



Introduction To Probabilistic Thinking

Madhav Kaushish

Member, ThinQ (www.schoolofthinq.com)

Email: madhav.kaushish@gmail.com

Phone: +91 99100 99677

This is an introduction to Probabilistic Thinking and not Probability Theory - it is not meant to present a rigorous, axiomatic theory, but a way of thinking about the world. The goal is to get students to be competent probabilistic thinkers able to deal with uncertainty. This introduction includes simple probabilities, probabilistic logic, distributions, bayesian reasoning, and bayesian updating.

Introduction

Was there a time when Dragons existed? Will aliens visit Earth? You might answer both of these questions with flat out 'no.' However, are you certain that you are correct? No, you cannot be. Aliens (organisms from planets outside the Earth) may exist and may visit us. Questions about the real world have to be answered with some level of skepticism. You may dismiss these questions. However, what about the following:

1. Will the BJP win the elections in Uttar Pradesh?
2. Will it rain tomorrow?
3. When did humans invent the wheel?
4. Are you your father's child?
5. Will my cousin give birth on the due date her doctor gave her?

As of late 2015, the BJP looks strong in UP. They won almost all the seats in the general elections. You might think the BJP will win given this evidence. However, what if SP and BSP combine with the Congress? Will that change your estimate? What if the economy crashes and there are huge job losses throughout the country? These are certainly not impossible scenarios to imagine. Hence, you cannot give the BJP a 100% chance of winning. You need to give a probabilistic estimate.

Turning to the question of whether you are your father's biological child (Notice the this is a question about the past while the previous was a question about the future). Lets remove the emotional aspect from this question, and lets deal with a person 'X.' Lets assume we know that X was not adopted and not a surrogate child or anything similar. Also, X's parents were married before X was born, and the father believes that he is X's biological father. Is X's father actually his biological father or did X's mother have an affair which resulted in X? This is an extremely complicated question which we can approach in various ways. However, it is a question to which, without DNA testing, we can only give a probabilistic answer. In fact, even with DNA testing we can only give a probabilistic answer.

Taking up the wheel question, this is even more complicated, since different time periods come with different probability estimates. We may put a high probability estimate on the time period 6500BC to 4500BC, which Wikipedia says has the earliest discovered wheel. However, maybe

earlier wheels have not been discovered. Also, what probability estimate would you put on the time period 6000BC to 5000BC?

I recently found out that my cousin sister is pregnant. The doctor has given her a due date of April 23rd. Due dates are calculated by adding 40 weeks to the date of conception. Does this mean that her child will be born on the 23rd of April? If I have to leave the city sound that time, will it be safe for me to book my tickets for the 24th of April so that I get to meet my new niece/nephew before leaving? Clearly, when a doctor says 23rd of April, there is some doubt as to the exact date. But, how much doubt? Would it be very surprising if the baby comes two days early, or one week late? According to various studies, only around 4% of women give birth on their due date. To ensure a 50% chance of witnessing a particular pregnancy, you need a 15 day window around the due date.

Did person X murder person Y? Does Z have cancer? Should I invest in the stock market? Should A get married to B? Should C buy a house or rent? Should D quit her job? These are all questions which are important to us, and all of them involve probabilities. Humans are notoriously bad with probabilities as evidenced by the huge attraction of lotteries and casinos (when I say casino's I mean games like roulette and not games like poker, which involve skill).

The question about rain is one in which we see probabilities being given to us on a daily basis. What does it mean when the weather report says 'there is a 70% chance of rain today.' In that case, if it doesn't rain, is the forecaster wrong? Not necessarily. There are at least two ways to look at it. One way is, assume you were a god who could replay today 10,000 times without changing anything important each time. In that case, what the weather forecaster is saying is that approximately 7,000 of these replays will have rain while 3,000 will not. Another way of interpreting it is that the weather forecaster will be willing to take a bet at 3:7 odds that it will rain, and at 7:3 odds that it will not. x:y odds means that for every y Rupees you bet, you expect to make a profit of x Rupees if you win the bet. So, the forecaster is willing to bet Rs.70 that it will rain in order to take Rs.100 if it rains (a profit of Rs.30). Equally, he is willing to bet Rs.30 that it will not rain in order to make Rs.100 if it does not rain (a profit of Rs.70).

Weather forecasting is an extremely hard example of probabilistic thinking. In order to understand probabilities, we have to start with much easier scenarios.

A Murder

Say you are a defense lawyer. You do not know whether your client is guilty of murdering her uncle. If you plead not guilty and the defendant is found guilty, she could face a death sentence or unto 30 years in prison (depending on the judgement of the judge). If you plead guilty, you will probably be able to get a deal by which your client will get 15 years in jail, reducible to 10 years for good behavior. You want to do what is best for your client. What call do you make?

Without any more information, it is very hard to make a call. Even 10 years in prison for an innocent 21 year old would be terrible. It would ruin her life. Then again, if she were guilty, 10 years would almost be a let-off.

You have a meeting with your client. She says she is innocent, but cannot remember anything specific from the night we are concerned with.

You want to get to know what your client is like as a person. You talk to her about her life and what she is interested in and so on. She **seems like a regular, nice person**. She **doesn't seem at all like the murders you have dealt with** in your 30 years of experience as a defense lawyer. You come to a judgement that **she is probably innocent**, at least of pre-meditated murder. Your judgement is based on your experience that **people like your client are very unlikely to have committed a murder**. That does not mean that you are certain that your client is innocent.

What is the evidence the prosecution has against your client? The first things you encounter in your search are the following:

1. Your client was seen that evening at her uncle's house by the neighbor.
2. The knife used to stab her uncle in the heart, which was still inside of her uncle when the police found him, was a flick knife your client had been gifted on her birthday to chisel wood with.

Both of these pieces of evidence can be countered. It is **not unusual** for somebody to visit their uncle. Also, if she was visiting her uncle, it is not completely crazy that the person who actually killed him took the knife off her and stabbed him with it. Even so, seeing this evidence **decreases your faith** in the innocence of your client. To see why it must decrease your faith in

your client's innocence, **consider the possibility** where the police were unable to find any evidence of your client being at her uncle's house. That would surely **increase your belief** in your client's innocence. So, even if it has a very slight impact, there ought to be some decrease in your **degree of confidence**.

Your next move is to talk to your client's doctor about her loss of memory. The doctor, an expert on amnesia, says that there are two **possible causes**:

1. Since the night in question was a few months ago, it is quite possible that your client just forgot a night which was not memorable.
2. The memory loss is a result of some trauma she faced that night

You try thinking back to nights 3 months ago. Most of them are just a blur. You only remember a few memorable occasions like your birthday. However, if that is true in the case of your client, how do you explain the neighbor's eyewitness statement and the knife?

You go to investigate the neighbor. She is a nice old lady who invites you in for tea. You ask her if she could have been wrong about seeing your client that night. She said that she was **quite certain** it was your client. You ask her where she was standing when she saw your client, and you decide to run an experiment. You tell her to keep looking into her neighbor's house and note the time that she sees you entering. Rather than going yourself, you send two of your employees. The first one looks nothing like you, but the second bears a resemblance. When you go back to the old lady, you see that she was fooled by your employee who looked a little like you, but not by the other one. So, it is **possible** that the old lady was wrong about your client being there that night. However, it is **most likely** that if she were fooled, she was fooled by somebody similar looking to your client.

What about the knife? You have a conversation with your client's employer, who says that your client was a forgetful person who kept leaving things all over the place. So, if she had gone to her uncle's house even a few days before the night in question, it is **possible that she left the knife** then. Since you have found holes in the evidence which made you **increase your belief** in your client's guilt, considering this possibility will **slightly decrease this degree of belief**.

So, you decide to construct three alternative narratives based on the two possible explanations suggested by the doctor, and one of your own.

Narrative 1 - The night was not memorable. If this is true, it is almost certain that your client did not kill her uncle. It is **highly unlikely** that murdering somebody isn't memorable. The implication of this narrative is that somebody else came to the uncle's house resembling your client. She found the knife and used it to kill the uncle.

Narrative 2 - Stress induced amnesia. If this is true, your client could have killed her uncle. However, she also may have just witnessed her uncle's murder. Both could result in these symptoms. If this is true, to prove your client's innocence, you would have to establish the presence of somebody else at the scene. If only your client were present, that would provide **enough evidence** for the court to convict her. Though, if Narrative 2 is true and your client is guilty, you may be able to show that the murder was not pre-meditated.

Narrative 3 - Your client is lying. She remembers that night perfectly well, but is just not coming forward with the story, either to protect herself or to help somebody else. If this can be established, your client is **probably guilty**.

The best possible scenario for you is to establish Narrative 1. If the prosecution can rule out Narrative 1, you need to start looking more into Narrative 2. Narrative 3 looks the worst for you.

Lets pause here and go through some of the things which have come up so far. You started off with the belief that your client is probably innocent given what you have observed about murderers in the past. **As evidence kept coming in, your degree of belief in your client's innocence kept going up and down.** This degree of belief/certainty can also be thought of as the probability of your client having murdered her uncle.

In the next few sections, we will go over the basic principals of probability.

Simple Probabilities

Suppose you have a sack of blue balls. You reach into the sack and pull out a ball. Is the ball blue or not? Well, assuming you didn't mix any other colored balls by mistake, we can be certain that the ball is blue. In other words, there is a 100% chance that the ball is blue.

Now, suppose a friend of yours adds some red balls to the sack. Now, when you pull out a ball, what color is it? Blue or not? Well, now you cannot be certain. Maybe you will pull out a ball your friend has added and it will be red. Or maybe you will pull out one of the original balls and it will be blue. However, you can be certain (once again assuming your friend did only add in red balls) that the ball will not be green. So, what is the chance that the ball will be either red or blue? Well, 100%. What about: what is the chance that the ball will be either red, blue or green? Still 100% since red and blue cover all the balls - green is just redundant.

However, what is the chance that the ball you pull out is actually blue? Well, that depends on how many blue balls you had originally had, and how many red balls your friend added. Say you had 999 blue balls and your friend added 1 red ball. Is the ball you pull out blue or red? Well, you can be quite sure it will be blue. However, you cannot had the complete certainty you had earlier. Now, what if you had 600 blue balls and you friend added in 400 red balls. Even now, you would say that it is more likely that the ball you pull out is blue. However, your certainty is reduced. What if there were 500 red and 500 blue balls? Now its a toss up. You can't really say.

Does the absolute number of red and blue balls matter or is it something else. What if the sack had 5000 blue balls and 5000 red balls? Would that be any different from 500 and 500 blue and red balls respectively? No it wouldn't. What seems to matter is the proportion of blue and red balls.

If there were 500 blue and 500 red balls, you would say that there is a 50% chance that the ball you pull out will be blue. By convention, when dealing with probabilities, rather than using percentages you tend to use portions of 1. So, 50% chance converts to a 0.5 probability. 100% chance is a 1 probability, while 0% chance is a 0 probability. Though this may be by convention, you will see that it is a useful convention when things get more complicated.

So, what does the case with 600 blue balls and 400 red balls translate to in terms of probabilities. There are a total of 1000 balls. 600 are blue. Hence we say there is a $600/1000$ probability that the ball is blue, which cancels down to 0.6.

What if I said there is a 0.6 probability of pulling out a blue ball and a 0.6 probability of pulling out a red ball from the same sack? That doesn't make sense, since they add up to more than one. In percentages that translates to over 100% chance of an event, which you would agree seems meaningless. (This is assuming that balls cannot be blue and red at the same time. We will deal with this case later) Lets write down a rule of probability (many of the rules written down will probably change as we go along):

Rule 1: The sum of probabilities of different events cannot be more than 1.

Now, what if I said that the probability of pulling out a blue ball is 0.5 and the probability of pulling out a red ball is 0.4, and there are no other colored balls. That means that assuming there were a thousand balls in the sack, 500 were blue and 400 were red. So, then what happens to the remaining 100? This scenario seems impossible. So, lets write down another rule:

Rule 2: The sum of probabilities of all possible events has to be 1.

Assume you were given a new sack with red, blue, green and yellow balls, and no others. You are told that there are 1000 balls in total, and there are 300 blue balls. So, the probability of the balls being blue is 0.3. So, what is the probability that the ball is not blue? Well it is equal to the probability that the ball is red, yellow or green. We know that since there are 300 blue balls, there must be 700 balls of those colors. So, the probability that the ball is not blue is 0.7. Another way of thinking about it is that since the probability of a ball being blue is 0.3, and (by Rule 2), the probabilities of all possible events have to sum up to 1, the probability of the ball being not blue is $1 - 0.3 = 0.7$. This leads us to another rule of probability:

Rule 3: If the probability of a particular event is x , then the probability of that event not happening is $1 - x$.

Given the same sack with the 1000 balls of 4 colors, assume you are told there are 300 blue balls and 300 red balls. What is the probability that the ball you pick out is either blue or red? Well, there are 600 balls which are either blue or red. So, the probability is 0.6. Notice that the

probability of blue is 0.3 and the probability of red is 0.3. What we have done is to just add them together. This leads to:

Rule 4: If the probability of an event happening is x and the probability of a different event happening is y , then the probability of either of the events happening is $x + y$.

Lets take a scenario with two sacks, each with 2 balls, one red and one blue. For each of them, you will agree that the probability of a ball you pull out being blue is 0.5. So, what is the probability that if you pull out a ball each from both the sacks, both the balls will be blue? For sure the probability has to be less than 0.5 - the probability of two of the same event happening has to be less than the probability of just one of them. The possible combinations you would get are: blue-blue, blue-red, red-blue, and red-red. In this case, these appear with equal probability, so the probability of blue-blue (which is what we are looking for) must be 0.25. Another way to think about it is: take a ball out of the first bag. There is a 0.5 probability that it is blue. If it is red, it doesn't matter what the next ball is. If it is blue, there is a 0.5 probability that the next ball is blue. So, the probability of both the events happening must be the product of the probabilities, which is $0.5 \cdot 0.5 = 0.25$. Now for another rule:

Rule 5: If the probability of an event happening is x and the probability of a different event happening is y , then the probability of both of the events happening is $x \cdot y$.

Dependent and Mutually Exclusive Events

Lets re-look at the last rule from the previous section, but taking a different scenario. Given one sack of 5 blue balls and 5 red balls, what is the probability that the first ball you pull out is blue and the second you pull out is red. There are two ways you could do this:

1. You take out the first ball, put it back, and then pull out the second.
2. You take out the first ball, and leaving it out of the sack, pull out the second ball

In the first case, the scenario is the same as the one in the previous section, and the probability will be 0.25. The second one is a little more problematic. In the second one, after you pull out the first ball, there are only 9 balls left, and not 10. So, what is the probability that the first ball is blue? That remains at 0.5. Now, the situation when the first ball is not blue is not relevant. So, we have to see what happens in the situation when the ball is blue. In that situation, there are 9 balls left 5 of which are red. What is the probability that the second ball is red? It is $5/9$. So, the probability that the first ball is blue and the second is red is $0.5 \cdot 5/9 = 5/18$. We have to revise Rule 5 to say:

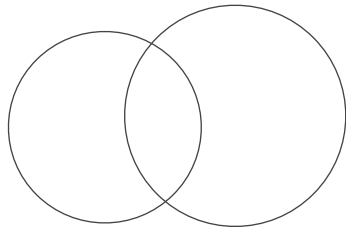
Rule 5': If the probability of an event happening is x , and when x has happened successfully, the probability of a different event happening is y , then the probability of both of the events happening is $x \cdot y$.

However, in 1 above how come we didn't need this caveat? That is because whether the first event happened or not did not effect the second event at all. They were '**independent events.**' We call the events in the second scenario '**dependent events.**'

Since we seem to be in the business of changing rules, lets now take a look at Rule 4. Say, somebody told you that there is a 0.8 chance of Putin winning the next Russian elections, and a 0.5 chance of Modi winning the next Indian elections. Lets assume for now that we trust the probabilities. By Rule 4, the probability of either Modi winning or Putin winning would be 1.3, which is impossible. These events seem like they are largely independent, so dependency is not the issue here. In order to see what is happening, lets unpack the question being asked.

What are the possible ways in which either Modi and Putin could win. Well, Modi could win and Putin could lose. Modi could lose and Putin could win, or Modi and Putin could both win. We call this complete description of all possibilities of an event ‘**the space of possibilities.**’

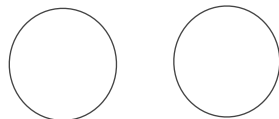
What is the probability that Modi wins and Putin loses. By Rule 3, the probability that Putin loses is 0.2. Since we are assuming these to be independent events, by Rule 4, the probability that Modi wins and Putin loses is $0.5 \cdot 0.2 = 0.1$. Similarly, the probability that Putin wins and Modi loses is $0.8 \cdot 0.5 = 0.4$. The probability that Modi and Putin both win is $0.5 \cdot 0.8 = 0.4$. These three events seem to be more similar to the events mentioned before rule 4 in the previous section. Maybe if we add them up to get 0.9, that would be correct. Lets look at this a different way. When we are adding the probability of Modi winning and Putin winning, we are double-counting the situation when both win. See this graphically:



The left circle represents the times when Modi wins, and the right represents the times when Putin wins. What we are doing when we are adding these up blindly is that we are double counting the space shared by them. So, lets revise rule 4:

Rule 4’: If the probability of an event happening is x and the probability of a different event happening is y , then the probability of either of the events happening is $x + y -$ (the probability of both the events happening together).

In the scenario in the previous section, there was no way that a ball could be blue and red at the same time (by stipulation). We call such events **Mutually Exclusive**. The left circle below is when the ball is blue and the right is when the ball is red. There is nothing shared.



The Modi-Putin event is **non-Mutually Exclusive**. It is very important to be careful in such situations. In the case of Modi-Putin above, it was easy to see that there was a problem since it

summed up to more than 1. However, say the probabilities were changed to 0.2 for Modi and 0.5 for Putin. The same problems would remain, but would be harder to spot.

Some Notation

The sentences we have been writing for the rules are quite cumbersome, and could get quite convoluted as things get more complicated. It is time to replace words with formal notation.

Let $P(A)$ be equivalent to saying: The probability of event A occurring. So, $P(\text{Picked Ball Being Blue})$ is the same as probability of the picked ball being blue.

$P(A|B)$ is probability of A given B occurred successfully - the probability that A happened assuming B actually happened.

$P(A \cup B)$ is the same as probability of either A or B occurring. $P(A \cap B)$ is probability of A and B both occurring.

$P(\sim A)$ is the same as probability of A not occurring.

The revised rules are as follows:

Rule 1F: If A_1, A_2, \dots, A_n are possible events, then $P(A_1) + P(A_2) + \dots + P(A_n)$ cannot be greater than 1

Rule 2F: If A_1, A_2, \dots, A_n are mutually exclusive events which make up the space of possibilities, then $P(A_1) + P(A_2) + \dots + P(A_n) = 1$

Rule 3F: $P(\sim A) = 1 - P(A)$

Rule 4F: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Rule 5F: $P(A \cap B) = P(A) \cdot P(B|A)$

This notation might look complicated, but you will get used to it. Also, as you can see, it saves a lot of writing!

Distributions

When you toss a fair coin 100 times, you would expect around 50 of the tosses to result in heads. However, would you expect exactly 50 tosses to result in heads? Would you be surprised if there were only 49 heads or if there were 51? Probably not, but you probably would be surprised if there were 95 heads and only 5 tails.

If we are right and our surprise levels match with low probability, then why is it that 95 heads is surprising but 49 heads is not?

In order to answer this question, let's move to a simpler example, with just 5 tosses. The translation of our surprise levels would be:

- 0 heads: Quite Surprised
- 1 head: Slightly Surprised
- 2 heads: Not too surprising
- 3 heads: Not too surprising
- 4 heads: Slightly Surprised
- 5 heads: Quite surprised

Now, the question we have to ask is: how many ways would we have arrived at a particular configuration.

For 0 heads, there is only one way - All tosses resulted in tails. For 1 head, there are 5 ways this could have happened - Any one of the five could have been heads, with all the others being tails. The following table lists the possibilities

Toss 1	Toss 2	Toss 3	Toss 4	Toss 5
H	T	T	T	T
T	H	T	T	T
T	T	H	T	T
T	T	T	H	T
T	T	T	T	H

For 2 heads, the possibilities increase substantially. It could be that the first and the last toss were heads, or the first and the second or it could have happened in many other ways. Lets list the possibilities:

Toss 1	Toss 2	Toss 3	Toss 4	Toss 5
H	H	T	T	T
H	T	H	T	T
H	T	T	H	T
H	T	T	T	H
T	H	H	T	T
T	H	T	H	T
T	H	T	T	H
T	T	H	H	T
T	T	H	T	H
T	T	T	H	H

There are 10 possibilities for 2 heads. What about for 3 heads:

Toss 1	Toss 2	Toss 3	Toss 4	Toss 5
H	H	H	T	T
H	H	T	H	T
H	H	T	T	H
H	T	H	H	T
H	T	H	T	H
H	T	T	H	H
T	H	H	H	T
T	H	H	T	H
T	H	T	H	H
T	T	H	H	H

Notice that 3 heads also has 10 possibilities, and we can create those possibilities by just replacing T with H and H with T in the 2 heads scenario. This is because T and H are equally likely to occur.

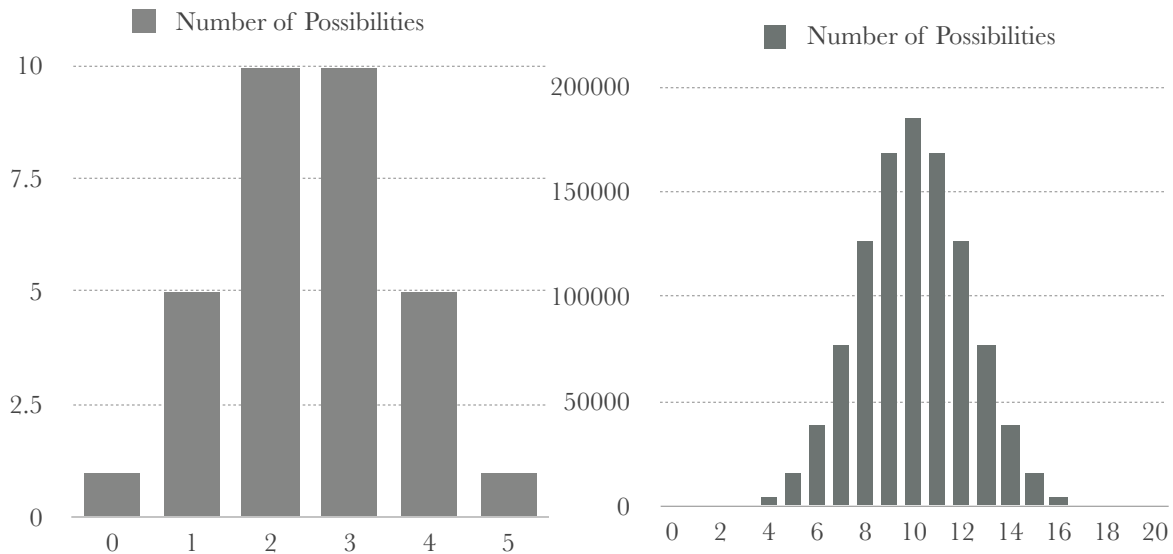
So, 4 heads will be the same as 1 head, and 5 heads will be the same as 0 heads. Lets put down what we have so far:

Number of Heads	Number of Possibilities
0	1
1	5
2	10
3	10
4	5
5	1

This is another example of a space of possibilities. There are a total of 32 possibilities. So, the probabilities of getting a particular number of heads are:

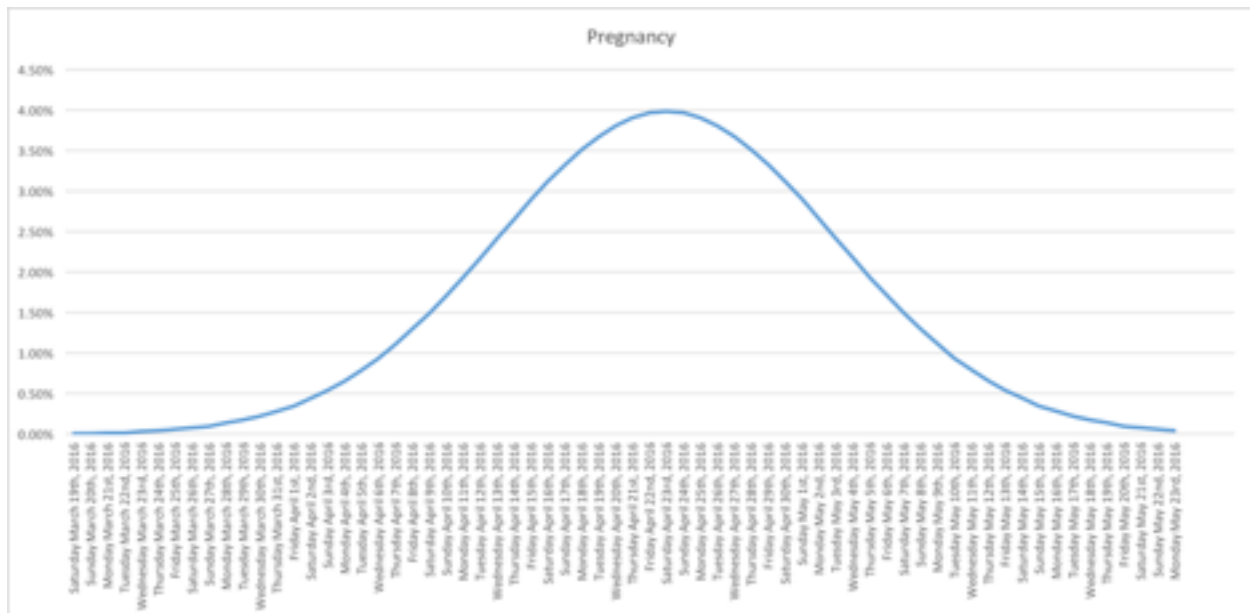
Number of Heads	Probability
0	$1/32$
1	$5/32$
2	$10/32$
3	$10/32$
4	$5/32$
5	$1/32$

All these in the above table are mutually exclusive events - you cannot get exactly 2 heads and exactly 4 heads in the same set of 5 tosses. So, if we were asking, ‘what is the probability that the number of heads is either exactly 2 or exactly 4,’ we can just add up the corresponding probabilities. If we plot this on a graph with the x-axis representing the number of heads, and the y-axis representing the frequency, we get the following graphs:



The one on the left is the graph with 5 coins, while the one on the right is when there are 20 coins. In the right one, you can clearly see that $P(12 \text{ Heads} \cup 13 \text{ Heads}) > P(10 \text{ Heads})$. In fact, there is less than 0.18 probability of getting exactly 10 heads, so getting exactly 10 heads is quite unlikely.

An interesting example of a distribution is of the probability of a particular person giving birth on a given day. The due date is calculated by the doctor as 40 weeks after conception. However, the probability of giving birth on that particular date is less than 0.04. Here is the distribution for somebody whose due date is the 23rd of April 2016:



Distributions which look like the ones above are called Normal distributions. They are symmetric graphs which increase till they hit a peak and then decrease. These graphs are also called Bell Curves. Such distributions are very useful in science since unbiased, random events like coin flips and dice rolls tend to follow this sort of distribution. If events do not follow such a distribution, that usually calls for an explanation.

For more on Distributions, see Basics of Statistics by Jarkko Isotalo. (<http://www.mv.helsinki.fi/home/jmisotal/BoS.pdf>)

Combining Probabilities

So far, we have seen very idealized examples. Such examples rarely occur in the real world. Most examples we encounter require us to combine various factors. If we wish to estimate who is going to win an next election, we have to put together various pieces of information such as poll data, historical precedent and so on.

Lets take some simple examples of combining probabilities.

Undergraduate Degrees

The following is data on Undergraduates in the UK published by the Guardian:

Students obtaining undergraduate degrees, 2011-12				
	Subject areas	Number of students obtaining degrees	% female	% male
1	Medicine & dentistry	10140	57.6	42.3
2	Subjects allied to medicine	67960	82.1	17.9
3	Biological sciences	42040	60.8	39.2
4	Veterinary science	900	79.4	20.6
5	Agriculture & related subjects	4490	63	37
6	Physical sciences	17975	42.6	57.4
7	Mathematical sciences	8895	42.2	57.8
8	Computer science	20060	17.4	82.6
9	Engineering & technology	30500	14.3	85.7
10	Architecture, building & planning	13895	29.7	70.3
11	Social studies	50355	64.5	35.5
12	Law	20440	61.7	38.3
13	Business & administrative studies	77280	51.4	48.6
14	Mass communications & documentation	13560	55.4	44.6
15	Languages	28705	68.9	31.1
16	Historical & philosophical studies	20825	54.4	45.6
17	Creative arts & design	49920	61.7	38.3
18	Education	38465	80.4	19.6
19	Combined	6710	60.1	39.9

It is easy to see that more females study education than males do, more males study engineering than females do, and that law has about an equal number of both genders. However, think about the following questions:

1. Somebody called Sonam (assume it is a name equally common amongst both males and females) emerges from a graduation ceremony jointly conducted by the Veterinary Science department and the Architecture, building and planning school. Is Sonam more likely to be male or female?
2. Somebody you overheard is either studying Business and administrative studies or mathematical sciences - you don't know which one. Given this information, is the person more likely to be male or female?

Taking the first question, the probability that Sonam gender is male or female in the individual schools is:

	Male	Female
Veterinary Science	0.206	0.794
Architecture, building and planning	0.703	0.297

There is no way that Sonam can study at both the places, so if we follow our earlier technique for dealing with mutually exclusive or, we can just add up the probability. However, that will result in a figure greater than 1 for the sum of probability of males and females. So, we will have to divide by 2 since we are adding up two different sets of events. That would result in:

$$P(\text{Male}) = 0.455$$

$$P(\text{Female}) = 0.545$$

So, what we get is that Sonam is more likely to be a girl. However, is that true? Can we just add up probabilities when we are dealing with two different situations? Notice, that when we divided by 2, one of the assumptions we were making was that the populations in the two departments was of the same size. However, that is clearly not the case as you can see in the data. This adding up would work if the two departments had the same size. Lets try to see what is actually happening by converting the percentages into numbers:

	Male	Female	Total
Veterinary Science	185	815	900
Architecture, building and planning	9768	4127	13895
Total	9953	4942	14895

After seeing the actual numbers rather than the percentages, we can clearly see that males outnumber females almost 2 to 1 in this population. Our initial estimate was wrong because we compared percentages without thinking about the relative size of the populations.

Now, addressing question 2, you can see that business and administrative studies has about the same number of males as it has females, with slightly more females than males. However, mathematical sciences has a larger percentage gap than business and administrative studies, favoring males. We should be worried about what happened in the first question, and make the conversion to numbers:

	Male	Female	Total
Business and Administrative Studies	37565	39715	77280
Mathematical Sciences	5140	3755	8895
Total	42705	43470	86175

Once again, we see that the percentage comparisons proved wrong and that the probability that the person is female is slightly more than the probability that the person is male.

Hit and Run¹

85% of cabs in a city are green and the remaining are blue. A person is run over by a cab at night. There is an eyewitness who says the cab was blue. To test the reliability of the eyewitnesses, tests are done under similar circumstances with blue and green cabs. The witness correctly identifies each one of the two colors correctly 80% of the time and wrongly 20% of the time. Assume the test is valid. Should the police continue the investigation with the assumption that the cab is blue or that it is green?

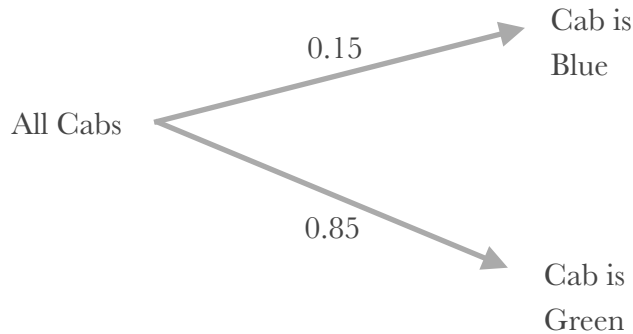
Basically, the question is asking whether the probability that the cab was blue is more or less than the probability that the cab was green.

If a blue car is put in from the eyewitness, it is identified correctly as blue 80% of the time. So, our first thought might be that the eyewitness was most likely correct, and the police should be looking for a blue car. However, out of all the cars in the city, 85% of them are green. So, if you had been asked the question without an eyewitness to tell you anything about the color of the

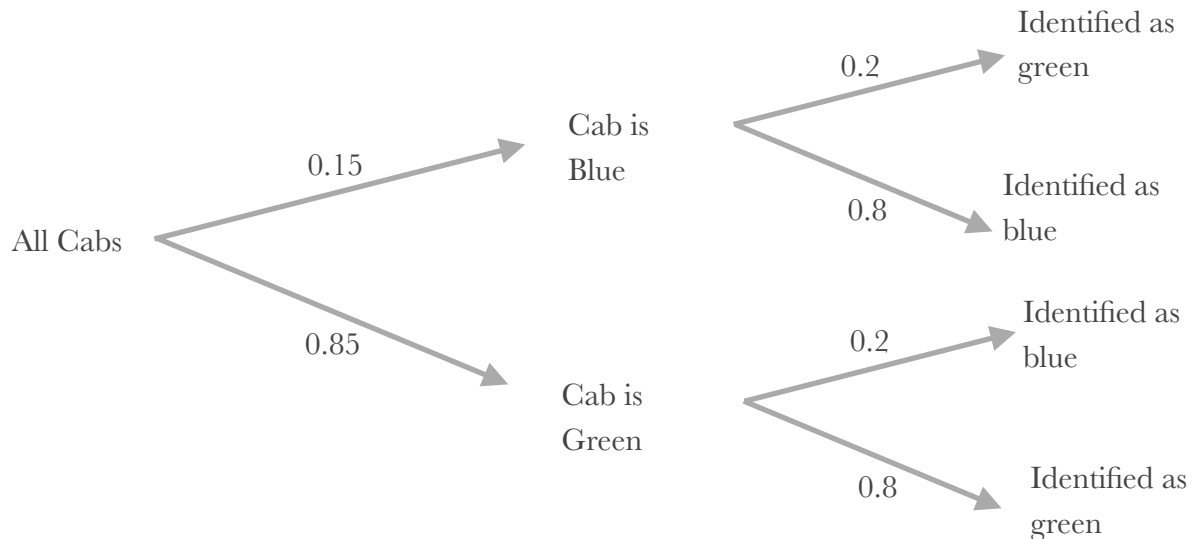
¹ This question is from Thinking Fast and Slow by Daniel Kahneman

cars, you would be right to have the suspicion that the car was most likely green. Does the percentage of green and blue cars in the population of cars have any impact. In order to see whether it does, let's attempt to unpack what is going on.

There are two possibilities for the color of the cab: it is either blue with a 0.15 probability or green with a 0.85 probability.



If the cab is blue, the eyewitness would have identified it as blue with a 0.8 probability and green with a 0.2 probability. If the cab was green, the eyewitness would have identified it as green with a 0.8 probability and blue with a 0.2 probability.



What we are interested in is situations where the cab is identified as blue, as that is the situation we find ourselves in. So, the following are the probabilities we are interested in:

$P(\text{Cab is blue} \cap \text{Cab is identified as blue})$

$P(\text{Cab is green} \cap \text{Cab is identified as blue})$

From rule 5F, $P(\text{Cab is blue} \cap \text{Cab is identified as blue}) = P(\text{Cab is blue}) \cdot P(\text{Cab is identified as blue} | \text{Cab is blue})$

$$P(\text{Cab is blue}) = 0.15$$

$P(\text{Cab is identified as blue} | \text{Cab is blue}) = 0.8$ which we can see from the tree drawn above

$$\text{So, } P(\text{Cab is blue} \cap \text{Cab is identified as blue}) = 0.8 \cdot 0.15 = 0.12$$

$$\text{Similarly, } P(\text{Cab is green} \cap \text{Cab is identified as blue}) = 0.85 \cdot 0.2 = 0.17$$

So, $P(\text{Cab is green} \cap \text{Cab is identified as blue}) > P(\text{Cab is blue} \cap \text{Cab is identified as blue})$

Hence, it is more likely that the cab in question was green.

This question underscores the importance of judges having a good understanding of probability. A situation which looks at first glance clear cut turns out not to be. A highly reliable eyewitness, who gets things right 80% of the time might still be wrong due to the relative proportions of the constituents of the population. Lets move from legal matters to medical matters

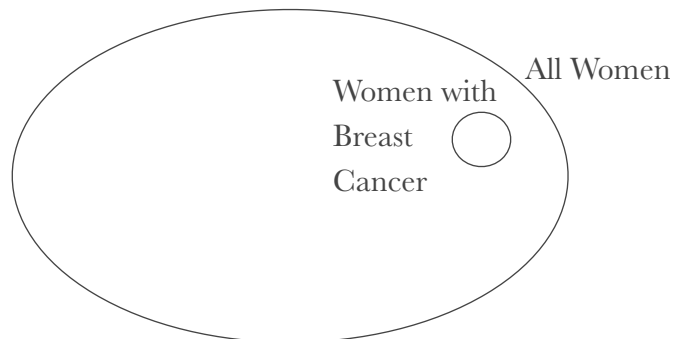
Breast Cancer²

We have the following information

- A. 1% of women have breast cancer (and therefore 99% do not)
- B. 80% of mammograms detect breast cancer when it is there. (true positive)
- C. 9.6% of mammograms detect breast cancer when it's not there (false positive)

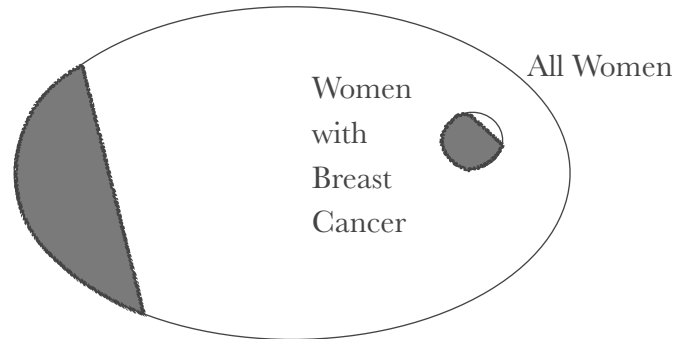
In their regular yearly checkups, somebody you know has tested positive for breast cancer in a mammogram. What is the probability that she actually has breast cancer?

See the following Image:

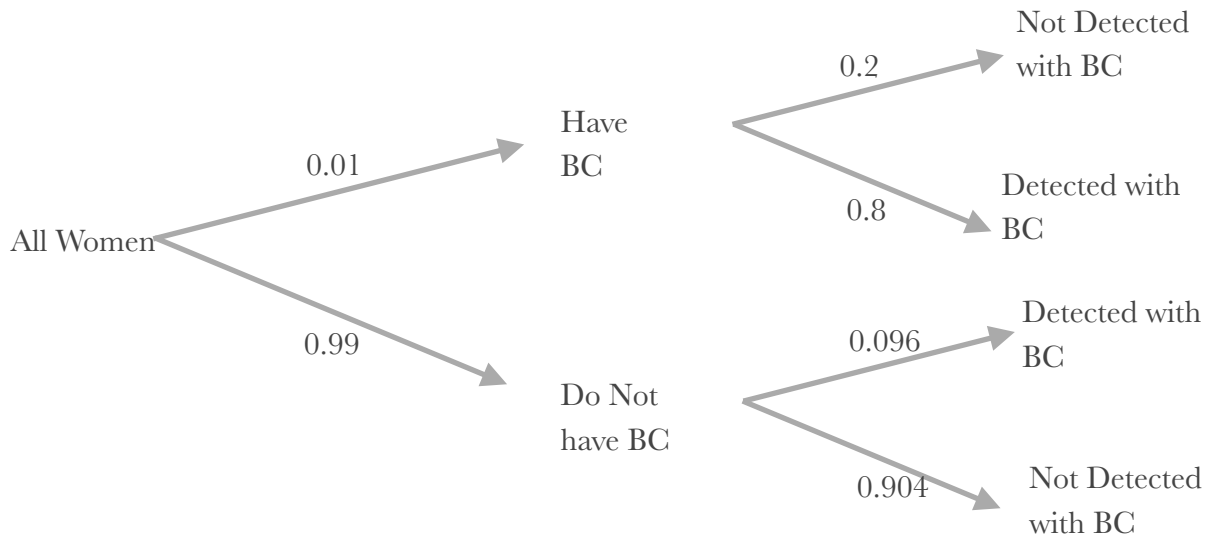


² From <http://www.yudkowsky.net/rational/bayes>

The Outer ellipse represents all women. The inner circle represents women with breast cancer (it is not completely to scale). If a woman has taken a mammogram, the mammogram detects breast cancer 80% of the time when it is there, and 9.6% of the time when it is not there.



The gray bits in the image above represent those who, if they take a mammogram, would be detected as having breast cancer. Though the image is not completely to scale, we can see that it is significantly more likely that the person detected with breast cancer does not have it. Lets look at the same thing using the representation from earlier (BC = Breast Cancer)



Once again, we are interested in situations where the person has been detected with Breast Cancer. We are interested in:

$$P(\text{Has BC} \cap \text{Detected with BC})$$

$$P(\text{Does not have BC} \cap \text{Detected with BC})$$

$$\begin{aligned} P(\text{Has BC} \cap \text{Detected with BC}) &= P(\text{Has BC}) \cdot P(\text{Detected with BC} | \text{Has BC}) \\ &= 0.01 \cdot 0.8 = 0.008 \end{aligned}$$

$$\begin{aligned}
P(\text{Does not have BC} \cap \text{Detected with BC}) &= P(\text{Does not have BC}) \cdot P(\text{Detected with BC} \mid \text{Does not have BC}) \\
&= 0.99 \cdot 0.096 = 0.095
\end{aligned}$$

So, the probability that your friend has breast cancer when it is detected, will be given by:

$$\begin{aligned}
&\frac{P(\text{Has BC} \cap \text{Detected with BC})}{P(\text{Does not have BC} \cap \text{Detected with BC}) + P(\text{Has BC} \cap \text{Detected with BC})} \\
&= 0.008 / 0.103 \\
&= 0.078
\end{aligned}$$

The probability that the person does not have breast cancer is not slightly more, but over 10 times more. This is why it is important for doctors to put together various symptoms before coming to a conclusion. Just the mammogram with no other symptoms is not a good judge of whether a person has breast cancer.

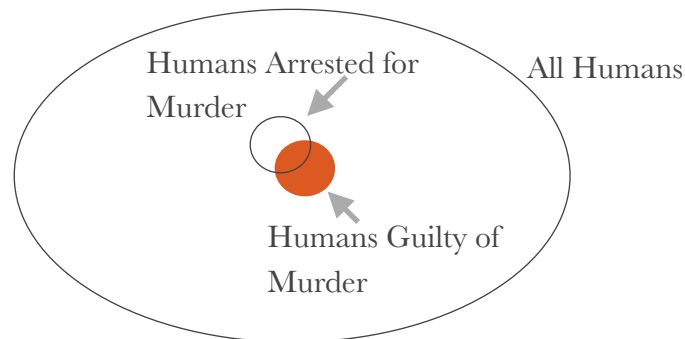
What these three examples show is the importance of the **Base Rate** of the the event in the population. Base Rate is the probability of an event happening in the population without any conditions. The base rate of breast cancer would be 0.01 - before the condition of mammograms is added on. The base rate of green cars in the city is 0.85.

Bayesian Updating

Evaluating whether a claim is true or what decision to take in a particular situation is not usually as straightforward as the examples in the last section. In the real world, situations tend to resemble the lawyer-murder example from the introduction. In that example, as in most other cases, you are not given information all at once. So, you have to constantly update your degree of belief when confronted with new information.

Going back to the murder example, what would be your degree of belief in your client's guilt/innocence before you knew anything about the client apart from the fact that your client has been arrested for murder? The question I'm asking is: what is the appropriate base rate to pick in this situation?

Should we say your client has a 0.00001 chance of being guilty since that is the base rate of murderers in the population? Or, is the appropriate base rate the probability of a person who has been arrested for murder being guilty? The latter seems the better choice, but why? The reason is that there is probably a strong correlation between those who are arrested for murder and those who are guilty.



While the above diagram is not to scale, what it represents is that the base rate of guilt amongst those who are arrested is far more than the base rate of those who are guilty amongst the entire population, even though there are many innocent people who are arrested and many guilty people who are not.

Assume the probability of being guilty when arrested is 0.3. Now, you find out that your client is female. Assume you are told that women make up only 20% of all murderers. Does that change anything?

Lets unpack this. We already are restricted to a world consisting of only those humans who are arrested for murder.