8
Probabilistic reasoning

The set of tools under discussion in this part – which so far have included easy cases and lumping – all offer ways to handle complexity by discarding information. The next tool in this set is probabilistic reasoning. Instead of trying to force the world to give us exact and complete information, we accept that our information will be incomplete and use probability to quantify the incompleteness.

8.1 Is it my telephone number?

To introduce the fundamental ideas, here is an example from soon after I had moved to England sometime in the last century. I was talking to a friend on the phone, and he needed to call me back. Having just moved to the apartment, I was still unsure of my phone number (plus, British phone numbers have a strange format). I had a guess, which I thought was reasonably likely, but I was not very sure of it. To test it, I picked up my phone, dialed the candidate number, and got a busy signal.

▶ *Given this experimental evidence, how sure am I that the candidate number is my phone number? To make the question quantitative: What odds should I give?*

This question makes no sense if one uses the view of probability as long-run frequency. In that view – alas, the most common view – the probability of, say, a coin turning up heads is 1/2 because 1/2 is the limiting

do you discard information when working with easy cases?

Yes, all the information in between the easy cases! its like using boundary conditions to figure out what the middle section must be, without ever actually looking at it

Discarding? Meaning: eliminating information we have already been given to make the problem an "easy case"? I think that clears up a lot of what we have been doing lately.

Maybe you should use the phrase "lossy" since you use it everywhere else when talking about this.

I agree, it would make things more consistent.

At least one adjective in there to make sure it's clear that we're not discarding anything important!

how do you know that your information is incomplete?

This seems like an idea that we have been using all semester, even if it does fit best under "lossy" methods

I like this concept. I do feel like I am always throwing away information in this class to calculate something. It is pretty cool when the calculations end up working out.

I wouldn't say we throw out info in all the methods – we just choose numbers that are easy to deal with. In the end, we get a sense of what the answer tells us, which is much more important then a precise number you don't understand

Still, we don't have a solid basis for how close we should be to the actual result, besides (perhaps) gut instinct and some idea of how wide our estimates were along the way. This (I'm guessing) will give us a chance to quantify our wrongness.

8
Probabilistic reasoning

The set of tools under discussion in this part – which so far have included easy cases and lumping – all offer ways to handle complexity by discarding information. The next tool in this set is probabilistic reasoning. Instead of trying to force the world to give us exact and complete information, we accept that our information will be incomplete and use probability to quantify the incompleteness.

8.1 Is it my telephone number?

To introduce the fundamental ideas, here is an example from soon after I had moved to England sometime in the last century. I was talking to a friend on the phone, and he needed to call me back. Having just moved to the apartment, I was still unsure of my phone number (plus, British phone numbers have a strange format). I had a guess, which I thought was reasonably likely, but I was not very sure of it. To test it, I picked up my phone, dialed the candidate number, and got a busy signal.

▶ *Given this experimental evidence, how sure am I that the candidate number is my phone number? To make the question quantitative: What odds should I give?*

This question makes no sense if one uses the view of probability as long-run frequency. In that view – alas, the most common view – the probability of, say, a coin turning up heads is 1/2 because 1/2 is the limiting

Funny, I just did this not an hour or two ago to calculate my 6.055 grade so. We don't have all our memo grades for the rest of the semester, so the best estimate is an average of what we already have. Assuming we get better at writing memo's over the semester, this method would give a lower bound on our grade.

So in essence we are quantifying that "off factor" that we have been discounting through rounding up and then down throughout the semester?

I haven't read this section yet, but it would be great to finally look at these rounding errors

i disagree with the assumption that our grade will stay the same or get better. as we progress through the semester, we use up our free extensions, etc. thus, doing the same as you've always done, towards the end of the semester, will hurt you more.

Read this section for Friday's lecture (memo due Friday at 9am).

I really like this shift to am deadlines. I really have a tendency to do work later at night than the previous deadlines were set. Now its much easier to get my memos in on time

I agree. I also really liked this section title. It's quite inviting (no offense to the sections on drag...).

Assumin he has time to read before lecture I agree. If this means he doesn't or can read fewer comments, I don't mind getting my work in before 10.

Hahaha, this is a funny phrase - it deceptively makes it sound as though it were very long ago.

I like this little story

8
Probabilistic reasoning

The set of tools under discussion in this part – which so far have included easy cases and lumping – all offer ways to handle complexity by discarding information. The next tool in this set is probabilistic reasoning. Instead of trying to force the world to give us exact and complete information, we accept that our information will be incomplete and use probability to quantify the incompleteness.

8.1 Is it my telephone number?

To introduce the fundamental ideas, here is an example from soon after I had moved to England sometime in the last century. I was talking to a friend on the phone, and he needed to call me back. Having just moved to the apartment, I was still unsure of my phone number (plus, British phone numbers have a strange format). I had a guess, which I thought was reasonably likely, but I was not very sure of it. To test it, I picked up my phone, dialed the candidate number, and got a busy signal.

▶ *Given this experimental evidence, how sure am I that the candidate number is my phone number? To make the question quantitative: What odds should I give?*

This question makes no sense if one uses the view of probability as long-run frequency. In that view – alas, the most common view – the probability of, say, a coin turning up heads is 1/2 because 1/2 is the limiting

What format could the phone numbers be in? I thought all countries had a country code (it just happens that the US has a 1) and then then after that comes an area code and the specific location digits.

Yes, but they're chunked differently. We're very well-trained to remember phone numbers as 3 digits, then 3 digits, then 4 digits. Trying to remember a different number or blocking of digits makes it much harder.

but couldn't you re-chunk it in your head? i often end up doing stuff like that.

Maybe he's referring to how it's chunked differently depending on location? http://en.wikipedia.org/wiki/Telephone_numbers_in_the_United_Kingdom

UK phone numbers, like phone numbers in many European countries, have a variable-length area code (which starts with a 0) and a variable-length number. I think it's now true that all numbers in one area have the same length, but I don't know if that was always true. So, the system makes sense but is strange to an American.

Really, they have variable length as well? That seems so...odd! Wow. They must think that our version is too simplistic when they come here.

Do you have any idea how phone numbers from instant messaging clients are generated? For instance, when I receive a text message from my friend using AIM, it's almost always 25060.

I think that it is funny that you called yourself to figure out if you remember your phone number correctly- did that really happen?

It's a true story, and is how I tested my guess. The example became Exercise 3.13 in David Mackay's textbook _Information Theory, Inference, and Learning Algorithms_ (available online too). [And It was David with whom I was talking.]

I never knew that you could call yourself. Interesting...

On cellphones...how do you think you get to your voicemail?

8
Probabilistic reasoning

The set of tools under discussion in this part – which so far have included easy cases and lumping – all offer ways to handle complexity by discarding information. The next tool in this set is probabilistic reasoning. Instead of trying to force the world to give us exact and complete information, we accept that our information will be incomplete and use probability to quantify the incompleteness.

8.1 Is it my telephone number?

To introduce the fundamental ideas, here is an example from soon after I had moved to England sometime in the last century. I was talking to a friend on the phone, and he needed to call me back. Having just moved to the apartment, I was still unsure of my phone number (plus, British phone numbers have a strange format). I had a guess, which I thought was reasonably likely, but I was not very sure of it. To test it, I picked up my phone, dialed the candidate number, and got a busy signal.

▶ *Given this experimental evidence, how sure am I that the candidate number is my phone number? To make the question quantitative: What odds should I give?*

This question makes no sense if one uses the view of probability as long-run frequency. In that view – alas, the most common view – the probability of, say, a coin turning up heads is 1/2 because 1/2 is the limiting

did you use a cell phone and a land line or something? How did you call with a phone that was already being used?

I was assuming you were still on the phone with your friend at the time. Maybe you should point out that you hung up with him first.

You can call your own number with a land line. Some people do it to talk to other people in different rooms in the same house. If you only have one phone connected in that house then you would probably get a busy signal but if you have multiple phones then the other phones will ring.

Didn't know you could do that.. sounds pretty lazy though.

Right, I hung up, then used the same phone. I didn't have a cell phone [and still don't have one].

Just curious- is this actually what ran through your mind after the incident?

Yes! It also helped that I had moved to England and joined the Bayesian inference group in the physics department.

Hehe cute

As a gambler, I love odds-making.

8

Probabilistic reasoning

The set of tools under discussion in this part – which so far have included easy cases and lumping – all offer ways to handle complexity by discarding information. The next tool in this set is probabilistic reasoning. Instead of trying to force the world to give us exact and complete information, we accept that our information will be incomplete and use probability to quantify the incompleteness.

8.1 Is it my telephone number?

To introduce the fundamental ideas, here is an example from soon after I had moved to England sometime in the last century. I was talking to a friend on the phone, and he needed to call me back. Having just moved to the apartment, I was still unsure of my phone number (plus, British phone numbers have a strange format). I had a guess, which I thought was reasonably likely, but I was not very sure of it. To test it, I picked up my phone, dialed the candidate number, and got a busy signal.

▶ *Given this experimental evidence, how sure am I that the candidate number is my phone number? To make the question quantitative: What odds should I give?*

This question makes no sense if one uses the view of probability as long-run frequency. In that view – alas, the most common view – the probability of, say, a coin turning up heads is 1/2 because 1/2 is the limiting

I'd guess about 100% chance

There's still very possible that it's some one else's busy signal though. I would guess closer to 50%.

so we're talking about the probability that you dialed a number that wasn't yours, it was busy, given that it worked (you at least got a busy signal) and you knew most of the numbers. I'd think that it was pretty unlikely that all this could happen. What if you called again in 20 mins?

I was thinking it would be 75-80% accurate. I think it's be right about 95% if you called it again in 20 mins. 99% in an hour.

I agree with your assessment on how it would go. Theoretically, would it ever be 100% sure to be your number?

Yea, like said above, I would just wait 20 mins and call again. That way if its busy again, you're almost certain that it's your number.

I trust Sanjoy's memory so I'd say 99%.

Woah this is a cool relatable example - I feel like this could very well happen to me if I moved somewhere else...

I smell bayesian probabilities coming on.

I thought this is a great introduction to the section

I agree. Not only is it something that can easily be understood, it is also something that relates to everyone.

I agree. It got me interested very quickly.

I agree too...this little story is a great way to get people engaged and interested in the new topic of probabilistic reasoning.

Are you saying you are quantifying probability?

I think it means he is using probability to quantify uncertainty.

8

Probabilistic reasoning

The set of tools under discussion in this part – which so far have included easy cases and lumping – all offer ways to handle complexity by discarding information. The next tool in this set is probabilistic reasoning. Instead of trying to force the world to give us exact and complete information, we accept that our information will be incomplete and use probability to quantify the incompleteness.

8.1 Is it my telephone number?

To introduce the fundamental ideas, here is an example from soon after I had moved to England sometime in the last century. I was talking to a friend on the phone, and he needed to call me back. Having just moved to the apartment, I was still unsure of my phone number (plus, British phone numbers have a strange format). I had a guess, which I thought was reasonably likely, but I was not very sure of it. To test it, I picked up my phone, dialed the candidate number, and got a busy signal.

▶ *Given this experimental evidence, how sure am I that the candidate number is my phone number? To make the question quantitative: What odds should I give?*

This question makes no sense if one uses the view of probability as long-run frequency. In that view – alas, the most common view – the probability of, say, a coin turning up heads is 1/2 because 1/2 is the limiting

In case anyone else was confused by the term "long run frequency"- it refers to the idea that as the number of trials approaches infinity, the relative frequency will converge exactly to the probability lim (x-> inf) N_x/N_t

Hmm, so is it similar to the Law of Large Numbers, then?

yes...Long-Run Frequency is another name for the Law of Large Numbers. ....an explanation of Long-Run Frequency would probably be useful in the book. (maybe in a side note?)

the relative frequency of what will converge?

Is this like considering only end-behavior?

why does it make no sense? i'm confused as to what your point is here.

You can't just look at frequencies, it wouldn't work. You need to approach it differently

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \quad (8.1)$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

Comments on page 2

Typo: should be evaluation.

should actually be "evaluating" i think, not "evaluation".

Right, or "evaluation of"

I feel that this sentence could be worded a little better

Yeah me too, I think it's saying that you can't extract that 1/2 from continuous flips, the 1/2 is from the probability of a single flip.

I didn't understand that's what it was trying to say until I read your comment.

Also: Rosencranz and Guildenstern are Dead.

i did a double take and now understand what you mean, but when i first read it, the word "series" triggered the idea of infinite series in my head, since i'm familiar with them for probability.

typo "evaluating"

This is interesting. I dont know much about probability, and normally when I think of probability this is what I think of...

This seems to be a very strange thing to look for since we are not given imuch information on what we want to claculate.

This reminds me of the way we have to establish that we're dealing with an "ensemble of identically-prepared systems" in thermodynamics, since we can't repeat the same process on a single system.

we could estimate the number of phone numbers in England, and the percentage of those in use at that moment

Yeah, I was going to go digit by digit (How sure are you of the first digit, how sure are you about the second, etc). Given that there are 10 choices for each digit... etc.

I feel like it would be easier just to try the same number again later, then the chance of it being busy both times if it isn't your number is way low.

I am a bit confused by what this term means can you define it

he discusses this idea in the following paragraph

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

but if you called the number over and over again, and it was always busy, wouldn't that make you more confident in your guess? or am I interpreting the frequentist interpretation incorrectly...

I really like that idea- I didn't even think of it

I agree. I would think repeating the call would be helpful in that someone may answer the number on a future call, which would prove the number was incorrect? You could even treat the person being on the line (creating the busy signal) as a fixed probability based on the average person's phone use. So I think I may have also missed something in the question.

That's a good idea. But it's not the same as repeating the same experiment many times, because with each phone call you're collecting additional evidence.

the way that I read the first full sentence here, it made it seem like frequentist interpretation would be explained in the next sentence (or paragraph), but then the question/answer that you pose after made me re-think this...I went back to the beginning of the paragraph to try and figure it out (and it took a couple of reads).

I'm still not entirely sure why it makes _no_ sense...I mean if you try enough times with reasonable timing between calls, you will reach the point where you're pretty damn sure it's the right number.

I guess what I'm saying is that i'd rather see a short sentence explaining than a question/answer that's just repeating the same thing.

couldnt you call the number multiple times? the more you call, then the probablility approaches 1 that it is your number

that's a very different situation, because you are gaining certainty in your answer (yes or no) over time. in the coin drop, your answer is a fraction, and your calculations converge to that fraction over time.

Indeed. The actual probability of the situation is either 0 or 1 - the distinction needs to be drawn between the probability that the phone is yours and how strongly you believe the number you gave is the correct number.

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

I would figure it would be similar to picking a phone number at random within the same area code, potentially one in millions

This was also my first impression, but I think the fact that he wasn't completely unsure makes it difficult to quantify. For instance, I'm sure he'd know if his number was 111-1111, or something like this.

Even ignoring the simple cases for memorable phone numbers, its still a whole lot of possibilities...I feel like the work that would have to be put into eliminating obvious choices wouldn't change the number of options significantly

I agree, I think it has a lot to do with the number of potential choices presented, and how one could sort and limit them to reach a probabalistic answer.

Or is it a property of the flipper? What if the flipper was really good and could reproduce his flip of the coin with enough accurary that it became more likely for a certain side to come up. What if the flip were preformed by a robot. Then would the results of a coin toss be deterministic. this would be an interesting social experiment. Would people still consider a coin toss a random even if performed by a machine?

I'm pretty sure this is the idea behind pseudo-randomness. Most people still consider that to be "random," despite it not being by any means random, only seeming so.

The more you know about the situation, the more you know about H versus T. Friends of my parents knew Persi Diaconis well; he's a famous statistician and also a magician. I remember at dinner parties that he could make the coin come up any way he wanted. So, if he said, "I'm going to toss it in the air, it'll flip a lot, and it'll come up heads," my probability of heads would be pretty high – even though the coin was fair.

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

this is going to be a tough concept to get out of my head, because that's definitely how i view coin toss probability

i don't think you should get it out of your head...it's a system that works in this situation (and works _very_ well too).

Just don't expect the method to always work.

Just think of it as: the coin is going to come up heads or it's going to come up tails. To an extent (at least in Newtonian mechanics), the outcome is predetermined – you just have no reason to suspect one outcome more than the other based on your current knowledge (for a fair coin, of course).

I am wondering why this is so. You say that the interpretation is incorrect but don't really explain why.

I'm a bit unclear if you're also saying that the placement is incorrect for the telephone example, or for the coin also. If it is also for the coin, I don't understand why either.

yeah i'm also really confused about why the placement is incorrect...I feel like a different word should be used rather than "placement" to make it less confusing.

i'm confused by this paragraph as well. i dont have a really good grasp of probability to begin with, and i dont understand the differences being outlined here.

does this mean that probability cannot be placed in the physical system itself?

I think he's saying in this type of problem you cannot.

It certainly seems right that in the phone number example the probabilities are not objective. But the text seems to endorse the stronger view that probabilities are never objective. But aren't the probabilities in quantum mechanics objective (at least, on some interpretations of quantum mechanics)? And even with the coin flip, you might think that the symmetry of the coin is an objective, physical property of the coin that can ground objective probabilities of heads and tails. Why not be a pluralist and say that some probabilities are objective and some are subjective?

y is it incorrect?

and moreover, what is then correct?

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

Is this saying that the frequentist approach mutually exclusive from the Bayesian interpretation of probability? If so, if you had enough data in this specific case, wouldn't the frequentist approach apply also?

I take it Sanjoy is a Bayesian.

All of the lead-up he does to Bayes' theorem is incontrovertible - it's simple math. There's only the sense of "being a Bayesian" because the probability of E given H, for example, is rarely obvious. If you have solid rationale for the numbers you use, being a "non-Bayesian" just doesn't even make sense. Though, as with all statistics, they can be used poorly and to the benefit of whoever is paying the bill of the statistician, so you have to analyze all the assumptions being made.

Definitely!

i just wanted to say that i find this reading super interesting! :)

This is a really good layman's definition of Bayesian probability. I wish it was stated this clearly in 6.01/6.042.

typo

Might be nice here to just provide a 2X2 table to quickly compare the frequentist and Bayesian probability definitions. (2 cells would just be 'Bayesian' or 'Frequentist)

just "certainty" seems clearer than "state of certainty"

I disagree. I don't think you want to use "certainty," since that implies that you are certain about things. A "state of certainty" reflects how certain you are. It's subtly different.

OK, I'll go along with that. I'm just used to hearing certainty as "degree of certainty". "State" implies you're in or out, certain or not, which is also not what we mean. I think degree of certainty would be better here, and a comparative google search has "degree of certainty" 100x more common and more related to probability than "state of certainty".

I really like the "degree of certainty" phrase as well. It does give a feel of a measure more than "state" does.

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

This reminds me of a friend of mine who likes to claim that all probability is 50%-50% – either it will happen, or it won't!

Is this directly related to the person's assumptions? therefore different assumptions result in differences in calculating assumptions

Right (if you include background knowledge as part of the assumptions)!

Could you clarify this? I would push back and say Bayesian probability is objective in the sense that you have defined formulas that you use.

But elements in Bayes' rule are subjective, like picking a prior probability.

Yes, everything hinges on the prior. In 6.041 we kind of just take the prior for granted, but in reality, all of the calculations depends on the subjectivity of the prior probability.

Does this mean that the probability of something happening can change as we gain more knowledge of it, or even that it will change?

Yes, as you know more, the probability will change. From an example I gave in an earlier comment: If you know that someone is gong to flip a coin, and that's all you know, you'll say P(heads)=1/2. But if you find out that the "someone" is a talented magician who has said, "I will flip it high in the air and it'll come up heads,", you might say P(heads)=0.9 or 0.95. (I knew someone who could do that.)

this paragraph is explained a lot better.

This sums up pretty nicely why bayesian probability is subjective. Its based on our own current beliefs.

I'd never heard this idea explained so quickly and yet make so much sense. I wish someone had said it this way earlier

It also makes the whole Monty Hall thing make more sense to me. But yeah, this is a really clear explanation.

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

wait can we see an example right here before moving on to make sure we understand?

I think what he means here is that given everyone's different levels of knowledge, everyone may come up with a different probability for the same problem

How does this relate to coin flipping? Don't we all agree it is 1/2? How is that subjective?

Does this mean Bayesian inference assumes nothing probablistic can be independent?

I think the coin-flipping is used as an example because we know the outcome. Suppose another person knew something about how coins degrade that make one side significantly more probably as time approaches infinity. I feel like they would have a subjective and more accurate approximation of the probability.

This idea is still vague to me even after reading the case multiple times. Maybe its because I don't have much experience in probability.

Its sort of equivalent to the more you know the better you can assume things...like why Sanjoy can make all these really fast guesses in class that confound most of us. If you know more about the situation, assuming its more complicated than how many times a coin has been flipped, it tells you more about what you should assume the future probabilites will converge to.

I feel like the first idea is kind of in the second idea too. Collecting information is also subjective; people choose what kind of information to collect.

And there can be a bias as to what information is available to be collected. This is especially true in surveys; certain types of people are more likely to respond.

Aren't these two related? the more personal information is on an individual basis, should the range of expectation change?

yes, they are _very_ related...however they are not the _same_. 1 = it is, where 2 = it can change.

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

This is a key point I think.

So if this subjective probability allows each person to create a different set of probabilities, is this due to different understandings of the world or different understanding of the evidence, or different obvs. evidence?

I believe it's due to the latter: evidence from observations.

It seems like these are the types of subconscious probabilities we run through our minds every day, interesting

I think it's a mix of all three, with the most prevalent being a different understanding of the evidence.

I really like this way of ending the paragraph. This sentence is concise and very clear, leaving the reader with the this important idea that evidence affects probability values.

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

I don't really understand how this would apply to coin tossing? No matter our personal thoughts about tossing a coin, the probability of heads is always 1/2... I can see how this would apply to the phone number case because there we have to accept that the probability of that number being ours is extremely high because the other option is extremely rare... however, unless the example is something like that, I can't really grasp the application of this idea..

You bring up an interesting point.

I believe that it can be applied to the coin toss example in the following way:

Person 1 has no insight, evidence, or experience with tossing coins. He might believe that 'heads' will appear more often than tails.

Person 2 works at an arcade and spends much of his day flipping coins. In his experience, he knows that 'heads' and 'tails' appear at nearly the same frequency.

As such, "nature" has already determined the probabilities for everything in our world. It is through our experiences that we make our own interpretations that, in some easier cases, we are able to match nature's probabilities.

> "As such, "nature" has already determined the probabilities for everything in our world. It is through our experiences that we make our own interpretations that, in some easier cases, we are able to match nature's probabilities."
>
> ...well put! Thank you

The more you know about the situation, including coin tossing, the more you know about H versus T. What follows is an example repeated from an earlier comment – I wish NB had a way of connecting threads together...

Friends of my parents knew Persi Diaconis well; he's a famous statistician and also a magician. I remember at dinner parties that he could make the coin come up any way he wanted. So, if he said, "I'm going to toss it in the air, it'll flip a lot, and it'll come up heads," my probability of heads would be pretty high – even though the coin was fair.

I little table of the variables and what they mean would be helpful here.

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

This paragraph is extremely clear and concise.. thanks!

Yeah. While I'd like to think most of us have already come across this before, it is definitely useful as a refresher.

what is this symbol? i'm not familiar with it.

It translates to "given" as explained in the rest of the sentence.

It means H given E

Yup, basically if E MUST have happened, what is the likelihood that H happened either as a result or as well?

what would you base this probability assignment off of? (I'm assuming it's a number between 0 and 1, so where on the scale would you place it?)

This is the probability we're actually interested in (the probability of our hypothesis being true given our data), so we're not going to pick a value for it. (See the last sentence of the page.)

Do you need some statement that the evidence is true?

Here, E has been observed, so it's "true" insofar as we have seen it happen.

typo Bayes'

i hate to be the grammar nazi here, but technically, it should actually be Bayes's. an s-apostrophe is used for s's that come from the plural form. an s-apostrophe-s is used for words and names that naturally end in s, such as Bayes.

Actually, there seems to be support for both ways. I've usually heard it pronounced "Bayes rule" not "Bayeses rule" (maybe the reason for the original form here?), so this consideration applies: * If the singular possessive is difficult or awkward to pronounce with an added sibilant, do not add an extra s; these exceptions are supported by The Guardian,[17] Emory University's writing center,[18] and The American Heritage Book of English Usage.[19] Such sources permit possessive singulars like these: Socrates' later suggestion; James's house, or James' house, depending on which pronunciation is intended. (caveat: from wikipedia, whose entry has "Bayes'")

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

i feel like in the phone number case it would not be as helpful to have a formula since the evidence and the initial probability are so subjective anyway

You could have a diagram of the sample space and an event to make this more clear.

what do you mean by "mental world"?

typo? hypotheses

I'm more used to seeing this probability written as the union of H and E, is the symbol that looks like an upside down u, don't know if its worth introducing though.

I'm also more used to that from 6.041, but we use the & notation in 8.044. I guess it's just a matter of convention.

yeah I think the union sign would be better since it is a more wide spread convention when talking about probability.

Conditional Probability

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \qquad (8.1)$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

This is a great refresher for those who have seen probability before, but I wonder if this moves too quickly for someone who has never taken 6.041/18.440.

I don't think so. I felt this section was relatively self-contained. If the reader is still shaky, there are literally tons of probability books out there to which he can refer.

Is there a good illustration (marbles?) for the joint probability being equal to P(A given B) * P(B)?

I find the best way to think about it is sequentially using a tree. Also one can always go back to the coin toss example. For example, whats the probability of flipping two heads in a row. P(H on 1st toss & H on 2nd toss) = P(H on 1st) * P(H on 2nd | heads on 1st) = (1/2)*(1/2) = 1/4.

Its probably a too simple example since the probability of the 2nd head is the same no matter what was the previous toss. Its more apparent when a loaded coin is involved.

I feel like for what we are doing in this class, this concise explanation is fit... It is a nice refresher for those who have taken a course in probability and I feel like it is a clear enough summary for people who are unfamiliar with the topic. Any longer and I agree that it would take away from the point of the section.

I have never taken probability and had to re-read the paragraph but feel confident to move on.

I have never taken a probability class and I agree it is a bit much but I generally think that is ok to understand except I don't get why the two paths to H&E must be equal?

This was very difficult for me to follow. Maybe you could include a short example to illustrate this point (such as the marble example mentioned earlier).

This explanation is really helpful (especially when reflecting on old 6.041 work).

I also took 6.041 a couplte semesters ago, but it took me reading this paragraph very slowly to really understand this line of thought. I think even breaking it up into multiple paragraphs would help clarify the ideas here.

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

I agree it may take away from the focus of this to go on at more length about this, but I was also confused by this having never taken a course in probability...

or the quick example of diseases and positive or negative tests would provide a bit of clarity for the uniformed reader.

I have never taken a course on probability before and I admit it is a little difficult to follow.

While I think having a probability introduction or refresher for a few pages would be helpful to those students who haven't had previous exposure to probability, it would probably (no pun intended) detract from the focus in this text, and in this section in particular. I think it would be best to have an addendum or an appendix with a probability intro or refresher to those students who need it.

I do think it was a bit rushed. It's not so much that the material is too difficult or anything, but just the way that the paragraph is written seems very condensed.

It took me a while to fully understand the notation. I feel that removing P(H&E) and instead just using P(H|E) with the explanation would be better.

This notation does take a bit to understand but I think the way it's current done is important because it allows for uniformity throughout the process.

It might also help the students who have never taken 6.041 or the equivalent.

Are these two probabilities in fact always equal to each other? Because it really doesn't seem that way to me.

I feel like we could have reached this conclusion with much less explanation. I think it made it seem more complicated than it is.

I agree. P(E|H)P(H)=P(EˆH)=P(H|E)P(E) and done.

For people who have seen probability in the recent past or have a very intuitive grasp of it a really quick explanation would suffice, and for those that under stand it this paragraph can easily be skipped over. I think the positives of it being here for anyone that's confused about this step outweighs the negatives of not doing so as understanding is necessary for the rest of the section.

I find it easier to see this if there is also an "=P(E,H)" here for the joint probabilit

proportion of heads in an ever-longer series of tosses. However, for evalutin the plausibility of the phone number, this interpretation – called the frequentist interpretation – makes no sense. What is the repeated experiment analogous to tossing the coin repeatedly? There is none.

The frequentist interpretation places probability in the physical system itself, as an objective property of the system. For example, the probability of 1/2 for tossing heads is seen as a property of the coin itself. That placement is incorrect and is the reason that the frequentist interpretation cannot answer the phone-number question. The sensible alternative – that probability reflects the incompleteness of our knowledge – is known as the Bayesian interpretation of probability.

The Bayesian interpretation is based on a two simple ideas. First, probabilities reflect our state of certainty about a hypothesis. Probabilities are explicitly *subjective:* Someone with a different set of knowledge will use a different set of probabilities. Second, by collecting evidence, our state of certainty changes. In other words, evidence changes our probability assignments.

In the phone-number problem, the hypothesis H is that my candidate number is correct. For this hypothesis, I have an initial or prior probability $P(H)$. After collecting the evidence E – that when I dialed this number the phone was busy – I make a new probability assignment $P(H|E)$ (the probability of the hypothesis H given the evidence E).

The recipe for using evidence to update probabilities is known as Bayes theorem. To derive it, imagine that the mental world contains only two hypothesis H and $\overline{H}$, with probabilities $P(H)$ and $P(\overline{H}) = 1 - P(H)$, and that we have collected some evidence E. Now write the joint probability $P(H\&E)$ in two different ways. $P(H\&E)$ is the probability of H being true and E occurring. That probability is, first, the product $P(H|E)P(E)$ – namely, the probability that H is true given that E occurs times the probability that E occurs. $P(H\&E)$ is, second, the product $P(E|H)P(H)$ – namely, the probability that E occurs given that H is true times the probability that H is true. These two paths to $H\&E$ must produce identical probabilities, so

$$P(E|H)P(H) = P(H|E)P(E). \tag{8.1}$$

Our goal is the updated probability of the hypothesis, namely $P(H|E)$. It is given by

Is there a proof for this?

I am always fascinated by how 6.055 incorporates so many different disciplines together.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}. \quad (8.2)$$

Similarly, for the opposite hypothesis $\overline{H}$,

$$P(\overline{H}|E) = \frac{P(E|\overline{H})P(\overline{H})}{P(E)}. \quad (8.3)$$

Using the ratio $P(H|E)/P(\overline{H}|E)$, which is known as the odds, gives an even simpler formula because $P(E)$ is common to both probabilities and therefore cancels out:

$$\frac{P(H|E)}{P(\overline{H}|E)} = \frac{P(E|H)P(H)}{P(E|\overline{H})P(\overline{H})}. \quad (8.4)$$

The right side contains the factor $P(H)/P(\overline{H})$, which is the initial odds. Using O for odds,

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})}. \quad (8.5)$$

This result is Bayes theorem (for the case of two mutually exclusive hypotheses).

In the fraction $P(E|H)/P(E|\overline{H})$, the numerator measures how well the hypothesis H explains the evidence E; the denominator measures how well the contrary hypothesis $\overline{H}$ explains the same evidence. Their ratio, known as the likelihood ratio, measures the relative value of the two hypothesis in explaining the evidence. So, Bayes theorem has the following English translation:

$$\text{updated odds} = \text{initial odds} \times \text{relative explanatory power}. \quad (8.6)$$

Let's see how this result applies to my English telephone number. Initially I was not very sure of the phone number, so $P(H)$ is perhaps 1/2 and $O(H)$ is 1. In the likelihood ratio, the numerator $P(E|H)$ is the probability of getting a busy signal given that my guess is correct (given that H is true). If my guess is correct, I'd be dialing my own phone using my phone, so I would definitely get a busy signal: $P(E|H) = 1$. The hypothesis of a correct number is a very good explanation of the data.

The trickier estimate is $P(E|\overline{H})$: the probability of getting a busy signal given that my guess is incorrect (given that $\overline{H}$ is true). If my guess is

Comments on page 3

haha ok here's what I was looking for.

I'm really glad we're getting into Bayes' theory. I had thought it would be very useful given the work we've been doing.

I like that we are doing probability, but I'm still trying to figure out how it will work in to approximations.

I completely disagree, I find myself routinely estimating probabilities in everyday life. All the little decisions we make, like whether to take the longer route with less traffic, depends on the probabilities of the situation.

While I am familiar with conditional probability, I think that a short recap on it might be helpful for other readers, particularly outside of MIT.

I disagree...I think the line "the probability that H is true given that E occurs" on the previous page is good enough

The whole paragraph before this one was a short recap

so in this problem, the opposite hypothesis is that it isn't your number correct?

I think thats the case...and then it would become the likelihood that the phone was busy given that it was NOT his phoneline

Do you think a short introduction to probability before this lecture would be helpful?

I agree. While I understand the concepts of probability presented in this section, it would be very hard to understand without some sort of background in probability.

I think that just a tiny bit more introduction to some of the terms would be good, but too much intro, and it would detract from the point, might be better as a appendix.

I think its safe to assume students in college in engineering fields will have some experience or at least an intuitive feel for basic probability, so I think the intro included is sufficient.

So maybe I'm just being too hesitant. But this still works even though P(H_bar) is = 1-P(H)?

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}. \tag{8.2}$$

Similarly, for the opposite hypothesis $\overline{H}$,

$$P(\overline{H}|E) = \frac{P(E|\overline{H})P(\overline{H})}{P(E)}. \tag{8.3}$$

Using the ratio $P(H|E)/P(\overline{H}|E)$, which is known as the odds, gives an even simpler formula because $P(E)$ is common to both probabilities and therefore cancels out:

$$\frac{P(H|E)}{P(\overline{H}|E)} = \frac{P(E|H)P(H)}{P(E|\overline{H})P(\overline{H})}. \tag{8.4}$$

The right side contains the factor $P(H)/P(\overline{H})$, which is the initial odds. Using O for odds,

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})}. \tag{8.5}$$

This result is Bayes theorem (for the case of two mutually exclusive hypotheses).

In the fraction $P(E|H)/P(E|\overline{H})$, the numerator measures how well the hypothesis H explains the evidence E; the denominator measures how well the contrary hypothesis $\overline{H}$ explains the same evidence. Their ratio, known as the likelihood ratio, measures the relative value of the two hypothesis in explaining the evidence. So, Bayes theorem has the following English translation:

$$\text{updated odds} = \text{initial odds} \times \text{relative explanatory power}. \tag{8.6}$$

Let's see how this result applies to my English telephone number. Initially I was not very sure of the phone number, so $P(H)$ is perhaps 1/2 and $O(H)$ is 1. In the likelihood ratio, the numerator $P(E|H)$ is the probability of getting a busy signal given that my guess is correct (given that H is true). If my guess is correct, I'd be dialing my own phone using my phone, so I would definitely get a busy signal: $P(E|H) = 1$. The hypothesis of a correct number is a very good explanation of the data.

The trickier estimate is $P(E|\overline{H})$: the probability of getting a busy signal given that my guess is incorrect (given that $\overline{H}$ is true). If my guess is

In this reading, "odds" refers to a ratio, for example "1 to 10". However, in common vocabulary or day-to-day colloquialism, I've heard the word "odds" being used as: "The odds are 1 in a million", which would be equivalent to "1 to 999,999".

This is a slight difference that didn't get specified clearly until the last page, when you said "1 to 10". It might make a difference occasionally, but I guess usually it won't matter that much.

The day-to-day term you mentioned is a misuse of the term. When i think of odds, I think of betting, like horse racing, where they say "the odds of a horse winning is 3 to 2".

Yeah, the "odds" refers to the favorability of choosing one outcome as opposed to another (or others). Odds in favor of an event are p/(1-p).. The odds of choosing a day of the week and choosing sunday is (1/7)/ (1-1/7)= 1/6; whereas, the probability of choosing a sunday is just p= 1/7. (info from wikipedia)

Thanks, the day of the week example really helped clarify this for me.

I read most of the comments about this, but I am still confused, I would think this could use much more explanation.

I am a little worried about using this to solve problems by approximating...I'm finding it hard to keep straight in my head.

So in reading other comments on this reading so far, it seems like everyone has a different opinion of what the difference between probability and odds are. Personally, I was taught that odds is always written as a ratio of the number of times an event will happen to the number of times an event won't happen. (i.e. if the probability of an event happening is 1/3, then the odds are 1:2.) But since there seems to be so much confusion, I think a detailed explanation of the definition of odds used here would be useful.

wait maybe this won't be so foreign to me – this seems like something i learned in high school stats.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}. \quad (8.2)$$

Similarly, for the opposite hypothesis $\overline{H}$,

$$P(\overline{H}|E) = \frac{P(E|\overline{H})P(\overline{H})}{P(E)}. \quad (8.3)$$

Using the ratio $P(H|E)/P(\overline{H}|E)$, which is known as the odds, gives an even simpler formula because $P(E)$ is common to both probabilities and therefore cancels out:

$$\frac{P(H|E)}{P(\overline{H}|E)} = \frac{P(E|H)P(H)}{P(E|\overline{H})P(\overline{H})}. \quad (8.4)$$

The right side contains the factor $P(H)/P(\overline{H})$, which is the initial odds. Using O for odds,

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})}. \quad (8.5)$$

This result is Bayes theorem (for the case of two mutually exclusive hypotheses).

In the fraction $P(E|H)/P(E|\overline{H})$, the numerator measures how well the hypothesis H explains the evidence E; the denominator measures how well the contrary hypothesis $\overline{H}$ explains the same evidence. Their ratio, known as the likelihood ratio, measures the relative value of the two hypothesis in explaining the evidence. So, Bayes theorem has the following English translation:

$$\text{updated odds} = \text{initial odds} \times \text{relative explanatory power}. \quad (8.6)$$

Let's see how this result applies to my English telephone number. Initially I was not very sure of the phone number, so $P(H)$ is perhaps 1/2 and $O(H)$ is 1. In the likelihood ratio, the numerator $P(E|H)$ is the probability of getting a busy signal given that my guess is correct (given that H is true). If my guess is correct, I'd be dialing my own phone using my phone, so I would definitely get a busy signal: $P(E|H) = 1$. The hypothesis of a correct number is a very good explanation of the data.

The trickier estimate is $P(E|\overline{H})$: the probability of getting a busy signal given that my guess is incorrect (given that $\overline{H}$ is true). If my guess is

I am still a little unsure of the difference between probability and odds, as well as how we get from this equation to that.

Let p the probability that an event occurs. Then the probability that the event does not occur = 1 - p. The odds of an event occurring refers to the ratio of the probability that the event occurs divided by the probability that the event does not occur. That is, the odds = p/(1-p).

Im a little confused about just throwing in the O for odds.

I agree. To clarify why I'm confused, the odds of getting a head for a coin is 1/2. But P(H)/P(T) = 1. Maybe I have the definition wrong...

I've been using Baye's thm a lot recently- definitely have not seen it like this, or explained like this.

For some reason, I don't remember dealing with this in 6.041. Yet this form of the theorem seems very useful.

isn't it harder in this case to guess these other probabilities than just guessing the entire probability

I'm getting lost and would like to see an example soon

This is a great explanation of Bayes' theorem

I concur, I always knew it just because I knew it and never thought about the reasoning behind it like this.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}. \qquad (8.2)$$

Similarly, for the opposite hypothesis $\overline{H}$,

$$P(\overline{H}|E) = \frac{P(E|\overline{H})P(\overline{H})}{P(E)}. \qquad (8.3)$$

Using the ratio $P(H|E)/P(\overline{H}|E)$, which is known as the odds, gives an even simpler formula because $P(E)$ is common to both probabilities and therefore cancels out:

$$\frac{P(H|E)}{P(\overline{H}|E)} = \frac{P(E|H)P(H)}{P(E|\overline{H})P(\overline{H})}. \qquad (8.4)$$

The right side contains the factor $P(H)/P(\overline{H})$, which is the initial odds. Using O for odds,

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})}. \qquad (8.5)$$

This result is Bayes theorem (for the case of two mutually exclusive hypotheses).

In the fraction $P(E|H)/P(E|\overline{H})$, the numerator measures how well the hypothesis H explains the evidence E; the denominator measures how well the contrary hypothesis $\overline{H}$ explains the same evidence. Their ratio, known as the likelihood ratio, measures the relative value of the two hypothesis in explaining the evidence. So, Bayes theorem has the following English translation:

$$\text{updated odds} = \text{initial odds} \times \text{relative explanatory power}. \qquad (8.6)$$

Let's see how this result applies to my English telephone number. Initially I was not very sure of the phone number, so $P(H)$ is perhaps 1/2 and $O(H)$ is 1. In the likelihood ratio, the numerator $P(E|H)$ is the probability of getting a busy signal given that my guess is correct (given that H is true). If my guess is correct, I'd be dialing my own phone using my phone, so I would definitely get a busy signal: $P(E|H) = 1$. The hypothesis of a correct number is a very good explanation of the data.

The trickier estimate is $P(E|\overline{H})$: the probability of getting a busy signal given that my guess is incorrect (given that $\overline{H}$ is true). If my guess is

This is actually quite interesting - this is different from the Bayes' theorem one normally sees: P(A|B) = P(B|A)P(A)/P(B), but they are certainly interrelated.

Actually, it's the exact same thing we see 'normally' from Bayes' theorem. If you scroll to the bottom of page 152, you'll see the equation in the form you described above. All he did here was put Bayes' theorem into a ratio of odds. They're more than interrelated, they're the same thing!

Yeah i don't think i've ever seen it in the context of odds before.

Same, I've never seen it in the context of odds, but this is a really interesting application of Bayes theorem

Agreed! I've derived it before, but never this way and never with such an obvious example

Bayes'

I've seen Bayes theorem before, but it's nice to think of it as a clear cut ratio of which hypothesis is right.

While I agree with this statement, I feel that "explains" isn't quite the proper word here. Perhaps something more like "predicts" would be better. Because fundmantally, this is the probability that E occurs given H is true, which translates to the probability that we would find evidence E if our hypothesis were true. "Explains" sounds as if our hypothesis were already true.

I agree with this. I dont really understand what "explains" would mean here. the above statement clears things up a little but its still a little fuzzy. I am not quite sure what exactly this ratio tells us in laymens' terms.

While I happen to agree with a Bayesian outlook on probabilities, I think it might be useful to mention the problem frequentists have with the Bayesian approach, namely that we need some p(H) before making any measurements, and it's hard to have a probability of a hypothesis without any evidence whatsoever.

To make it easier to follow and understand, maybe instead of separating the explanation and the example, you could merge them together.

Though, in my opinion this is one of the easier to follow and more interesting sections we have had.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}. \quad (8.2)$$

Similarly, for the opposite hypothesis $\overline{H}$,

$$P(\overline{H}|E) = \frac{P(E|\overline{H})P(\overline{H})}{P(E)}. \quad (8.3)$$

Using the ratio $P(H|E)/P(\overline{H}|E)$, which is known as the odds, gives an even simpler formula because $P(E)$ is common to both probabilities and therefore cancels out:

$$\frac{P(H|E)}{P(\overline{H}|E)} = \frac{P(E|H)P(H)}{P(E|\overline{H})P(\overline{H})}. \quad (8.4)$$

The right side contains the factor $P(H)/P(\overline{H})$, which is the initial odds. Using O for odds,

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})}. \quad (8.5)$$

This result is Bayes theorem (for the case of two mutually exclusive hypotheses).

In the fraction $P(E|H)/P(E|\overline{H})$, the numerator measures how well the hypothesis H explains the evidence E; the denominator measures how well the contrary hypothesis $\overline{H}$ explains the same evidence. Their ratio, known as the likelihood ratio, measures the relative value of the two hypothesis in explaining the evidence. So, Bayes theorem has the following English translation:

$$\text{updated odds} = \text{initial odds} \times \text{relative explanatory power}. \quad (8.6)$$

Let's see how this result applies to my English telephone number. Initially I was not very sure of the phone number, so $P(H)$ is perhaps 1/2 and $O(H)$ is 1. In the likelihood ratio, the numerator $P(E|H)$ is the probability of getting a busy signal given that my guess is correct (given that H is true). If my guess is correct, I'd be dialing my own phone using my phone, so I would definitely get a busy signal: $P(E|H) = 1$. The hypothesis of a correct number is a very good explanation of the data.

The trickier estimate is $P(E|\overline{H})$: the probability of getting a busy signal given that my guess is incorrect (given that $\overline{H}$ is true). If my guess is

i'm having trouble seeing the connection between the written words and the formulas. here i dont know what explanatory power means.

I think this is a really helpful summary–now I can go back and think through the series of formulas to make sense of everything

This is a great summary!

Wow, i never thought of conditional probability in terms of odds. This should help me remember that formula.

I agree. This word-equation is extremely helpful. And thinking about Bayes' Theorem in terms of odds is a much more intuitive way of thinking about it, at least for me.

I don't think this is as clear for people who have not taken a probability class

i really like the way this equation is set out, so that it's easy to read. thank you for the formatting (even if it is mostly LaTeX)

I agree that this way of stating the general idea of Bayes theorem is really helpful after having gone through the derivation, seeing as it gives you a way to think about it more intuitively rather than in term of the formula.

This is somewhat unclear because you have juxtaposed odds and probabilities, which on first pass, are pretty similar...

It would be nice to define odds as relative probabilities with a short equation O(H) = P(H)/P(barH)

Wouldn't O(H) be also equal to 1/2? I am confused how a probability like that can be equal to 1

Its explained later in the paragraph

I had to go back and think about this for a few seconds. Maybe spell it out even more clearly. P(H)=P(Hbar)=1/2, so 1/2 / 1/2=1 or something

I agree, it would be nice to see this more clearly listed so it could be compared more easily.

I agree with what others have been saying. It is confusing to read that O(H) is 1. What does that mean?

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}. \qquad (8.2)$$

Similarly, for the opposite hypothesis $\overline{H}$,

$$P(\overline{H}|E) = \frac{P(E|\overline{H})P(\overline{H})}{P(E)}. \qquad (8.3)$$

Using the ratio $P(H|E)/P(\overline{H}|E)$, which is known as the odds, gives an even simpler formula because $P(E)$ is common to both probabilities and therefore cancels out:

$$\frac{P(H|E)}{P(\overline{H}|E)} = \frac{P(E|H)P(H)}{P(E|\overline{H})P(\overline{H})}. \qquad (8.4)$$

The right side contains the factor $P(H)/P(\overline{H})$, which is the initial odds. Using O for odds,

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})}. \qquad (8.5)$$

This result is Bayes theorem (for the case of two mutually exclusive hypotheses).

In the fraction $P(E|H)/P(E|\overline{H})$, the numerator measures how well the hypothesis H explains the evidence E; the denominator measures how well the contrary hypothesis $\overline{H}$ explains the same evidence. Their ratio, known as the likelihood ratio, measures the relative value of the two hypothesis in explaining the evidence. So, Bayes theorem has the following English translation:

$$\text{updated odds} = \text{initial odds} \times \text{relative explanatory power}. \qquad (8.6)$$

Let's see how this result applies to my English telephone number. Initially I was not very sure of the phone number, so $P(H)$ is perhaps 1/2 and $O(H)$ is 1. In the likelihood ratio, the numerator $P(E|H)$ is the probability of getting a busy signal given that my guess is correct (given that H is true). If my guess is correct, I'd be dialing my own phone using my phone, so I would definitely get a busy signal: $P(E|H) = 1$. The hypothesis of a correct number is a very good explanation of the data.

The trickier estimate is $P(E|\overline{H})$: the probability of getting a busy signal given that my guess is incorrect (given that $\overline{H}$ is true). If my guess is

Are these just guesses or is there some reasoning?

How different is the British phone system that allows you to come up with a probability of 1/2? Or did you choose 1/2 because it brought about easier numbers for estimation?

How did you come up with 1/2? This seems very arbitrary, is it just a guess? I would have thought the chance might be lower, given all of the possible ways there are to misremember something.

I've been itching for an example. Hopefully this will answer my questions

I'm a little confused on why we needed P(H)?

We needed P(H) to find O(H), which is just P(H) divided by P(H_bar)

I really like this example; I was having a hard time understanding how to implement this, but this example has given me more confidence.

I agree this example shows exactly how to approximate everyday probabilities. Without this example, I wouldn't have really known how to approach this kind of problem.

incorrect, I'd be dialing a random person's phone. What is the probability that a random phone is busy? I figure it's similar to the fraction of the day that my phone is busy. In my household, the adults use the phone maybe 1 hour per day (and the children are not yet able to use a phone). So the busy fraction is perhaps the ratio of 1 hour to 24 hours, or 1/24. But that's an underestimate. At 3am I would not do the experiment – in case I am wrong and wake someone up. Equally, I am not often on the phone at 3am. A more reasonable denominator is probably 10 or 12 hours, making the busy fraction roughly 0.1. In other words, $P(E|\overline{H}) \sim 0.1$. The hypothesis of an incorrect number is not a very good explanation for the data. The relative explanatory power, which is the likelihood ratio, is

$$\frac{P(E|H)}{P(E|\overline{H})} \sim \frac{1}{0.1} = 10. \tag{8.7}$$

Therefore, the updated odds are

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})} = 10, \tag{8.8}$$

or 10-to-1 odds in favor of the number (at the start the odds were 1 to 1). The guess become very plausible!

Is it really a _random_ person's phone? It's just not yours. Maybe "random" should be used carefully in a section on probability.

I think it's fine to use the term "random." He's not defining any random variable here or getting more complicated, and randomness is inherent in probability. If you're really worried about it I guess you could use the term "arbitrary" in its place.

hmmm. You're probably (no pun intended) right. Since he has some idea of what the number is, the number dialed is biased to similar numbers.

P(E| H) assumes that it is not his phone. Given that it's not his phone, it is presumably a random phone (that necessarily exists, since we got a busy signal – i.e. it can't be an unused number).

I think the above argument is not that it will be some other phone, but rather that semantically, it's not random since the numbers involved probably restrict it to a particular geographic location (presumably near where he lived), etc. However, it is of little or no importance to the reading itself.

ah! I like this method, i was gonna say it's almost impossible to calculate this and I had no idea how to start it

I know, i was thinking about how many people might be on the phone at a given time, and how many possible phone numbers there are... and trying to figure out some probability that way.

This shows the simplicity in estimation that we sometimes lose track of with all of our formal methods. apply the problem to our own lives and it turns out we can come up w values that would be really hard to calculate

I'm confused about why the fraction of the time you use your phone is roughly the probability that you called someone else's busy number. Could you elaborate more on why this is a good approximation?

I think it's because for this part of the problem, he says to assume we've guessed incorrectly and dialed someone else's number. Therefore, if we assume all people use the phone for roughly the same fraction of the day,this fraction will equal the probability that any phone number we call will be busy.

incorrect, I'd be dialing a random person's phone. What is the probability that a random phone is busy? I figure it's similar to the fraction of the day that my phone is busy. In my household, the adults use the phone maybe 1 hour per day (and the children are not yet able to use a phone). So the busy fraction is perhaps the ratio of 1 hour to 24 hours, or 1/24. But that's an underestimate. At 3am I would not do the experiment – in case I am wrong and wake someone up. Equally, I am not often on the phone at 3am. A more reasonable denominator is probably 10 or 12 hours, making the busy fraction roughly 0.1. In other words, $P(E|\overline{H}) \sim 0.1$. The hypothesis of an incorrect number is not a very good explanation for the data. The relative explanatory power, which is the likelihood ratio, is

$$\frac{P(E|H)}{P(E|\overline{H})} \sim \frac{1}{0.1} = 10. \tag{8.7}$$

Therefore, the updated odds are

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})} = 10, \tag{8.8}$$

or 10-to-1 odds in favor of the number (at the start the odds were 1 to 1). The guess become very plausible!

I've seen Bayes' theorem used in genetics problems; its really interesting to see you estimating the probability values whereas normally the majority of the problem concerns finding these values.

i wonder how cellphone usage affects this. at my house, most long conversations are held on personal cell phones, so as to not clog up the main family line.

I don't think this assumption holds across all households. I think that a more detailed way to do this calculation is to do an estimation of single households vs. family households and then break the family household down to say, an average of 4 people- 2 adults and 2 children. For the children, it might be reasonable to assume that one is old enough to use the phone, and the other isn't, but an even more detailed calculation would involve thinking about the distribution of ages of children.

This way of going about approximating the probability is very cool! It would have been near impossible to estimate the probability using amount of people, area codes, time spent on the phone etc. Just a plain 'ol people talk about 1/24 of the day makes things so simple!

This seems like an underestimate to me, if you consider businesses who use their phones constantly.

I was going to point out this very thing

This is an extremely interesting approximation that you make because you take into account some important factors that in the middle of a problem I would have never thought about (such as people being asleep at 3am and not performing the call during that time)

i'm actually very surprised you took those factors into account. since there's so many details we usually, in this class, throw away because we're only concerned with order of magnitude approximations.

Exactly what i thought of! Except I didn't realize this is considered "the odds"

Can we please stop boxing entire paragraphs? Thanks.

This was a good job of tying everything in. It makes a lot more sense after reading this.

I don't fully follow/agree with the reasoning here- I would have looked at the time of day it was rather than trying to average it over the entire day

incorrect, I'd be dialing a random person's phone. What is the probability that a random phone is busy? I figure it's similar to the fraction of the day that my phone is busy. In my household, the adults use the phone maybe 1 hour per day (and the children are not yet able to use a phone). So the busy fraction is perhaps the ratio of 1 hour to 24 hours, or 1/24. But that's an underestimate. At 3am I would not do the experiment – in case I am wrong and wake someone up. Equally, I am not often on the phone at 3am. A more reasonable denominator is probably 10 or 12 hours, making the busy fraction roughly 0.1. In other words, $P(E|\overline{H}) \sim 0.1$. The hypothesis of an incorrect number is not a very good explanation for the data. The relative explanatory power, which is the likelihood ratio, is

$$\frac{P(E|H)}{P(E|\overline{H})} \sim \frac{1}{0.1} = 10. \tag{8.7}$$

Therefore, the updated odds are

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})} = 10, \tag{8.8}$$

or 10-to-1 odds in favor of the number (at the start the odds were 1 to 1). The guess become very plausible!

This seems awkward and too loing, but I'm not sure how to fix it.

I actually think it's helpful. I never thought to modify the guess based on when people are more likely to use their phone.

this seems like a very concise way to calculate the odds. originally I had thought about a tree diagram with number of people in a range of close area codes *land lines* usage all as the odds not in your favor. but this was well explained!

I agree! Sorry, I can't figure out how to un-question this. But it was a very clear way of figuring out the value without resorting to trees or lots of calculations!

wouldn't doing the experiment at a time like 3am make the experiment easier because not as many people are using the phone at this hour...thus raising the probability that a busy signal is yours?

I really enjoyed this section. I haven't really done anything with probabilities since 6.041 so it's nice to see this knowledge to good use.

But why didn't we take into account the # of other random people? It seems a lot more likely that you would get a busy signal not your own because there are so many other possible people you could have called.

I think this is looking at the probability of any busy signal, not just from calling your own phone.

Could you also frame this in terms of a probability? I naturally think in probability, and a 10/11 probability is much more intuitive to me than 10:1 odds.

does that give a 1/11 chance now (.09)?

yes.

No – it gives us a 10/11 probability that it is in fact his number.

The (subjective) probability that it's not his number has been reduced to 1/11.

incorrect, I'd be dialing a random person's phone. What is the probability that a random phone is busy? I figure it's similar to the fraction of the day that my phone is busy. In my household, the adults use the phone maybe 1 hour per day (and the children are not yet able to use a phone). So the busy fraction is perhaps the ratio of 1 hour to 24 hours, or 1/24. But that's an underestimate. At 3am I would not do the experiment – in case I am wrong and wake someone up. Equally, I am not often on the phone at 3am. A more reasonable denominator is probably 10 or 12 hours, making the busy fraction roughly 0.1. In other words, $P(E|\overline{H}) \sim 0.1$. The hypothesis of an incorrect number is not a very good explanation for the data. The relative explanatory power, which is the likelihood ratio, is

$$\frac{P(E|H)}{P(E|\overline{H})} \sim \frac{1}{0.1} = 10. \tag{8.7}$$

Therefore, the updated odds are

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})} = 10, \tag{8.8}$$

or 10-to-1 odds in favor of the number (at the start the odds were 1 to 1). The guess become very plausible!

Interesting - not sure how I feel about using "odds" instead of simple probabilities like the original Bayes' rule gives, but I guess it's easier for someone who has no probability experience to wrap their head around.

I have had little probability experience and it seems that it would be much easier to go directly to the probability of calling another person's busy phone and get the same answer. But I imagine this is a framework for more complex estimates.

I've never fully understood the idea of "odds" and have always favored simple probabilities as a more intuitive way to understand the certainty.

Aren't odds and probabilities the same thing said a different way. Why do you favor one?

I agree, i think mixing the two makes it a bit more confusing for someone who is used to using one or the other (I've been taught probabilities so it seems a bit uneccessary to head this route).

To me, probability has always been much more intuitive. It very clearly tells me how likely an event is based on how many times it would happen given 100 trials. I definitely don't have this intuition for odds, but maybe it's just because I've used them less? Do other people have a good intuition for different values of odds?

This stuff is really cool. I've always wanted to learn about probability and odds but never had time in my schedule for 6.041.

If you really like this stuff then you should definitely take a probability course at MIT before you graduate.

Definitely something I want to do as well, probability has always confused me a bit as I've never had a background in it but it's pretty amazing

I really enjoyed this reading as well. I think I may end up taking a probability class because of it.

This is a different way than you expressed odds before.

incorrect, I'd be dialing a random person's phone. What is the probability that a random phone is busy? I figure it's similar to the fraction of the day that my phone is busy. In my household, the adults use the phone maybe 1 hour per day (and the children are not yet able to use a phone). So the busy fraction is perhaps the ratio of 1 hour to 24 hours, or 1/24. But that's an underestimate. At 3am I would not do the experiment – in case I am wrong and wake someone up. Equally, I am not often on the phone at 3am. A more reasonable denominator is probably 10 or 12 hours, making the busy fraction roughly 0.1. In other words, $P(E|\overline{H}) \sim 0.1$. The hypothesis of an incorrect number is not a very good explanation for the data. The relative explanatory power, which is the likelihood ratio, is

$$\frac{P(E|H)}{P(E|\overline{H})} \sim \frac{1}{0.1} = 10. \qquad (8.7)$$

Therefore, the updated odds are

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})} = 10, \qquad (8.8)$$

or 10-to-1 odds in favor of the number (at the start the odds were 1 to 1). The guess become very plausible!

Can someone explain me what "10-to-1 odds" means? I usually mix this part up (and I don't trust my gut on this one).

I think 10 to 1 odds here means there are 10 times as many chances that you'd be calling your own phone (correct phone number) than calling a random phone (incorrect phone number).

thanks for explaining that, I always get confused when talking about "odds" as well!

10 to 1 means the same as 1 out of 11.

Overall a clear summary of Bayesian statistics. I've seen it once before and never used it much, and it was quite easy to follow here. I'm now kind of excited to think about day-to-day probabilities this way.

It also explains why Bayesian statistics can be so controversial - because there is (as with much of estimation) some guesswork.

did you have time to go through all this analysis as you were trying to tell your friend your phone number? just wondering

"became" or "becomes"?

or "has become"?

Good catch.

so 10:1 is "very" plausible? I wonder what the cutoff is....its a huge increase over the initial guess but 10:1 doesn't seem surefire to me!

This reading was very easy to follow and the content was very interesting.

were you right?

I felt like the initial question got convoluted in the explanation. I got a little confused on how the number being busy correlated to it actually being the right number. Maybe because I don't have much probability experience.

no, the initial question was "how certain am i that getting the busy signal means i got the right number" and the conclusion is, if i get a busy tone, i'm 10x more likely to have gotten the number right

incorrect, I'd be dialing a random person's phone. What is the probability that a random phone is busy? I figure it's similar to the fraction of the day that my phone is busy. In my household, the adults use the phone maybe 1 hour per day (and the children are not yet able to use a phone). So the busy fraction is perhaps the ratio of 1 hour to 24 hours, or 1/24. But that's an underestimate. At 3am I would not do the experiment – in case I am wrong and wake someone up. Equally, I am not often on the phone at 3am. A more reasonable denominator is probably 10 or 12 hours, making the busy fraction roughly 0.1. In other words, $P(E|\overline{H}) \sim 0.1$. The hypothesis of an incorrect number is not a very good explanation for the data. The relative explanatory power, which is the likelihood ratio, is

$$\frac{P(E|H)}{P(E|\overline{H})} \sim \frac{1}{0.1} = 10. \tag{8.7}$$

Therefore, the updated odds are

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})} = 10, \tag{8.8}$$

or 10-to-1 odds in favor of the number (at the start the odds were 1 to 1). The guess become very plausible!

i don't really understand how your level of certainty in your guess didn't play a factor at all in our calculations.

usually it doesn't, it's just kind of an afterthought

In the end, this example was not basic enough to explain what issues I have. Although I can memorize a formula, it is difficult to master one. I wanted to see easy cases as to why P(E|H)P(H) = P(H|E)P(E).

That's a good idea to include in the reading. I'll give an example in class.

i agree. i still dont feel very confident about using this.

just curious, was the number that you dialed indeed your number?

don't leave us hanging!

Agreed. Now we're all curious.

It was right! And after that I wrote it down so I wouldn't have to guess again.

I really liked this reading. I felt like I was able to follow this reading a lot better than some of the more recent ones, particularly the ones on drag.

I too found this section much easier to follow. Granted it's an introductory section, I like that it didn't require much prior knowledge to grasp the example.

Yeah I agree. I'm excited to see how we end up using this. Hopefully I'll understand this section better since the introduction was easier for me to follow.

I agree, it definitely helps when the introduction to a new section is very clear, easy to follow and interesting. I leaves the reader eager to jump into the next section!

I agree. This section was very clearly presented and the example was simple and illustrative.

I also agree - this had a good balance of math and variables and explanation .

It's also easier to get a grasp of coin tosses, probability, and phones than drag coefficients and viscosity.

incorrect, I'd be dialing a random person's phone. What is the probability that a random phone is busy? I figure it's similar to the fraction of the day that my phone is busy. In my household, the adults use the phone maybe 1 hour per day (and the children are not yet able to use a phone). So the busy fraction is perhaps the ratio of 1 hour to 24 hours, or 1/24. But that's an underestimate. At 3am I would not do the experiment – in case I am wrong and wake someone up. Equally, I am not often on the phone at 3am. A more reasonable denominator is probably 10 or 12 hours, making the busy fraction roughly 0.1. In other words, $P(E|\overline{H}) \sim 0.1$. The hypothesis of an incorrect number is not a very good explanation for the data. The relative explanatory power, which is the likelihood ratio, is

$$\frac{P(E|H)}{P(E|\overline{H})} \sim \frac{1}{0.1} = 10. \tag{8.7}$$

Therefore, the updated odds are

$$O(H|E) = O(H) \times \frac{P(E|H)}{P(E|\overline{H})} = 10, \tag{8.8}$$

or 10-to-1 odds in favor of the number (at the start the odds were 1 to 1). The guess become very plausible!

Would we get examples for a case with more than just 2 options? How would that play out.

We'll do a famous example in lecture (Monty Hall).