PHL B20 Lecture notes
Topic 4: Probability and belief

** Belief and uncertainty

We have been working with the view that belief drives out uncertainty: 'it will rain' is incompatible with 'it might not rain', and since 'it will rain' is equivalent to 'I believe it will rain', 'I believe it will rain' is incompatible with 'it might not rain'.

Uncertainty about an issue is a matter of lack of belief on either side. I am uncertain whether the number of trees is even. I don't believe that the number of trees is even, and I don't believe it isn't even. It might be even, and it might not be even.

** Uncertainty and possibilities

We have been thinking of uncertainty and belief in terms of *possibilities*. If it might rain, then there is at least some possibility that it will rain. If I believe it will rain, then from my point of view, there is no possibility it won't rain.

(Here we don't mean 'possibility' in the broadest sense: I always acknowledge it's *possible* it won't rain, in the sense that this scenario makes sense, unlike 2+2=100000000, which doesn't make sense. Rather, when I believe it will rain, I don't take those possibilities seriously.)

We use a diamond to represent possibility: 'it might rain' = '<>(it will rain)'. And we use a box to represent certainty: 'it will certainly rain' = '[ ](it will rain)'. <>P = ~[ ]~P (exercise: convince yourself of this).

Accordingly, uncertainty about P = <>P & <>~P. This is different from <>P & ~<>P (exercise: convince yourself of this).

If we think of possibilities in terms of regions of a space of situations that, for all I know, might be actual, we can think of <> as meaning (at least) *some of this space* and [ ] as meaning *all of this space*:

```
                  |============================|
                  |============================|    <-------  everywhere Q
                  |============================|
                  |============================|              = certainty
                  |============================|              = [ ]Q
                  |============================|              requires somewhere Q
                  |============================|              (namely <>Q)


left side ~P ---> |=============|=============|    <-------
                  |=============|=============|              right side P
= might not P     |=============|=============|
= <>~P            |=============|=============|              = might be that P
= somewhere,      |=============|=============|              = <> P
        ~P        |=============|=============|              = somewhere P
                  |=============|=============|
```

** Probability as a measure of certainty

If Sam ate the whole cake, we can say Sam ate 100% of the cake, and if Sam ate none of the cake, we can say Sam ate 0% of the cake.

Similarly, if Sam ate at least some of the cake, we can ask which percentage of the cake Sam ate: 45%? 100%? 1%? 99%?

Percentages are a more fine-grained and precise way of thinking about notions like some, all, and none. They make distinctions between different levels of some-but-not-all: 99% is closer to *all* than 45%, which is closer to *all* than 1%, which is in turn very close to *none*.

So since we are thinking of certainty that P as in *all* possibilites, P; and of certainty that ~P as in *no* possibilities, P; and of uncertainty about P as *some* possibility of P and *some* possibility of ~P -- we can also think of *levels of uncertainty*.

Someone might be uncertain about P in many ways: some are very close to certainty; others are very close to certainty of ~P; others are in between.

Probability is a way of measuring this level of certainty: a probability like 99% is a state very similar to certainty, while a probability like 1% is very similar to certainty-not; a probability like 45% is close to pure uncertainty (though a bit on the side of certainty-not).

** 'Subjective' probability

The probabilities we are thinking of are 'subjective': measures of personal certainty. Just as you are certain about things I know nothing about (your mother's name, for example), people might differ in their measures of personal certainty: you have been looking out the window and seen dark clouds -- you have a probability of 85% that it will rain. By contrast, my most recent news is a weather forecast from yesterday which said it would likely be sunny all day: I have a probability f 30% that it will rain.

** The mathematics of probability

A person's probability is a *function* or a *mapping*. The input is a "proposition" and the output is a percentage value.

For example, for Sam's probability that it will rain, we can write C[Sam](it will rain). If Sam is 86% certain it will rain, we say C[Sam](it will rain) = .86. (The 'C' is for 'credence' -- level of how in*cred*ible ~P is.)

Mo has a different probability -- 40% -- that it will rain: we write this C[Mo](it will rain) = .4.

A probability function obeys two rules: (1) Sam is certain that P just if C[Sam](P) = 1; (2) if in Sam's view, P and Q can't both happen, then C[Sam](P or Q) = C[Sam](P) + C[Sam](Q).

Let's try this on an example.

Sam is rolling a die: Sam thinks it is 'fair': there is a 1/6 chance of any of the numbers coming up on any given roll. Sam is certain that some number 1 through 6 will come up: C[Sam](the outcome is between 1 and 6) = 1 -- that's the first rule of probability.

Sam is certain that 3 and 4 won't both come up in a single roll: so according to the second rule, C[Sam](it's 3 or it's 4) = C[Sam](it's 3) + C[Sam](it's 4). Is this true? Intuitively, Sam thinks the probability of either 3 or 4 is 1/3 -- which is 1/6 + 1/6. This is what the second rule required: good.

We can use this second rule to calculate out Sam's probability of P or Q when P is *compatible* with Q. In general, there are three *incompatible* ways for P or Q to come about: (a) P alone (b) Q alone (c) P and Q together (when P and Q are incompatible, there is no possibility of (c), so the second rule of probability lets us ignore this factor). So under these circumstances, C(P or Q) = C(P alone) + C(Q alone) + C(P and Q together).

Let's try this on an example. Say that a value is 'low' if it is between 1 and 3. What is Sam's probability that a roll will be either low or even? Here the outcomes which are (a) low alone are {1,3}; the outcomes which are (b) even alone are {4, 6}; and the outcomes which are (c) low and even are {2}. So C[Sam](it's low or even) = C[Sam](it's 1 or 3) + C[Sam](it's 4 or 6) + C[Sam](it's 2) = 1/3 + 1/6 + 1/3 = 5/6.

Note that this value is different from C(it's low) + C(it's even). That first value is 1/2, while the second value is 1/2. Summing those, we get 1. That would be the wrong answer! There is some possibility that the roll is neither low nor even. This happens when the roll comes up 5. That has chance 1/6, so the remaining amount of chance is 5/6 -- the correct value. If we got the wrong value of 1, this would be because we ignored the fact that a roll of 2 is *both* low *and* even. We would be double-counting it if we just added up the probability of low and the probability of even.

Moral: ***when calculating probabilities, avoid double-counting!!!***

** Conditional probability

Often we are interested not so much in what the chance of P is, but in how the chance of P depends on some other assumptions. What is the chance that it will rain *assuming someone seeds the clouds?* What is the chance that it will rain *assuming a volcanic eruption sends a lot of dust in*? What is the chance that it will rain *if high winds blow the clouds away*? These may all have different answers.

We use a special notation to talk about this *conditional* probability. We write Sam's probability that P *assuming that Q* as follows: C[Sam](P/Q). Pronounce that: 'Sam's probability of P given that Q' or 'Sam's credence of P on Q' or the like.

The value of C(P/Q) is C(P&Q)/C(Q). MEMORIZE THIS EXTREMELY IMPORTANT FORMULA!

Why is this?

Considering our die, what is C(it's even/it's low)? The formula says this is C(it's even&it's low)/C(it's low). This is (1/6)/(1/2) = (1/6)*2 = 1/3.

Intuitively, the answer is 1/3: of the low values {1,2,3}, only one of those is even. There are three of them, and the die is fair; so the chance is 1/3. That is what the formula predicts. Why?

The formula squares with the intuitive understanding for the following reason. In computing the value intuitively, we did the following: ignore all the possibilities but those where it is low, and then determine which fraction of *those* are the ones in which it is even.

Now, when we are ignoring all those other possibilities, we are in effect treating all the possibilities in which the value is low as *all* the possibilities. We are treating that region as 100% of the possibilities. This is why we divide by 1/2: doing that *cancels out* the fact that the size of this region is *in fact* 1/2. (To see it another way, note that if the size of Q is very small, cancelling this fact out to treat its size as 1 requires multiplying by a very large number. How large exactly? Exactly 1/(the size of Q), because (the size of Q)*(1/(the size of Q)) is always = 1.)

Then, in order to determine which portion of the possibilities we are treating as all of them are the ones in which the value is even, we need to learn how many *of those* possibilities are ones in which it is even -- in other words, how many of the possibilites are ones in which the roll is both low and even. This is why we multiply by this value.

Another example: considering our die, what is C(it's even/it's high)? The formula says this is C(it's even&it's high)/C(it's high). This is (1/3)/(1/2) = (1/3)*2 = 2/3.

Note: ***changing Q changes C(P/Q)!!*** the second value is twice as high as the first, because out of the possibilities in which it is high, *relatively many* are possibilities in which it is even; by comparison, out of the possibilities in which it is low, *relatively few* are possibilities in which it is even.

Note: C(it's even/it's low) + C(it's even/it's high) = 2*(C(it's even/it's low or it's high)). Why? I leave this as an exercise.

Note: ***C(P/Q) is often different from C(Q/P)!!!*** Compare: C(it's not 6/it's even) = (1/3)/(1/2) = 2/3. By contrast, C(it's even/it's not 6) = (1/3)/(5/6) = 2/5. The effect is due to the following: although the numerator C(P&Q) will be the same for C(P/Q) and C(Q/P), the denominator will be either C(Q) or C(P), respectively: these values can easily differ. In the case under consideration, C(it's even) < C(it's not 6). For this reason, C(it's not 6/it's even) is *larger* than C(it's even/it's not 6).

** Bayes's Theorem

Let's switch from P and Q to E and H -- for /evidence/ and /hypothesis/.

The values of C(H/E) and C(E/H) are related as follows:

$$C(H/E) = C(E/H)*C(H)/C(E).$$

That is "Bayes's Theorem". MEMORIZE IT.

Why is it true? Well, as we saw, C(H/E) and C(E/H) share a numerator -- C(E&H) -- but differ in their denominator -- C(E) versus C(H), respectively. For this reason, we can turn the latter into the former by multiplying the latter by *its own* denominator and dividing by the *other's* denominator: in other words, by multiplying by C(H) and dividing by C(E) -- as the formula says.

Here is what the formula means, intuitively. Suppose I am considering a theory about this die: perhaps that it is a fair die. I roll it a number of times, observing the outcomes. How certain should I be that the die is fair? The formula gives three factors in answering this: (a) my confidence should go up when my theory says that the evidence is likely (when C(E/H) is high); (b) my confidence should go up when I just consider the theory to be likely in advance (when C(H) is high); (c) my confidence should get a boost when the evidence is surprising (when C(E) is low).

Why? Well, (b) is pretty clear. When C(H) is high, this means that H is a plausible hypothesis in advance of any testing. Of course this should give a theory an advantage over a hypothesis that is not plausible in advance of any testing.

And (a) is also pretty clear. If my theory makes the data into a great big surprise, then my theory does not do so well at predicting the data. What I want out of a theory is one that says 'here's what will probably happen', and then it does happen. I want my theories to make it so I won't be surprised. A theory that keeps saying 'X is really unlikely' and then X keeps happening is a lot worse than one that keeps saying 'X is really likely' and then X keeps happening.

The rationale for (c) is not quite as clear. Look at it this way though. Imagine if it weren't there. Then term (a) would do no work! This is because we could give extremely vague characterizations of the data. *Every* theory predicts 'something or other will happen'. That does indeed keep happening! So every theory gets the highest possible score on C(E/H), term (a), by predicting this. If we just liked theories to the extent that the value of (a)*(b) was high, we would like theories to the extent that 1*(b) is high: initial plausibility would mean everything, prediction would mean nothing.

Dividing through by C(E) lets us avoid this problem. It gives bonuses to theories that predict *precisely specified* data over theories that get the data only *approximately right*. In the extreme case, C(something or other will happen) = 1: that is the vaguest possible characterization of the data, and predicting it is always less valuable than saying 'the readout will indicate 98.58498789' -- if that in fact happens. That outcome is precise, and in that sense surprising when it happens. If a theory predicts this precise outcome in term (a), then it gets a big bonus in term (c) as well.

** Bayes's Theorem in the real world

The murderer has rare blood type ABCD: only 1 person in 1000 has it. The Crown has presented evidence that Fred has blood type ABCD. Looks pretty bad for Fred!

You are Fred's defense attorney. What do you say?

Bayes's Theorem to the rescue. Let H = Fred is the murderer; let E = Fred has blood type ABCD. What is needed to convict is a high value of C(H/E): the hypothesis needs to be likely (to the reasonable person), given the evidence presented by the Crown.

Bayes's theorem says that this value is equal to:

C(E/H)*C(H)/C(E).

What value does this have?

Well, C(E/H) is 1. We know the murderer has blood type ABCD; so given that Fred is the murderer, we can be 100% confident that Fred has blood type ABCD.

And we might think that C(E) is very low. This is a bit hard to evaluate, but think of it this way: not knowing anything in advance about Fred's blood type, how likely would we think it is that he has type ABCD? Since only 1 person in 1000 has it, I suppose the chance of Fred being one of those people is 1 in 1000.

So far so bad: high factor (a), low factor (c).

But what is factor (b)? Suppose there are 1 million men in town. And suppose that the method of the local police department in the face of a murder is to grab some guy at random off the street and railroad him. Then how likely is it up front that Fred is the murderer? Well, only 1 person in 1 million is that person: up front, not knowing anything else, we would say the police have 1 chance in a million of getting the murderer by their method. Since Fred is the person they got, we should say that Fred has 1 chance in 1 million of being the murderer.

So very low factor (b). Multiplying (a) times (b) divided by (c) gives us 1 in 1000. Fred goes free!

** Bayesianism

Bayesianism is a theory about *learning*: about what is required as a response to evidence if one is not to count as *unreasonable*.

It says the following. Suppose my probability that H on the assumption that E is 19%. And then suppose I learn that E is true. My probability that H should then be ... 19%!

This might seem to have no substance whatsoever to it. But that is actually an illusion. Suppose that on Monday, my credence that H is 90%. But: I also think that if E happens, H would turn out to be a lot less likely -- only 19% likely. (Most of E -- 81%, to be precise -- is in that last 10% of probability that ~H. Exercise: suppose in addition that if ~H, E is certain. How likely do I think E is on Monday?) And then, on Tuesday, I learn that E. In this case, my credence that H should drop way down: all the way down to 19%.

Think about what Bayesianism says as the following:

```
                    |=============|=============|
                    |=============|=============|
top: ~E             |=============|=============|
                    |=============|_______________|
                    |_______________|=============|
                    |=============|=============|      bottom: E
                    |=============|=============|
                    |=============|=============|

                    left side: H          right: ~H
```

This is a picture of my credences on Monday. On Monday, I think H and ~H are equally likely; I think that ~E is a bit more likely than E; and of the four combinations, C(~E&H) > C(~E&~H) = C(E&~H) > C(E&H) -- going clockwise from the top right, we start with the most likely and end with the least likely.

Suppose that on Tuesday, I learn that E. Then, says Bayesianism, here is all I should do: ignore all possibilities in which ~E. I shouldn't do anything else: no internal pushing around of possibilities is required. Just wipe away the top part of the chart and I will be fine; do anything else and I will not be fine.

So on Tuesday my credences should look like this:

```
                     _______________|=============|
                    |=============|=============|      all: E
                    |=============|=============|
                    |=============|=============|

                    left side: H          right: ~H
```

Notice the following: (i) there are no more possibilites in which ~E; (ii) among the possibilities in which E, no redistribution has gone on -- the ratios within that region of ~H to H remain the same; (iii) now, as a result, H and ~H have gone from being equally likely to being differentially likely: I now consider H to be less likely than ~H.

That is what Bayesianism says you should do when you get evidence. The terminology for doing this is *conditionalizing*. The terminology is chosen for the following reason: if on Monday, my *conditional* credence of H on E is x; and on Tuesday I learn that E; my *unconditional* credence of H should now be x. New (unconditional) credence is reset to what old conditional credence is: credence is *conditionalized*.

** David Lewis's insurance policy

Lewis argues that (a) if you are inclined to do anything other than conditionalize you are inclined to get crazy with your insurance policies; (b) if you are inclined to get crazy with your insurance policies you are irrational; so (c) if you are inclined to do anything other than conditionalize you are irrational.

Here is a simplified version of how Lewis's argument for (a) works.

Suppose Bill is going to try a barrel roll on Tuesday just if it doesn't rain. On Monday, Bill buys an insurance policy for $1000 with the following provisos: (i) if it rains, he gets his $1000 back and the policy is cancelled out; (ii) if it doesn't rain and the barrel roll fails, he gets $5000 back; (iii) if it doesn't rain and the barrel roll succeeds, he gets nothing back.

This would be a good insurance policy to buy if Bill's Monday credences look like this:

```
                    left top: succeed                    right top: fail

                                  80%                20%
                    |======================|====|
                    |======================|====|
top: ~rain          |======================|====|
                    |======================|====|
                    |======================|====|
                    |_________________________|_____|
                    |==========================|        bottom: rain
                    |==========================|
```

So suppose they do (it doesn't matter what the likelihood of rain is).

Now, suppose Tuesday rolls around, and Bill learns that it does not rain. The Bayesian says Bill's credences should now look like this:

```
                    left: succeed                        right: fail

                                  80%                20%
                    |======================|====|
                    |======================|====|
all: ~rain          |======================|====|
                    |======================|====|
                    |======================|====|
                    |_________________________|_____|
```

If they do, we can ask: how much is Bill's insurance policy now worth to him? Answer, by simple cost-benefit calculation (multiply the payoff in each case times the likelihood of the case and then add all these figures up): 80%*$0 + 20%*($5000) = $1000 -- the amount he paid for it. Clearly in this case he should hold on to the policy.

But suppose that instead, Bill gins up his confidence for the barrel roll. He has been reading the self-help books that say you gotta believe in yourself. So he decides to be certain he will succeed. He ignores the possibilities of failure. In that case, his credences look like this:

```
                    all: succeed

                                  100%
                    |======================|
                    |======================|
all: ~rain          |======================|
                    |======================|
                    |======================|
                    |_________________________|
```

Now how much is Bill's insurance policy worth to him? Answer: 100%($0) = $0.

At this point, Bill's insurance agent calls up and says 'how bout we cancel that thing out and I buy you a drink after you succeed?' What will Bill do?

He will say yes: a drink in the future is worth more than $0, which is the value of the policy to him now.

That completes stage (a) of David Lewis's argument. Bill has done something different from conditionalizing, and as a result is likely to grossly undervalue an insurance policy he bought.

Now to stage (b). Remember: BILL HASN'T TRIED HIS BARREL ROLL YET. He has no basis for being certain that he will succeed! His only evidence is that it is not raining. By his earlier lights, if so, his chance of failure is 20%. He got no further evidence to the contrary: he just decided to ignore the chance of failure. His evidentiary basis is not sufficient for certainty that he will succeed. So it would seem as if he should hold on to the insurance policy! Certainly trading it for a drink at this point is a choice his earlier self would completely disapprove of. And it completely makes a hash of the entire point of having bought the insurance policy in the first place.

We could imagine a canny insurance agent who has been manipulating Bill the whole time: he saw him reading the self-help books, knew about his desire to perform a barrel roll, etc. Bill turned out to be an easy mark.

This might suggest that Bill is irrational.

This is only a simplified version of David Lewis's argument for conditionalization. I only covered a single way to not conditionalize. Any time one's credence of H after learning E ends up either higher or lower than one's earlier conditional credence of H on E, something more or less analogous can be done to rip one off -- so Lewis argues (we'll skip the details).

** Inference to the best explanation and the Bayesian

How shall we think about IBE from the perspective of subjective probability?

Consider Jo, who hears little noises in the wall. Jo considers two possibilities as explanations of the noise: mice are making the noise; the CIA is making the noise. Both of those are coherent. But the mice explanation strikes Jo as better (in whatever sense): so Jo 'infers to', or comes to believe, that explanation -- comes to believe that there are mice in the wall.

Perhaps what this means, in probabilistic terms, is the following.

(A) Both 'it's mice' and 'it's the CIA' are coherent; and evidently Jo considers them both worth taking seriously (since Jo *does* take them seriously); and evidently Jo considers them all compatible with 'there are little noises in the wall'. So it seems as if Jo has some positive credence in (there are noises and the CIA is causing them).

(B) When Jo infers to the best explanation -- comes to believe that mice are causing the noises -- the result is a *belief*: Jo becomes certain that the mice are causing the noises. If so, C[Jo](it's the CIA) goes to zero.

Now notice that if this is right, then IBE is not a kind of conditionalization. To see this, consider the following pictures:

```
                 left top: mice            right top: CIA

                               80%                 20%
                 |======================|====|
                 |======================|====|
top: noises      |======================|====|
                 |======================|====|
                 |======================|====|
                 |__________________________|_____|
                 |==========================|       bottom: ~noises
                 |==========================|
```

That is Jo's system of credences prior to hearing the noises. (Notice that the existence of the right top region -- noises & CIA -- with a positive size is supported by the reasoning in (A), above.)

If after hearing the noises, Jo conditionalizes, Jo's system of credences would look like this:

```
                 left top: mice            right top: CIA

                               80%                 20%
                 |======================|====|
                 |======================|====|
all: noises      |======================|====|
                 |======================|====|
                 |======================|====|
                 |__________________________|_____|
```

But if Jo instead performs IBE, the reasoning in (B), above, says that Jo's credences will look like this:

```
                 all: mice

                               100%
                 |======================|
                 |======================|
all: noises      |======================|
                 |======================|
                 |======================|
                 |__________________________|
```

At this point we can advance David Lewis's style of argument to the effect that this is irrational: suppose that Jo paid $1000 for an insurance policy to go into effect just if there are noises, to pay off $5000 just if the CIA is in the wall; prior to hearing the noises this looks like a good policy, worth $1000 if there are noises; after making the IBE it looks worthless, despite Jo's having not acquired any further evidence beyond that to determine the policy will go into effect; we can imagine a schemy insurance agent who foresees this and connives to cheat Jo out of $1000 ...

That is an argument that IBE is irrational. But earlier we argued that we engage in IBE all the time! So this means we are always irrational????? What to say about this is a puzzle: more in Topic 6.