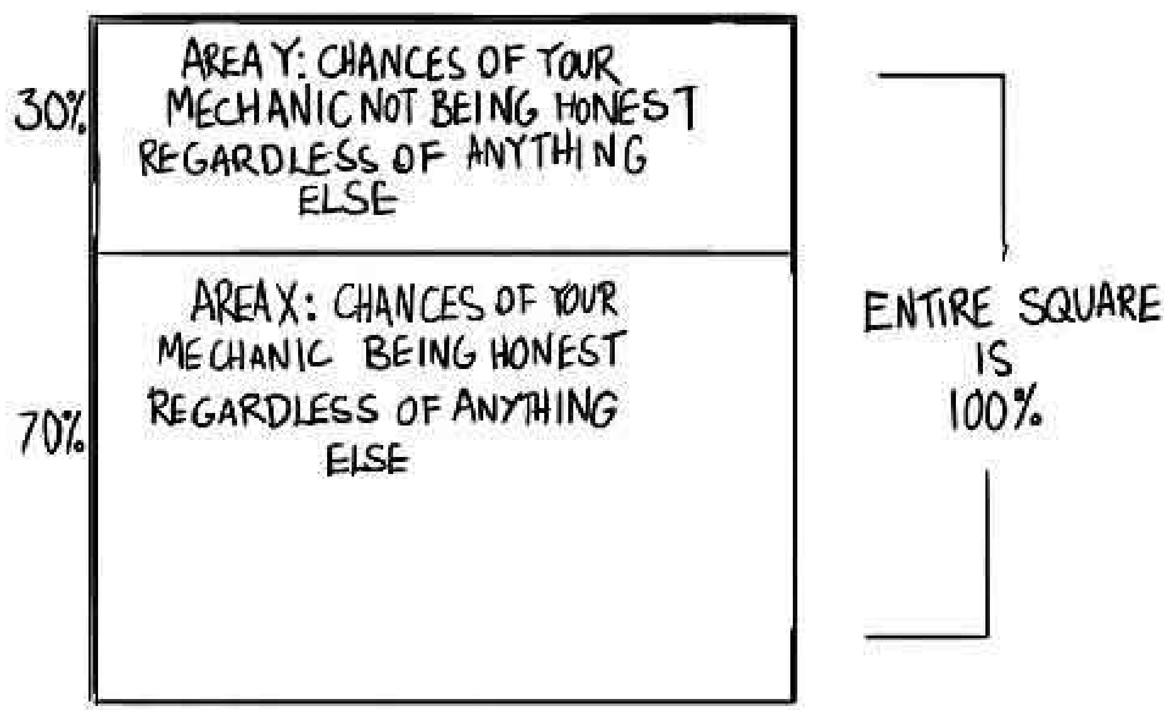
After crunching the numbers, we can see that the odds are *greater* for your mechanic being dishonest than honest. The odds are 7:9. The left (7) represents the chance of him being honest, and the right (9) represents the chance of him being dishonest. Knowing this, you'll want to be careful and inspect everything when you pick up your car. You might also want to bring a friend who knows about mechanics to the shop with you.

Visual Approach

Here is the same problem but with a visual explanation. Same steps as the first scenario!

Step 1:

Draw a square. The inside of the square represents 100%. Now, draw a horizontal line through the square to create an area that represents the chances of your mechanic *being honest* regardless of anything else.



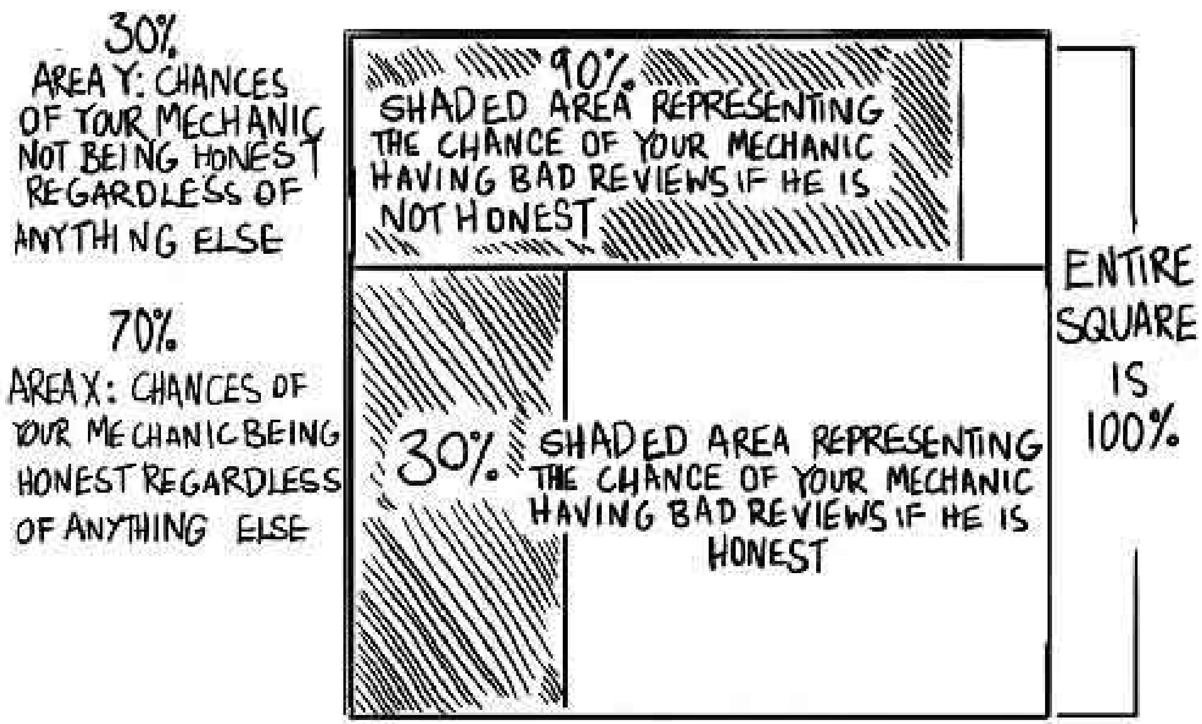
In Step 1 of our example we guess 70%, so we will draw a line that creates a space of roughly 70% of the entire square. We'll call this Area X. Got it? Okay, now the remaining 30% of the square represents the possibility of your mechanic *not being honest* regardless of anything else. We'll call this Area Y.

We now have two areas with numbers.**to keep it simple all our numbers will be whole between 1 and 10 and have *no decimal place*.

- Area X = 7
- Area Y = 3

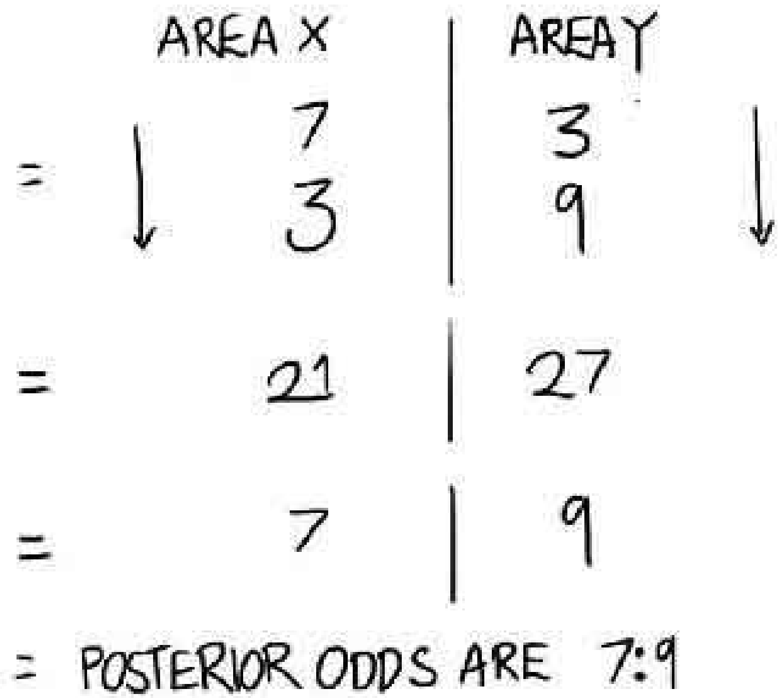
Step 2:

Now, if your mechanic is honest, how likely is it that he would have bad reviews? *Shade in* Area X with your answer. In our example, we chose 30%. What about if he is not honest? How likely is it he would have bad reviews? *Shade in* Area Y to show this. In our example, we chose 90%.



Step 3:

We now have four numbers. Let's line them up as an ratios, then multiply down and simplify to find our answer!



Conclusion:

Our conclusion is the exact same as when we solved this non-visually. We can see that the odds are *slightly greater* for your mechanic being dishonest than honest. The odds are 7:9. How do you read this fraction? The left (7) represents the chance of him being honest, and the right (9) represents the chance of him being dishonest. Knowing this, you'll want to be careful and inspect everything when you pick up your car. You might also want to bring a friend who knows about mechanics to the shop with you.

Reference: Definitions, Notations, and Proof

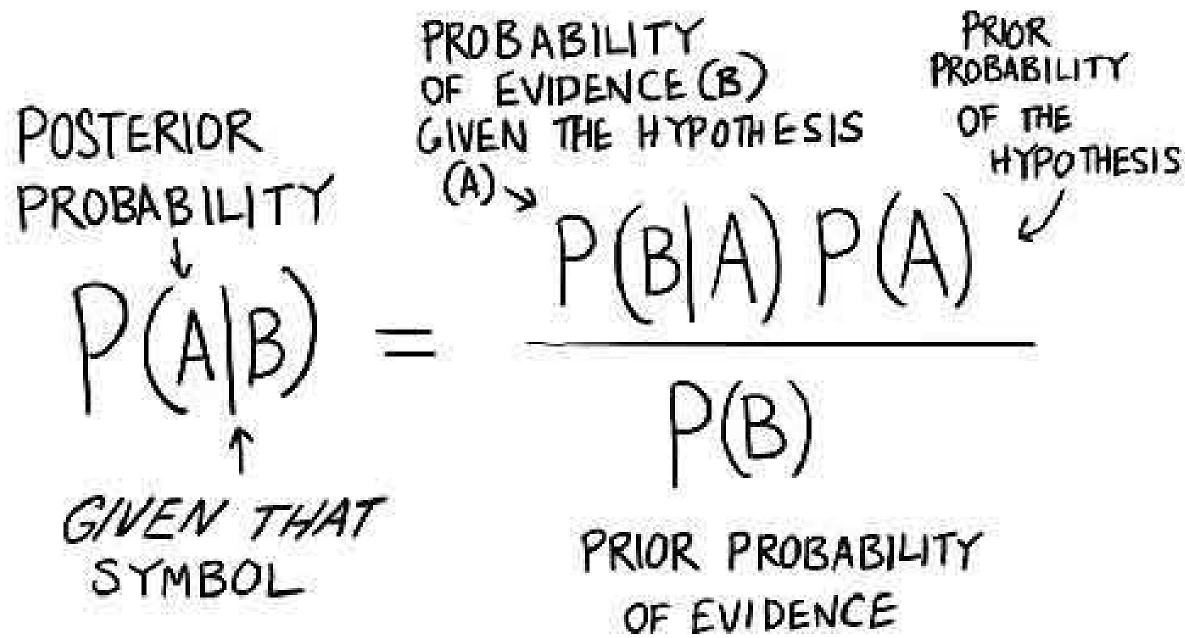
Welcome to our expanded reference section. Feel free to skip around to find what you need. Here's what we cover in this section:

Definitions

1. Bayes' Theorem Formula

We have expanded on the original definitions given in the Basic Overview. All of the definitions from both formulas in the Basic Overview still apply. We are simply building on this by adding in the terms that we will define below.

2. Probability



Probability is the chance that something will happen. Technically speaking, it is a branch of mathematics that deals with calculating the likelihood of an event's occurrence. Probability is expressed as a number between 1 and 0.

- A probability of 1 means that there is a 100% likelihood of an event occurring.

- A probability of 0 means that there is a 0% likelihood of an event occurring.

Classic Coin Toss Example:

- *Question:* Let's say you have a coin. If it is a fair (properly weighted) coin and tossed 3 times, what is the probability of 3 heads?
- *Answer:* There are 8 possibilities if the coin is tossed three times. In this example, *T* stands for Tail and *H* stands for Head.

- Possibilities: HHH, HHT,HTT,HTH,TTT,TTH,THH,THT
- All outcomes are equally possible, so the probability of tossing 3 heads in a row is ⅛, or 12.5%.

3. Conditional Probability

Conditional probability is the probability of event "A" being true *given that* event "B" is true. It helps us answer the question: *How does the probability of an event change if we have new evidence (or new information)?* Conditional probabilities always involve *two independent events,* "A" and "B".

Classic Coin Toss Example Con't:

- *Question:* In the example above we concluded that the probability of tossing 3 heads in a row is ⅛, or 12.5%. Now, what if we were given *new information, or new evidence* that the first coin tossed was a head? How would this impact our probability?
- *Answer:* We are now dealing with a *conditional probability* because we are taking into account other conditions (evidence or information). In this example, knowing that the first coin toss was a head is our new information. Now, instead of there being 8 combinations of coin tosses there are only 4. In this example, *T* stands for Tail and *H* stands for Head.

- Possibilities: HHH, HHT,HTT,HTH
- All outcomes are equally possible, so the probability of tossing 3 heads in a row is ¼, or 25%.

In Bayes' formula there are two conditional probabilities: P(A|B) and P(B|A).

- P(A|B) - This is the posterior probability. It is what we are searching for.

Classic Coin Toss Example Con't:

- Let A = Three heads in a row.
- Let B = First toss is a head.

How this fits together: We are wanting to know the probability of tossing three heads in a row *given that* the first toss was a head. In this example, *B* is the *new evidence* that alters or changes the probability of A.

- P(B|A) - This is called the *Likelihood* and is often confused with P(A|B). In real life, this is often the number we focus on and draw our conclusions from - but it is only one piece of the puzzle! (This error is called the Base Rate Fallacy). For example, say you have the flu and you research your symptoms. We'll label flu P(A) and symptoms P(B). You Google for some answers and find that 90% of people with the flu have your symptoms. This is P(B|A). Most people stop here and believe that they have the flu - but Bayes' Theorem tells us to take this a step further to find P(A|B).

Classic Coin Toss Example:

- Let A = Three heads in a row.
- Let B = First toss is a head.

How this fits together: We are wanting to know the probability of the first toss being a head *given that* three heads were tossed in a row. The answer of course is 1 or 100%.In this example, *B* is the *new evidence* that alters or changes the probability of A.

4. Posterior Probability

Also sometimes called *posterior odds,* the posterior probability refers to our changed belief *after* using Bayes' Theorem. It is what we initially define and work towards discovering. Technically, it is the probability of event A being true *given that* event B is true, which is expressed in the formula P(A|B). In addition, it is a conditional probability and a normalized weighted average.

5. Prior Probability

Sometime called *Prior Odds* or *Prior,* this refers to an initial belief *before* new evidence is introduced. In Bayes' Theorem:

- Both P(B) and P(A) are priors.
- The entire formula is divided by P(B) to *normalize* the answer. *Normalizing* can have a lot of different meanings depending on its usage. For simplicity in this case, it can be defined as summing an answer to 1.

6. Events

An *event* means one (or more) outcomes.

- Example:

- Tossing a coin and getting a head.
- Taking a cancer test and receiving a positive result.

There are three types of events:

- *Independent*: These events are not affected by other events.
- *Dependent*: These events can be affected by previous events.
- *Mutually exclusive*: Mutually exclusive events cannot happen at the same time. For example, when tossing a coin you can either have a head or tail, but not both at the same time.

Notation

There are many notations in Statistics, and Probability is no exception! Without going into too much detail, some of the basic notations you will see when using Bayes' Theorem include:

- P(A) = the probability that event A will occur.
- P(B) = the probability that event B will occur.
- P(A|B) = the conditional probability that event A is true *given that* event B is true.
- P(A') = the probability of the complement of event A.

- The *complement* of an event refers to the event not happening. For example, if an event is that it rains, the complement of that event is that it *does not rain*.

- P(A ∩ B) = the probability of the intersection of events A and B. In other words, the probability of events A and B both occurring.

- An *intersection* is a set of elements that belong to two sets. For example, if Set A has the elements {10,15,25} and Set B has the elements {5,10,15}, the intersection is {10,15} . In this example, 10 and 15 are the two elements that are common to both Set A and Set B. Visually, *intersections* are best demonstrated by using a Venn diagram. In a Venn diagram, the intersection is where both sets occur, or overlap. See Circle # 3 in any of the scenarios in Section 1 for an example.

Proof

At its most basic level, a proof is an argument that convinces other people that something is true. It proves *beyond any doubt* that something is true and is a written, step-by-step account of how a conclusion is reached. In mathematics, a proof is a deductive argument used to prove a statement. Maths.org provides a great overview of proofs.

There are simple and complex ways to demonstrate Proof of Bayes' Theorem. We have opted to show a simple method. If you are wanting a more thorough understanding of the proof, the University of Edinburgh offers a great explanation. This PDF from the University of Pennsylvania is also a fantastic overview.

1. First, let's define what we are proving. We are wanting to know the probability of A *given that* B has occurred.

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

2. Now, for our second step let us define what we already know. We know two things:

1. The first thing we know is that the following is true, and that the right-hand side of this statement can be read as: The probability of A, P(A) times the probability of B *given that* A has occurred P(B|A).

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

2. We also know that the following is true, and that the right-hand side of this statement can be read as: The probability of B, P(B) times the probability of A *given that* B has occurred P(A|B).

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

3. For our third step we equate the right hand side of both statements above.

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

4. Now, using basic algebra we can rearrange the formula to find Bayes' Theorem.

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

And that is Proof of Bayes' Theorem.

Recommended In-Depth Reading

If you want to know more about Bayes' Theorem there are a number of fantastic books that we would recommend. All the descriptions are taken from Amazon, and we provide links for you to click through to the Amazon website. Here we go, in no particular order.

The Theory That Would Not Die: How Bayes'' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy

By Sharon Bertsch McGrayne. Available on Kindle, Paperback, Hardcover, Audio.

Buy it on Amazon.

Amazon Description: *Bayes'' rule appears to be a straightforward, one-line theorem: by updating our initial beliefs with objective new information, we get a new and improved belief. To its adherents, it is an elegant statement about learning from experience. To its opponents, it is subjectivity run amok.In the first-ever account of Bayes'' rule for general readers, Sharon Bertsch McGrayne explores this controversial theorem and the human obsessions surrounding it. She traces its discovery by an amateur mathematician in the 1740s through its development into roughly its modern form by French scientist Pierre Simon Laplace.*

She reveals why respected statisticians rendered it professionally taboo for 150 years at the same time that practitioners relied on it to solve crises involving great uncertainty and scanty information, even breaking Germany's Enigma code during World War II, and explains how the advent of off-the-shelf computer technology in the 1980s proved to be a game-changer.

Today, Bayes'' rule is used everywhere from DNA de-coding to Homeland Security.Drawing on primary source material and interviews with *statisticians and other scientists, The Theory That Would Not Die is the riveting account of how a seemingly simple theorem ignited one of the greatest controversies of all time.*

Bayes'' Rule: A Tutorial Introduction to Bayesian Analysis

By Allen B. Downey. Available in Hardcover or Paperback Only

Buy it On Amazon

Amazon Description: *Discovered by an 18th century mathematician and preacher, Bayes'' rule is a cornerstone of modern probability theory. In this richly illustrated book, a range of accessible examples is used to show how Bayes'' rule is actually a natural consequence of common sense reasoning. Bayes'' rule is then derived using intuitive graphical representations of probability, and Bayesian analysis is applied to parameter estimation. As an aid to understanding, online computer code (in MatLab, Python and R) reproduces key numerical results and diagrams.*
The tutorial style of writing, combined with a comprehensive glossary, makes this an ideal primer for novices who wish to become familiar with the basic principles of Bayesian analysis.

The Signal and the Noise: Why So Many Predictions Fail - but Some Don't

By Nate Silver. Available on Kindle, Paperback, Hardcover, Audio.

Buy it on Amazon

Amazon Description: *Nate Silver built an innovative system for predicting baseball performance, predicted the 2008 election within a hair's breadth, and became a national sensation as a blogger—all by the time he was thirty. He solidified his standing as the nation's foremost political forecaster with his near perfect prediction of the 2012 election. Silver is the founder and editor in chief of FiveThirtyEight.com.*

Drawing on his own groundbreaking work, Silver examines the world of prediction, investigating how we can distinguish a true signal from a universe of noisy data. Most predictions fail, often at great cost to society, because most of us have a poor understanding of probability and uncertainty. Both experts and laypeople mistake more confident predictions for more accurate ones. But overconfidence is often the reason for failure. If our appreciation of uncertainty improves, our predictions can get better too. This is the "prediction paradox": The more humility we have about our ability to make predictions, the more successful we can be in planning for the future.

In keeping with his own aim to seek truth from data, Silver visits the most successful forecasters in a range of areas, from hurricanes to baseball, from the poker table to the stock market, from Capitol Hill to the NBA. He explains and evaluates how these forecasters think and what bonds they share. What lies behind their success? Are they good—or just lucky? What patterns have they unraveled? And are their forecasts really right? He explores unanticipated commonalities and exposes unexpected juxtapositions. And sometimes, it is not so much how good a prediction is in an absolute sense that matters but how good it is relative to the competition. In other cases, prediction is still a very rudimentary—and dangerous—science.

Silver observes that the most accurate forecasters tend to have a superior command of probability, and they tend to be both humble and hardworking. They distinguish the predictable from the unpredictable, and they notice a thousand little details that lead them closer to the truth. Because of their appreciation of probability, they can distinguish the signal from the noise.

Bibliography

Visual Introduction Part 1 & 2

Acts17Apologetics. "Understanding Bayes Theorem." *YouTube*. YouTube, 03 July 2014. Web. 06 Sept. 2016. Link

The Bayes' Theorem Formula: A Basic Overview

Wireless Philosophy. "CRITICAL THINKING - Fundamentals: Bayes Theorem [HD]." *YouTube*. YouTube, 22 Apr. 2016. Web. 06 Sept. 2016. Link

McRaney, David. "YANSS 073 – How to Get the Most out of Realizing You Are Wrong by Using Bayes Theorem to Update Your Beliefs." *You Are Not So Smart*. You Are Not So Smart, 22 Apr. 2016. Web. 06 Sept. 2016. Link

Kurt, Will. "A Guide to Bayesian Statistics." *Count Bayesie*. Will Kurt, 02 May 2016. Web. 06 Sept. 2016. Link

"Bayes Rule: Guide." *Arbital*. N.p., n.d. Web. 06 Sept. 2016. Link

Section 1

Wireless Philosophy. "CRITICAL THINKING - Fundamentals: Bayes Theorem [HD]." *YouTube*. YouTube, 22 Apr. 2016. Web. 06 Sept. 2016. Link

Bonilla, Oscar. "Visualizing Bayes Theorem." *Oscar Bonilla*. N.p., 01 May 2009. Web. 06 Sept. 2016. Link

Kurt, Will. "Use Bayes Theorem to Investigate Food Allergies." *Count Bayesie*. N.p., 22 Jan. 2016. Web. 06 Sept. 2016. Link

McRaney, David. "YANSS 073 – How to Get the Most out of Realizing You Are Wrong by Using Bayes Theorem to Update Your Beliefs." *You Are*

Not So Smart. You Are Not So Smart, 22 Apr. 2016. Web. 06 Sept. 2016. Link

Azad, Kalid. "An Intuitive (and Short) Explanation of Bayes Theorem."*BetterExplained*. N.p., n.d. Web. 06 Sept. 2016. Link

Base Rate Fallacy: Definition & Example." *Study.com*. Study.com, n.d. Web. 08 Sept. 2016. Link

Section 2

Wireless Philosophy. "CRITICAL THINKING - Fundamentals: Bayes Theorem [HD]." *YouTube*. YouTube, 22 Apr. 2016. Web. 06 Sept. 2016. Link

PatrickJMT. "Bayes Theorem and Cancer Screening." *YouTube*. YouTube, 03 Jan. 2013. Web. 06 Sept. 2016. Link

Ingersoll, George. "Stats 10 Bayes Theorem." *YouTube*. YouTube, 26 May 2014. Web. 06 Sept. 2016. Link

McRaney, David. "YANSS 073 – How to Get the Most out of Realizing You Are Wrong by Using Bayes Theorem to Update Your Beliefs." *You Are Not So Smart*. You Are Not So Smart, 22 Apr. 2016. Web. 06 Sept. 2016. Link

Section 3

Wireless Philosophy. "CRITICAL THINKING - Fundamentals: Bayes Theorem [HD]." *YouTube*. YouTube, 22 Apr. 2016. Web. 06 Sept. 2016. Link

Bonilla, Oscar. "Visualizing Bayes Theorem." *Oscar Bonilla*. N.p., 01 May 2009. Web. 06 Sept. 2016. Link

Section 4

"Thomas Bayes." *Wikipedia*. Wikimedia Foundation, n.d. Web. 06 Sept. 2016. Link

O'Connor, J.J, and E.F Robertson. "Thomas Bayes'." *Bayes Biography*. University of St. Andrews, Scotland, n.d. Web. 06 Sept. 2016. Link

McGrayne, Sharon Bertsch. *The Theory That Would Not Die: How Bayes Rule Cracked the Enigma Code, Hunted down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy*. New Haven: Yale UP, 2011. Print.

Tough, Paul. "A Speck in the Sea." *The New York Times*. The New York Times, 04 Jan. 2014. Web. 06 Sept. 2016. Link

Flam, F.D. "The Odds, Continually Updated." *The New York Times*. The New York Times, 29 Sept. 2014. Web. 06 Sept. 2016. Link

"Bayesian Search Theory." *Wikipedia*. Wikimedia Foundation, n.d. Web. 06 Sept. 2016. Link

"How to Apply Bayes Theorem to the Search for a Fisherman Lost at Sea."*Stack Exchange*. N.p., n.d. Web. 06 Sept. 2016. Link

"USS Scorpion (SSN-589)." *Wikipedia*. Wikimedia Foundation, n.d. Web. 06 Sept. 2016. Link

"Air France Flight 447." *Wikipedia*. Wikimedia Foundation, n.d. Web. 06 Sept. 2016. Link

"Naive Bayes Spam Filtering." *Wikipedia*. Wikimedia Foundation, n.d. Web. 06 Sept. 2016. Link

A Bayesian Approach to Filtering Junk E-mail." *Bayesian Spam Filter:*. Microsoft, n.d. Web. 06 Sept. 2016. Link

Lafrance, Adrienne. "The Hazard of Tesla's Approach to Driverless Cars." *The Atlantic*. Atlantic Media Company, July 2016. Web. 06 Sept. 2016. Link

"Google Self-Driving Car Project." *Google Self-Driving Car Project*. Google, n.d. Web. 06 Sept. 2016. Link

Rogers, Christina. "Ford Developing Fully Driverless Car." *WSJ*. Wsj.com, 16 Aug. 2016. Web. 06 Sept. 2016. Link

Chafkin, Max. "Uber's First Self-Driving Fleet Arrives in Pittsburgh This Month."*Bloomberg.com*. Bloomberg, 18 Aug. 2016. Web. 06 Sept. 2016. Link

"Self Driving Car." *CS221*. Stanford University, n.d. Web. 06 Sept. 2016. Link

Section 5

Kiersz, Andy. "DATING FOR BAYESIANS: Here's How To Use Statistics To Improve Your Love Life." *Business Insider*. Business Insider, Inc, 05 Nov. 2013. Web. 06 Sept. 2016. Link

Galef, Julia. "A Visual Guide to Bayesian Thinking." *YouTube*. YouTube, 16 July 2015. Web. 06 Sept. 2016. Link

Section 6

Freiberger, Marianne. "Why We Want Proof." *Plus.maths.org*. N.p., 10 Apr. 2015. Web. 06 Sept. 2016. Link

"The Definition of Probability." *Dictionary.com*. Dictionary.com, n.d. Web. 06 Sept. 2016. Link

Table of Contents

Table of Contents