Introduction to Mathematical Philosophy

Hannes Leitgeb, Stephan Hartmann

June 18, 2015

Contents

Week 5: Overview

Overview of Lecture 5: Confirmation

5.1 Introduction: Confirmation is a relation between a hypothesis (or theory) and a piece of evidence. We give a number of examples and outline various problems (such as the problem of underdetermination of theory by data) that help us sharpen the concept. Finally, we provide reasons for a quantitative (rather than a purely qualitative) explication of confirmation.

5.2 The Monty Hall Problem I: We introduce the basic ideas of probabilistic reasoning using the Monty Hall Problem as an example. This example illustrates that it is important to carefully think about what the learned evidence is. Different formulations of the evidence lead to different posterior probabilities. But presumably only one of them can be right. Which?

5.3 The Monty Hall Problem II: To settle this question, we run a computer simulation from which we conclude that one should condition on the logically strongest available evidence. This is the Principle of Total Evidence.

5.4 Confirmation: We introduce Bayesian Confirmation Theory which explicates confirmation as probability raising. A piece of evidence confirms a hypothesis if the probability of the hypothesis increases after learning the evidence. We illustrate this idea with the help of the Monty Hall Problem.

5.5 Final Remarks: We outline further applications of Bayesian Confirmation Theory in epistemology and philosophy of science and stress its close links to the psychology of reasoning.

Chapter 5

Week 5: Confirmation

5.1 Introduction (10:01)

Welcome to Lecture 5 of our Introduction to Mathematical Philosophy! You may remember that in Lecture 3 of this course, Hannes Leitgeb analyzed the concept of belief. This lecture focuses on a related concept – confirmation. Generally speaking, confirmation is a relation between a hypothesis, a piece of evidence, and a body of background knowledge. We would like to find conditions under which evidence confirms a hypothesis relative to background knowledge. This is important for our understanding of science, where empirical evidence is used, for example, to support claims about objects which are not directly accessible. Radio waves emitted by astronomical objects, for instance, are evidence for hypotheses about the composition of these objects.

But confirmation is also an issue in more mundane contexts. Take the way evidence is used in a court room: Here the prosecutor will present various pieces of evidence which she takes to support the hypothesis that the suspect is the murderer. She might, for example, point to a bloody knife, on it the finger prints of the suspect; Or she may cite from reports about an alleged affair between the suspect and the victim's wife. Similarly, think about the role of confirmation in medicine. X-ray scans, for instance, provide evidence that a patient has tuberculosis. As philosophers, we ask: what do all these instances have in common? Or: What, in general, is the relationship between a hypothesis H, a piece of evidence E, and a body of background knowledge K?

In our discussion, we will not explicitly mention the background knowledge. However, it is important to note that it is always there. We will make it explicit only when needed.

Let us start with two related worries. First, even if the evidence confirms the hypothesis, the hypothesis may nevertheless be false. The knife presented by the prosecutor may be evidence that the suspect committed the murder, while it is still possible that he is innocent. Maybe the knife also has other finger prints on it, and as it happens that another person committed the murder. It is also well known that medical tests are not fully reliable. Typically, there are false positives and false negatives. False negatives are test results which do not indicate a disease although the patient has it. False positives are test results which do indicate the disease although the patient does not have it. The rates of false positives and false negatives characterize the reliability of the device. Clearly, the goal is to minimize these (and to increase the reliability of the device). But false negatives and false positives usually can't be eliminated completely. Hence, confirming evidence is fallible: we have to be aware of the possibility that it misleads us.

Relatedly, but more radically, a possibly infinite number of alternative candidate hypotheses remains logically compatible with the available body of evidence, and none of them can be excluded on the basis of the evidence. To illustrate this point, let us assume that we investigate the relationship between two variables x and y and that we have collected three data points (see Figure 5.1).

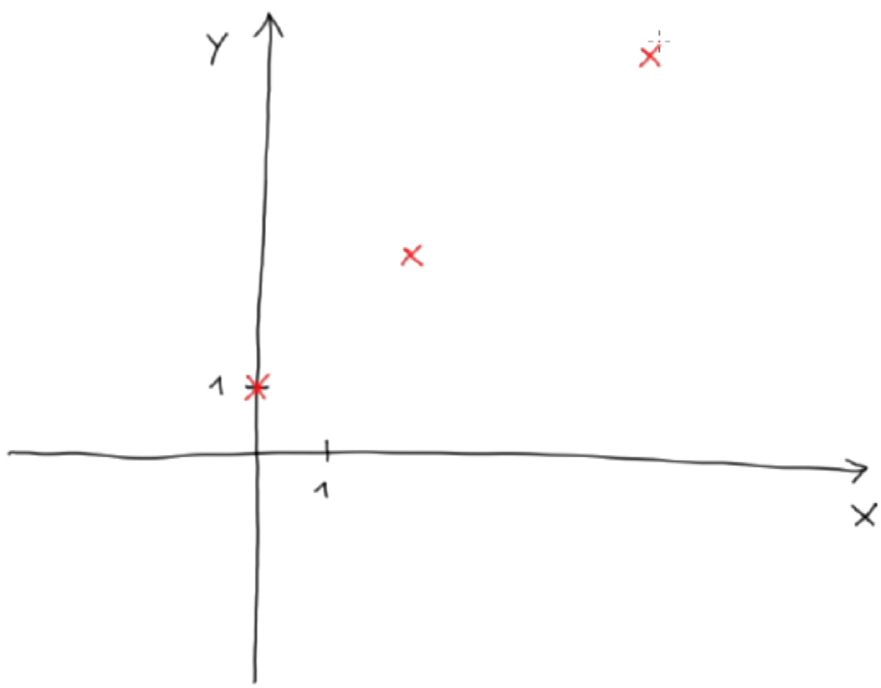

Figure 5.1: Three data points

The data are compatible with the hypothesis that x and y are related in a linear way. In fact, this hypothesis (let us call it H1) entails the data (see Figure 5.2).

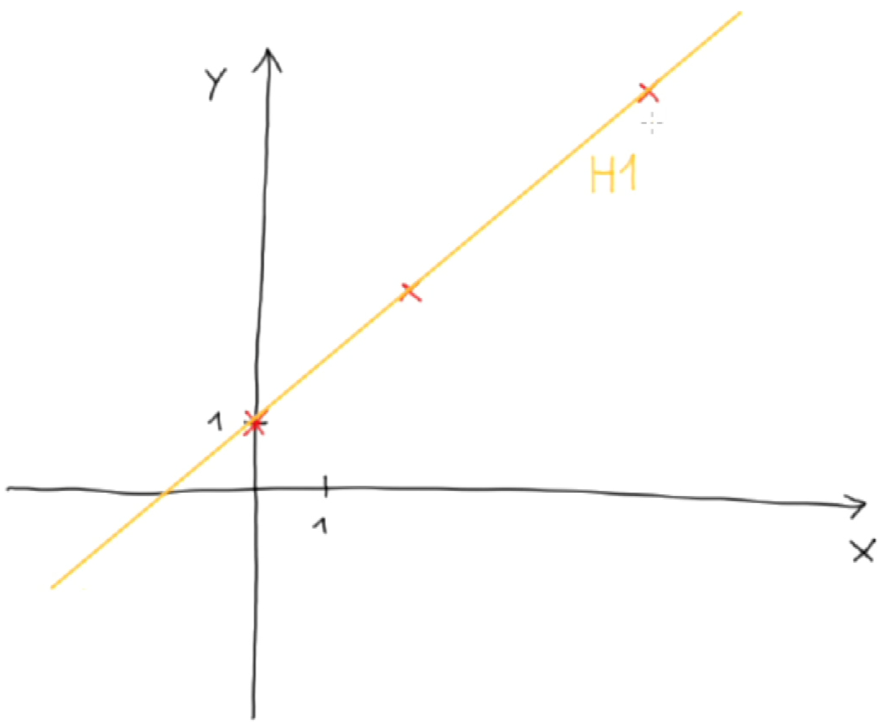

Figure 5.2: H1

But the data are also compatible with alternative hypotheses H2, H3, ... all of which entail the data, and all of which are apparently confirmed by the data (see Figure 5.3).

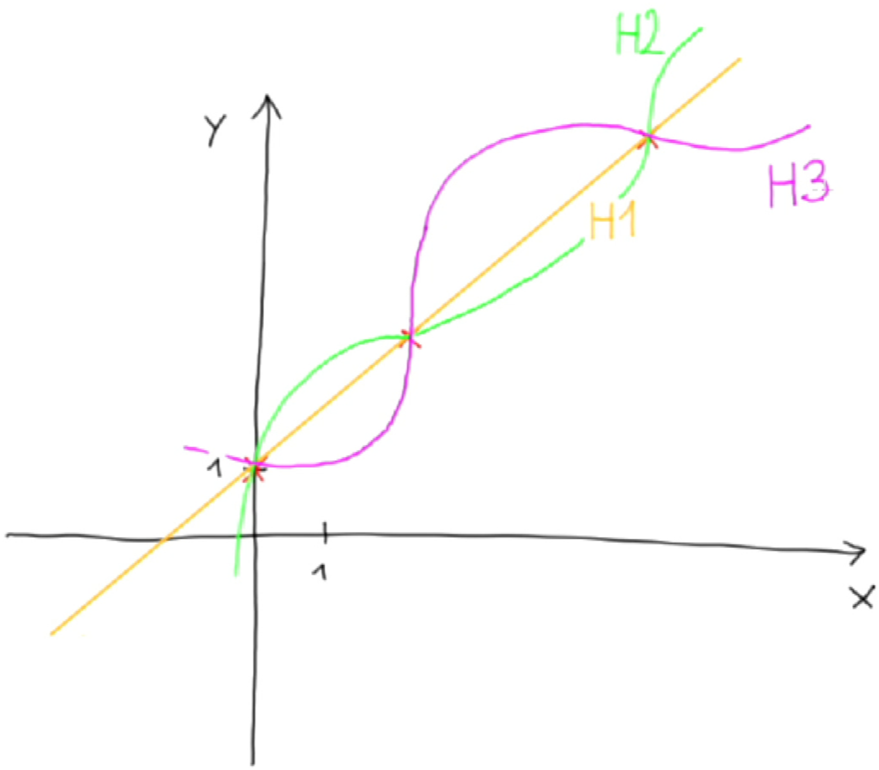

Figure 5.3: H2, H3

This is the problem of underdetermination of theory by data.

For more about the underdetermination thesis and its implications, read http://plato.stanford.edu/entries/scientific-underdetermination/.

What lessons shall we draw from this example? Clearly, if we explicate confirmation as a logical relation between a hypothesis and a piece of evidence, we have to accept that the evidence may confirm many different theories. Note that this way of putting it presupposes a qualitative concept of confirmation. E either confirms H, or it does not. Introducing a quantitative concept of confirmation solves this problem. In this case we can argue that H1 is supported by E much more strongly than H2. A reason for this might be that H1 is the simplest of the available hypotheses. If we argue like this, we introduce an additional aspect to the confirmation relation which is independent of the logical relation between the hypothesis and the evidence. Remember: All that deductive logic tells us is that E is implied by H1 and by H2. So here we have a first reason to prefer quantitative concepts of confirmation over qualitative ones.

More importantly, quantitative theories of confirmation are also more compatible with our actual ordinary and scientific reasoning. We often claim, for example, that some piece of evidence is stronger than another piece of evidence; Or we think that two pieces of evidence together provide stronger confirmation than just one. For instance, being presented with BOTH the bloody knife AND reports about the suspect's affair provides stronger support for the hypothesis that the suspect is the murderer than only one of these pieces of evidence would. Learning further pieces of evidence increases our conviction that the suspect is the murderer. This is psychologically realistic, and it also seems to make sense on normative grounds.

Quantitative theories of confirmation capture this insight. The most prominent quantitative theory of confirmation is Bayesian confirmation theory, which we will now introduce. To start with, consider the following problem:

Suppose you are on a game show, and you are given the choice of three doors: Behind one door is a car; behind the others, goats. You'd quite like to have the car, and have no use for goats (Alas, you have a garage, but no garden.). You pick a door, say No. 1, and the host (let's call him Monty), who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" What should you do? Is it to your advantage to switch doors?

Think about it for a moment and decide what to do.

Quiz 40:
Before we proceed, let us reflect upon the assignment of initial probabilities for H_1, H_2 and H_3. Our argument for assigning the same probability to H_1, H_2 and H_3 was something like this: If there is "no known reason for predicating for our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability." (J. M. Keynes, A Treatise on Probability, p. 42) Keynes called this rule the Principle of Indifference (ibid., p. 41). Unfortunately, this principle may lead to contradictory conclusions. Consider the following case. If there is no known reason for saying that my friend's new sports car is yellow rather than red, then, relative to this knowledge, it is equally probable that the sports car is yellow and that the sports car is red. We conclude that the probability of "The sports car is yellow" is 1/2 and that the probability of "The sports car is red" is 1/2. Similarly, we conclude that the probability of each of the propositions "The sports car is yellow" and "The sports car is orange" is 1/2. Show that this leads to a contradiction.
Solution

5.2 The Monty Hall Problem I (17:43)

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick door No. 1, and the host, Monty, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?

What should you do? Is it to your advantage to switch doors?"

To answer Monty's question, we distinguish three steps.

Step 1: (See Figure 5.4.)

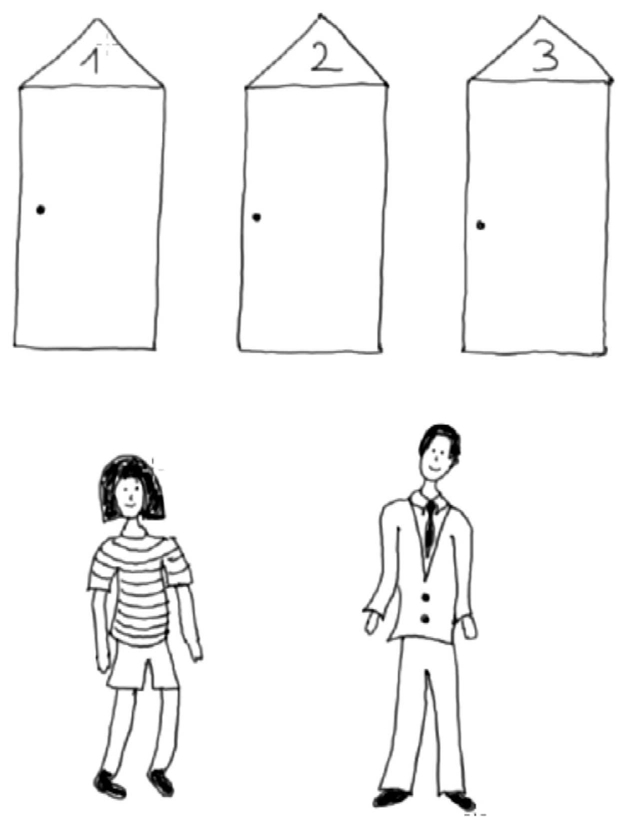

Figure 5.4: Monty explains the situation.

In the first step, Monty explains the situation and asks you to pick a door. You deliberate as follows: There is no reason to choose one door over the other, and so all doors are equally probable. That is, you assign a probability of 1/3 to each of the following hypotheses:

(Slide 1)

Step 1: Reason and Decide

What do we know initially?

(H_1) The car is behind door No. 1.

(H_2) The car is behind door No. 2.

(H_3) The car is behind door No. 3.

These three hypotheses are mutually exclusive and exhaustive, i.e. exactly one of them is true. You just do not know which, and this is why your initial probability assignment is

(Slide 2)

The three hypotheses are equally probable:

$P(\mathrm{H}_1) = P(\mathrm{H}_2) = P(\mathrm{H}_3) = 1/3$

Note that

$P(\mathrm{H}_1 \vee \mathrm{H}_2 \vee \mathrm{H}_3) = P(\mathrm{H}_1) + P(\mathrm{H}_2) + P(\mathrm{H}_3) = 1$

because the car is exactly behind one of the doors.

(Slide 3)

The three hypotheses are mutually exclusive and exhaustive:

$$P(\mathrm{H}_1 \vee \mathrm{H}_2 \vee \mathrm{H}_3) = P(\mathrm{H}_1) + P(\mathrm{H}_2) + P(\mathrm{H}_3) = 1$$

As all three hypotheses are equally likely, you pick a door at random. This means that if you were to play the game for a large number of times, you would pick door No. 1 in roughly 1/3 of the cases, door No. 2 in roughly 1/3 of the cases, and door No. 3 in roughly 1/3 of the cases. This time, you pick door No. 1 (see Figure 5.5).

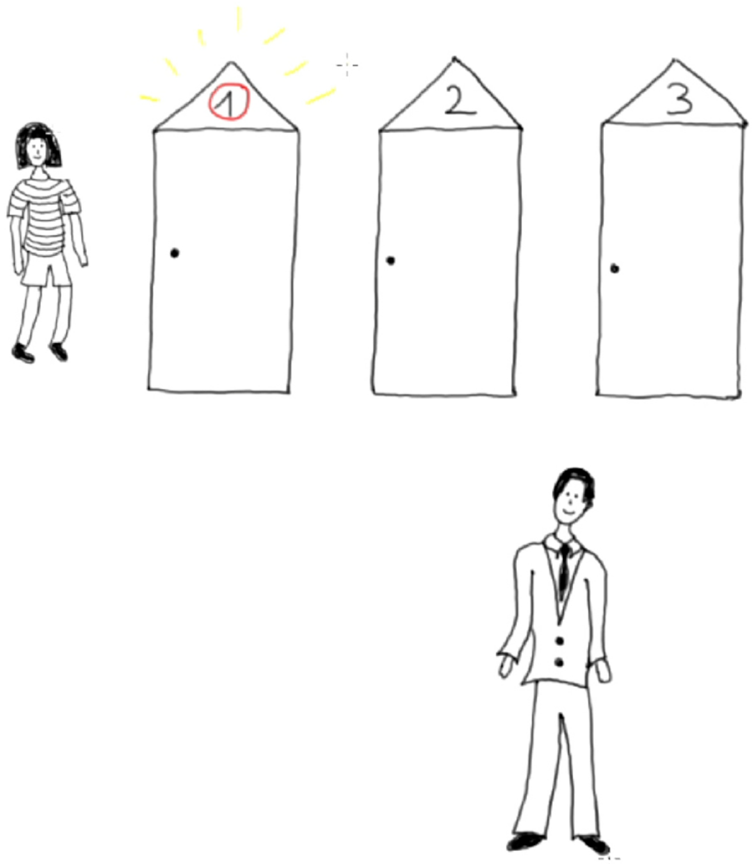

Figure 5.5: Decision for door No. 1

(Slide 4)

Step 2: Learn

What is it, precisely, that we learn?

Next, you learn that Monty opens door No. 3 and that, therefore, the car is not behind door No. 3. This is because the rules of the game only allow Monty to open a door with a goat behind it (see Figure 5.6).

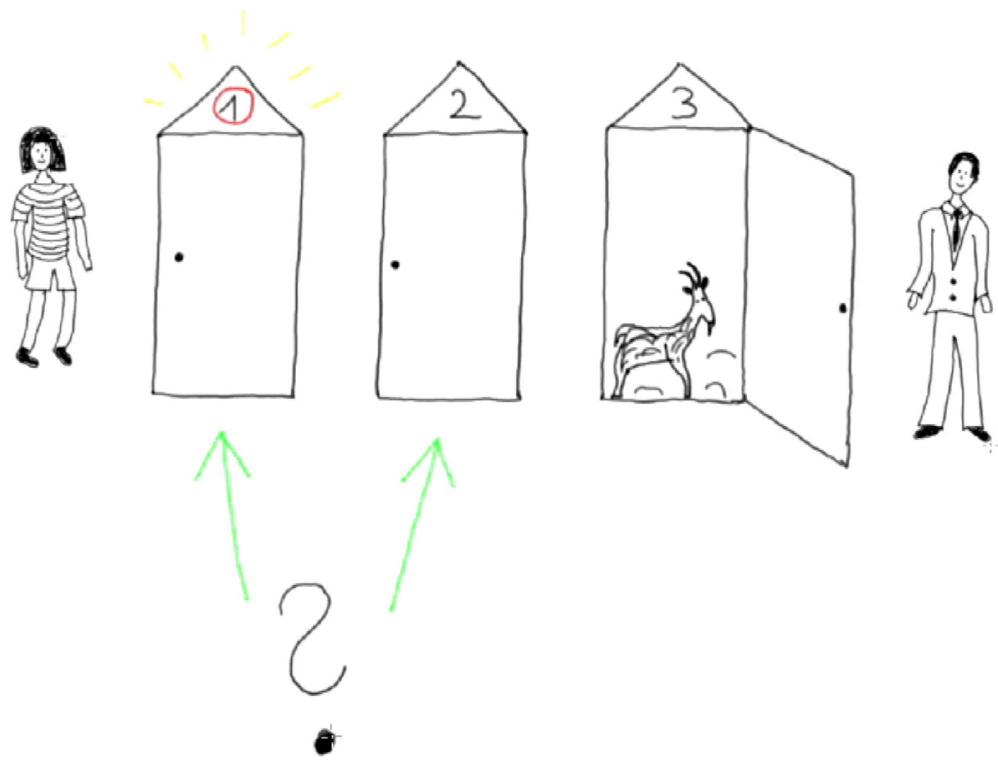

Figure 5.6: Monty opens door No. 3.

(Slide 5)

Step 3: Reason and Decide

Shall we switch doors?

The new information prompts you to change your beliefs. But how? Here are two ways of how you might reason.

Reasoning 1: Initially there were three possibilities for the location of the car: It is either behind door No. 1, or behind door No. 2, or behind door No. 3, and there is no reason to choose one door over the other. All possibilities are equally probable. Now you learn that the car is not behind door No. 3, i.e. you learn that it must be behind door No. 1 or door No. 2. And so there are only two possibilities left. Again there is no reason to choose one door over the other. Both possibilities are equally probable. Hence, switching does not increase the probability of winning the car. And so you decide not to switch.

Reasoning 2: The car is behind one of the three doors. If the car is behind door No. 1, Monty can open either door No. 2 or door No. 3. As you picked door No. 1, you should not switch. If the car is behind door No. 2, Monty will open door No. 3. This is the only choice he has, because he has to open a door with a goat behind it, and he is not allowed to open the door that you picked. Hence, you should switch to door No. 2. If the car is behind door No. 3, Monty will open door No. 2. Now, similarly, this is the only choice he has. Hence, you should switch to door No. 3. Summing up, switching is the right strategy in two out of three cases. Of course, you don't know which case you are in. But as all three cases are equally probable, switching increases the probability of winning the car.

Let us take stock. Both ways of reasoning sound plausible. However, only one of them can be the right one. But which one? Before addressing this question, I would like to make both ways of reasoning more precise and model both learning processes formally.

(Slide 6)

Bayes Theorem:

$$P_E(\mathrm{H}) = P(\mathrm{H}|\mathrm{E})$$

This is done by applying Bayes' Theorem, which tells us that the new probability $P_E(\mathrm{H})$ of a hypothesis H after having learned the evidence E is given by

$$P_E(\mathrm{H}) = P(\mathrm{H}|\mathrm{E})$$

Remember that the conditional probability $P(\mathrm{H}|\mathrm{E})$ is defined as follows:

(Slide 7)

The definition of conditional probability:

$$P(\mathrm{H}|\mathrm{E}) := \frac{P(\mathrm{H},\mathrm{E})}{P(\mathrm{E})} \text{ if } P(\mathrm{E}) > 0 (*)$$

$P(\mathrm{H},\mathrm{E})$ denotes the probability of the conjunction of the propositions H and E. Let us now calculate $P(\mathrm{E}|\mathrm{H})$ and use the symmetry of the conjunction:

(Slide 8)

The definition of conditional probability:

$$\begin{aligned} P(\mathrm{H}|\mathrm{E}) &:= \frac{P(\mathrm{H},\mathrm{E})}{P(\mathrm{E})} \text{ if } P(\mathrm{E}) > 0 \quad (*) \\ P(\mathrm{E}|\mathrm{H}) &:= \frac{P(\mathrm{E},\mathrm{H})}{P(\mathrm{H})} = \frac{P(\mathrm{H},\mathrm{E})}{P(\mathrm{H})} \text{ if } P(\mathrm{H}) > 0 \quad (**) \end{aligned}$$

Note: We use the notation $P(\mathrm{H},\mathrm{E}) = P(\mathrm{H} \wedge \mathrm{E})$ for the probability of the conjunction of two propositions.

Hence, comparing equation (*) with equation (**), we obtain:

(Slide 9)

Bayes Rule:

$$P(\mathrm{H}|\mathrm{E}) = \frac{P(\mathrm{E}|\mathrm{H})\, P(\mathrm{H})}{P(\mathrm{E})}$$

$P(\mathrm{H})$: prior probability of H

$P(\mathrm{E}|\mathrm{H})$: likelihood of E

$P(\mathrm{E})$: expectedness of E

$P(\mathrm{H}|\mathrm{E})$: posterior probability of H

This is Bayes' Rule. $P(\mathrm{H})$ is called the prior probability of H, $P(\mathrm{E}|\mathrm{H})$ is called the likelihood of the evidence given the hypothesis, and $P(\mathrm{E})$ is called the expectedness of the evidence. The prior probability is the probability of the hypothesis H before learning the evidence E. The likelihood measures how probable the evidence E is, given that the hypothesis is true. The expectedness of the evidence measures how much we expect the evidence to obtain before it actually obtains. $P(\mathrm{H}|\mathrm{E})$ is the posterior probability of the hypothesis, i.e. the probability of the hypothesis after having learned the evidence E.

To calculate $P(\mathrm{E})$, we apply the rule of total probability, i.e.

(Slide 10)

From the Rule of Total Probability and the definition of conditional probability, we obtain for the expectedness of E:

$$\begin{aligned} P(\mathrm{E}) &= P(\mathrm{E},\mathrm{H}) + P(\mathrm{E},\neg\mathrm{H}) \\ &= P(\mathrm{E}|\mathrm{H})\,P(\mathrm{H}) + P(\mathrm{E}|\neg\mathrm{H})\,P(\neg\mathrm{H}) \end{aligned}$$

If there are more than two mutually exclusive and exhaustive propositions H1, H2, ..., Hn, then this rule generalizes to

(Slide 11)

If $\mathrm{H}_1, \mathrm{H}_2, \ldots, \mathrm{H}_\mathrm{n}$ are mutually exclusive and exhaustive, then

$$\begin{aligned} P(\mathrm{E}) &= P(\mathrm{E}|\mathrm{H}_1)\,P(\mathrm{H}_1) + \cdots + P(\mathrm{E}|\mathrm{H}_\mathrm{n})\,P(\mathrm{H}_\mathrm{n}) \\ &= \sum_{i=1}^{n} P(\mathrm{E}|\mathrm{H}_\mathrm{i})\,P(\mathrm{H}_\mathrm{i}). \end{aligned}$$

Let us now apply this rule to the first way of reasoning discussed above. We learn:

(Slide 12)

Reasoning 1:

E: The car is not behind door No. 3.

E: The car is not behind door No. 3.

That is,

(Slide 13)

$\mathrm{E} = \neg\mathrm{H}_3 = \mathrm{H}_1 \vee \mathrm{H}_2$

We first calculate the likelihoods and obtain:

(Slide 14)

$P(\mathrm{E}|\mathrm{H}_1) = 1$

$P(\mathrm{E}|\mathrm{H}_2) = 1$

$P(\mathrm{E}|\mathrm{H}_3) = 0$

Here is how we get these assignments. If the car is behind door No. 1, then the probability that it is behind door No. 1 or behind door No. 2 is equal to 1. If the car is behind door No. 2, then the probability that it is behind door No. 1 or behind door No. 2 is also equal to 1. If the car is behind door 3, then the probability that it is behind door No. 1 or behind door No. 2 is equal to 0.

With these likelihoods and the prior probabilities $P(\mathrm{H}_1) = P(\mathrm{H}_2) = P(\mathrm{H}_3) = 1/3$ we first calculate:

(Slide 16)

$$P(\mathrm{E}) = 1 \cdot (1/3) + 1 \cdot (1/3) = 2/3$$

This is a plausible result, because the probability to find a goat (i.e. no car) behind one of the three doors (including door No. 3) is 2/3 as there are two goats and only one car.

As the likelihoods for H_1 and H_2 are the same (viz. equal to 1), the posterior probabilities must also be the same, i.e.

(Slide 17)

We obtain:

$$\begin{aligned} P_E(\mathrm{H}_1) = P_E(\mathrm{H}_2) &= \frac{1 \cdot 1/3}{2/3} \\ &= 1/2 \end{aligned}$$

And so we find indeed that the new probability of H_1 and the new probability of H_2 are equal to 1/2. As both are the same, switching is not to your advantage. It does not increase the probability of winning the car.

Let us now turn to the second way of reasoning. Here it is important to note that we now learn a different piece of evidence, viz.

(Slide 18)

Reasoning 2:

F: Monty opens door 3.

And we have to condition on F to calculate the new probabilities of H_1, H_2, and H_3. Again, we first calculate the likelihoods. Here they are:

(Slide 19)

$P(\mathrm{F}|\mathrm{H}_1) = 1/2$

$P(\mathrm{F}|\mathrm{H}_2) = 1$

$P(\mathrm{F}|\mathrm{H}_3) = 0$

Here is how we get these numbers. If the car is behind door No. 1, Monty can either open door No. 2 or door No. 3, because there is a goat behind both of them. He will apply a chance mechanism such as flipping a coin and pick door No. 3 with a probability of 1/2. If the car is behind door No. 2, Monty has to open door No. 3 as he cannot open door No. 1 (because you picked it) or door No. 2 (because the car is there). Hence $P(\mathrm{F}|\mathrm{H}_2) = 1$. Finally, $P(\mathrm{F}|\mathrm{H}_3) = 0$, because Monty is not allowed to open door No. 3, as this is where the car is.

We can now calculate the probability of F and obtain 1/2. That you assign a value of 1/2 to the expectancy of F makes sense. All you know is that Monty cannot open door No. 1, as this is the door that you picked. And so he can either open door No. 2 or door No. 3. You have no additional information that tells you which of them he will open, and so you assign $P(\mathrm{F}) = 1/2$. By contrast, if the car is behind door No. 2 or door No. 3, Monty does not have a choice: He must open door No. 3 or door No. 2 respectively.

Note that because of this, the likelihood $P(\mathrm{F}|\mathrm{H}_1)$ is now different from $P(\mathrm{F}|\mathrm{H}_2)$. $P(\mathrm{F}|\mathrm{H}_1)$ is half as big as $P(\mathrm{F}|\mathrm{H}_2)$. Hence the posterior probability of H_1 is half as big as the posterior probability of H_2. Note further that the probability of H_3 given the evidence is 0 and that the probabilities of all three posteriors sum to 1. Hence,

(Slide 20)

$$P_F(\mathrm{H}_1) = 1/3 \text{ and } P_F(\mathrm{H}_2) = 2/3$$

Hence, it is twice as likely that the car is behind door No. 2 than that it is behind door No. 1 (which you originally picked). So you should switch if you want to maximize your chance to win the car.

We would also have reached this result by calculating

(Slide 21)

Reasoning 2:

F: Monty opens door 3.

$P(\mathrm{F}|\mathrm{H}_1) = 1/2$

$P(\mathrm{F}|\mathrm{H}_2) = 1$

$P(\text{F}|\text{H}_3) = 0$

$$P_F(\text{H}_1) = 1/3 \text{ and } P_F(\text{H}_2) = 2/3$$

$$\begin{aligned} P(\text{F}) &= P(\text{F}|\text{H}_1)P(\text{H}_1) + P(\text{F}|\text{H}_2)P(\text{H}_2) + P(\text{F}|\text{H}_3)P(\text{H}_3) \\ &= (1/2) \cdot (1/3) + 1 \cdot (1/3) \\ &= 1/2 \end{aligned}$$

and by applying Bayes' Theorem. Take a moment to verify this.

Quiz 41:
To familiarize ourselves with Bayesian reasoning, consider the following example.
You cheated in a casino by smuggling in a loaded die. When the croupier finds out, you run away and escape with your sports car. As it happens, there are only red and yellow sports cars in the city. 90% of the sports cars in the city are red and 10% are yellow. The croupier identified the car as yellow. The police tested the reliability of the croupier under the same circumstances that existed in the night of the run away and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time. What should the police conclude about the probability that you escaped with a yellow sports car knowing that the croupier identified it as yellow?
Before you calculate, guess the result.
Solution

5.3 The Monty Hall Problem II (4:27)

We have presented two ways of reasoning, and both led to different posterior probabilities and different recommendations. On evidence E, switching is not advantageous, and on evidence F it is. Clearly only one way of reasoning can be right, but which? Put differently, which evidence should you condition on? Should you condition on evidence E, or should you condition on evidence F?

To decide this question, let us do an experiment. Let a chance mechanism decide behind which door of the three doors the car is put. Let a player then choose a door, and let Monty open another door. Monty then asks his question, and the player decides whether to switch or not. We then check whether the player won the car or not. We consider two players, following different strategies. Player 1 always switches, and player 2 never switches when Monty asks his question. If we run our experiment many, many times with both players, we can easily calculate the frequency with which each wins the car. Clearly, we will have to do this many, many times to get reliable results. After all we are dealing with a chance mechanism. Given that life is short, we don't do the experiments ourselves, but use a computer program instead which allows us to run many games in a short period of time. And here is the result (see Figure 5.7):

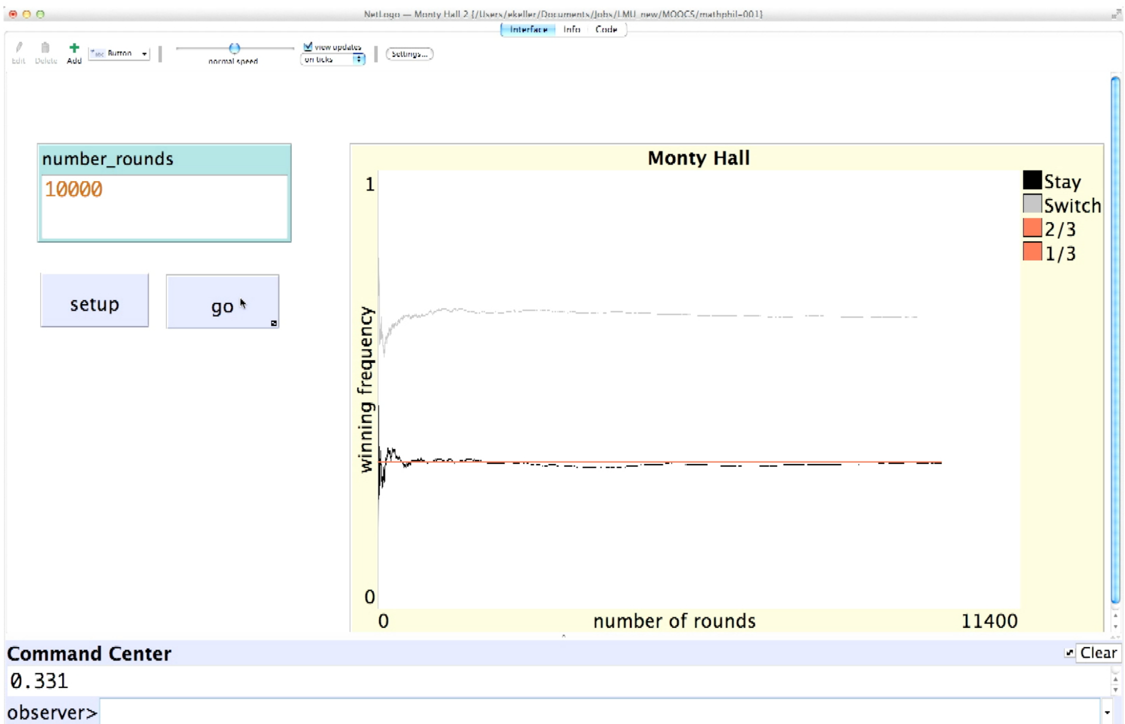

Figure 5.7: Computer simulation

The plot shows the winning frequency of player 1 (upper curve) and the winning frequency of player 2 (lower curve) as a function of the number of games. We see that the winning frequency of player 1 is larger than the winning frequency of player 2, and that the winning frequency of player 1 seems to converge to 2/3 and that the winning frequency of player 2 seems to converge to 1/3, which is the red line. Just as our second way of reasoning has it.

Hence, we have experimentally established that the second way of reasoning is the right one. It used the evidence F to update the probabilities of H_1, H_2 and H_3. But what is wrong with evidence E? Why should we update on F, and not on E? To answer this question, we first note that E is a logical consequence of F, but F is not a logical consequence of E. If Monty opens door No. 3 (F), then we can infer that the car is not behind door No. 3 (E).

However, if the car is not behind door No. 3 (E), then we cannot infer that Monty opens door No. 3 (F). He could also open door No. 2, if the car is behind door. No. 1. To sum up: F is logically stronger than E. It contains all the relevant information for the question at hand that is available to us, whereas E disregards some of the available evidence (because it is logically weaker than F). To properly change our beliefs, it is important to condition on all relevant information. This is the Principle of Total Evidence, which is associated with the name of Rudolf Carnap.

Quiz 42:
The simulation assumes that probabilities are limiting relative frequencies. In the long run, i.e. the more games we run, our numerical experiments will approximate the true probability. Note, though, that the identification of probabilities with relatives frequencies is not without problems. Have a look, for example, at
http://plato.stanford.edu/entries/probability-interpret/
which discuses various versions of the frequency interpretation and their problems (along with other interpretations of probability and their problems).
Here we consider a position called finite frequentism, according to which the probability of an attribute A in a finite reference class B is the relative frequency of actual occurrences of A within B. Here is what this means: Let us assume that a coin is tossed ten times with the following result (the "reference class"): $T, H, T, H, T, T, T, H, T, H$. Simple counting reveals that the attribute H occurred 4 times, and hence the relative frequency is $4/10 = .4$. Show that, in this example, $P(H) + P(T) = 1$. Next, show that this equation also holds if the coin is tossed N times (with n occurrences of H).
Solution

You might be curious to learn how the simulation was programmed. To do so, we used the software NetLogo, which is very convenient and easy to use. What is more, it is downloadable for free! Learn more about it, e.g. by checking out the Models Library (under Files) and start playing with it:
http://ccl.northwestern.edu/netlogo/

5.4 Confirmation (14:02)

The problem we have been discussing is commonly called the Monty Hall Problem, after the real game show host Monty Hall. When it was first posed in 1975, it caused a lot of discussion, also amongst professional mathematicians. Many people with PhDs refused to believe that switching is the optimal strategy in this game, so if this seems counterintuitive to you, you are in good company.

The Monty Hall Problem nicely illustrates a number of issues in probabilistic reasoning.

We will follow up on some of them below. Before, however, let us introduce the concept of confirmation. What do we mean when we say that evidence E confirms the hypothesis H relative to background knowledge K? According to the Bayesian theory of confirmation, which is currently the most popular account, E confirms H iff $P_E(\mathrm{H}) > P(\mathrm{H})$, i.e. if the posterior probability of H is greater than the prior probability of H. E disconfirms H, iff $P_E(\mathrm{H}) < P(\mathrm{H})$. This is plausible, because confirming evidence makes the hypothesis in question more probable, and disconfirming evidence makes it less probable. Consider again a medical test which reports either positive or negative, and which has small rates of false positives and false negatives. If the test result is positive, then we are more convinced that the patient has the disease, and if the test is negative, we are less convinced that the patient has the disease. The test confirms or disconfirms the hypothesis that the patient has a certain disease.

(Slide 22)

Confirmation – The Bayesian Way

E confirms H iff $P_E(\mathrm{H}) > P(\mathrm{H})$.

E disconfirms H, iff $P_E(\mathrm{H}) < P(\mathrm{H})$.

E is irrelevant for H, iff $P_E(\mathrm{H}) = P(\mathrm{H})$.

As you might have noticed, there is a third possibility. E may be irrelevant for H. This is so iff $P_E(\mathrm{H}) = P(\mathrm{H})$. In this case we have $P(\mathrm{H}|\mathrm{E}) = P(\mathrm{H})$, or, put differently, $P(\mathrm{H}, \mathrm{E}) = P(\mathrm{H})P(\mathrm{E})$. One also says that H and E are probabilistically independent in this case.

(Slide 23)

Irrelevance:

$$P(\mathrm{H}|\mathrm{E}) = P(\mathrm{H}) \Leftrightarrow P(\mathrm{H}, \mathrm{E}) = P(\mathrm{H})P(\mathrm{E})$$

Let us further illustrate these ideas with the Monty Hall Problem. Here we see that E confirms H_1 and H_2, but E disconfirms H_3. F is irrelevant for H_1, it confirms H_2, and it disconfirms H_3. Is this plausible? Well, it is obvious that E and F both disconfirm H_3. After learning E or after learning F the hypothesis H_3 is impossible. The car cannot be behind door No. 3 in these cases. It is also clear that E raises the probability of H_1 as well as of H_2, because there is one possibility less where the car is. But what about evidence F? Why is F irrelevant for H_1? Interestingly, F does not give us any new information which helps us to judge the truth or falsity of H_1. This is because Monty is simply not allowed to open door No. 1. You knew this beforehand, it is part of the rules, as Monty cannot open the door you have picked. So how could his choice of a door he opens have any implications for your belief in the truth or falsity of H_1? Hence, the probability of H_1 remains 1/3. F is irrelevant for H_1. Note furthermore that the new probability of H_3 is 0, and so the new probability of H_2 has to be 2/3 to make sure that all three new probabilities add up to 1. Hence, F confirms H_2.

Bayesian confirmation theory makes sense of a number of intuitive methodological principles. For example, there is the insight that surprising evidence E confirms a hypothesis H more than less surprising evidence. If we expected the evidence to obtain (i.e. if $P(\mathrm{E})$ is large), then observing E tells us much less about the truth or falsity of the hypothesis than if the evidence is unexpected (i.e. if $P(\mathrm{E})$ is small). It is easy to see why this is so.

(Slide 24)

We assume: E is a deductive consequence of H.

Hence, $P(\mathrm{E}|\mathrm{H}) = 1$.

Using Bayes Theorem and Bayes Rule:

$$P_E(\mathrm{H}) = \frac{P(\mathrm{H})}{P(\mathrm{E})}$$

We conclude: The smaller $P(\mathrm{E})$, the greater $P_E(\mathrm{H})$ for a fixed $P(\mathrm{H})$.

This is exactly what we wanted to show. A similar claim can also be made if the evidence is not a deductive consequence of H. Let E and F be evidence for a hypothesis H_2 and let $P(\mathrm{F}) < P(\mathrm{E})$. Let us furthermore assume that the likelihoods are equal, i.e. that $P(\mathrm{E}|\mathrm{H}_2) = P(\mathrm{F}|\mathrm{H}_2)$. Then $P_{\mathrm{F}}(\mathrm{H}_2) > P_{\mathrm{E}}(\mathrm{H}_2)$. F confirms H_2 better than E, because F is less expected than E. This is exactly the situation we find in the Monty Hall Problem. Given evidence F, the probability that the car is behind door No. 2 is greater than given evidence E.

You might wonder that the Monty Hall Problem is a rather special problem and that confirmation in science or in other contexts is different. And indeed, the Monty Hall problem is rather special. It assumes, for example, that the prior probabilities of all three mutually exclusive and exhaustive alternatives are equal. We have applied the Principle of Indifference and assigned the same prior probability to all three hypotheses because there is nothing that tells us that one hypothesis is more likely than another. But perhaps we have a sense that one is more likely than another. After all, we have picked door No. 1. So perhaps there is a reason for this, perhaps unbeknownst to us (maybe we heard goat-like sounds behind doors 2 and 3, while we thought to discern the sweet sound of a Porsche engine behind door 1). In this case, it is reasonable to set the prior probability of H_1 at a higher value than the prior probability of H_2 and H_3. We ask: Is switching then still the right decision?

(Slide 25)

Let $P(\mathrm{H}_1) = h_1, P(\mathrm{H}_2) = h_2$, and $P(\mathrm{H}_3) = h_3$.

Let $h_1 > h_2, h_3$.

Then we calculate:

(Slide 26)

$$P_F(\mathrm{H}_1) = \frac{1/2 \cdot h_1}{P(\mathrm{F})}$$

$$P_F(\mathrm{H}_2) = \frac{h_2}{P(\mathrm{F})}$$

$$P(\mathrm{F}) = 1/2 \cdot h_1 + h_2$$

$$P_F(\mathrm{H}_2) > P_F(\mathrm{H}_1) \text{ iff } h_2 > 1/2 \cdot h_1$$

In this case, switching is only advantageous, given our assessment of the situation, if we initially consider H_2 to be sufficiently probable (i.e. if $\mathrm{h}_2 > 1/2 \cdot \mathrm{h}_1$) - (i.e., if we are sufficiently skeptical of our ability to discern the sounds of a Porsche engine from goat sounds). Any player who assigns corresponding prior probabilities in a one-off situation should switch. Otherwise, she should not switch.

Given Monty was careful enough to soundproof the doors, it may sound unreasonable to assign prior probabilities other than the ones that follow from applying the Principle of Indifference. But in many cases the Principle of Indifference cannot be applied and one is left with subjective judgments of the reasoning agents.

Note that we have applied the Principle of Indifference twice in our analysis of the Monty Hall Problem. When calculating the likelihood $P(\mathrm{F}|\mathrm{H}_1)$, we have assumed that Monty has two options. He can either open door No. 2, or he can open door No. 3. Now let us suppose that although you believe that $P(\mathrm{H}_1) = P(\mathrm{H}_2) = P(\mathrm{H}_3)$, you believe that it is more likely that Monty will open door No. 3 than door No. 2. Hence you assign a likelihood

(Slide 27)

$$P(\mathrm{F}|\mathrm{H}_1) =: a, \text{ with } 1/2 < a < 1$$

(a is a randomization parameter.) Hence,

(Slide 28)

$$P(\mathrm{F}|\mathrm{H}_1) =: a, \text{ with } 1/2 < a < 1$$

$$\begin{aligned} P_F(\mathrm{H}_1) &= \frac{a \cdot 1/3}{P(\mathrm{F})} \\ P_F(\mathrm{H}_2) &= \frac{1/3}{P(\mathrm{F})} \\ P(\mathrm{F}) &= (1+a) \cdot 1/3 \end{aligned}$$

Hence, $P_F(\mathrm{H}_2) > P_F(\mathrm{H}_1)$. Note that this results holds for all $a \in (0, 1)$.

That is, whatever the value of the randomization parameter is, switching is always better. The strong assumption that we originally made and which led to the assignment of $a = 1/2$ is therefore not necessary to derive the recommendation that switching is advantageous. Our analysis is robust with regard to this assumption.

Quiz 43:
(1): Show that E confirms H if and only if E disconfirms ¬H.
(2): Show that E confirms H if and only if ¬E disconfirms H.
(3): Le H be a hypothesis and E be a piece of evidence. The likelihood ratio is defined as follows:

$$L := \frac{P(\mathrm{E}|\mathrm{H})}{P(\mathrm{E}|\neg\mathrm{H})}.$$

Show that E confirms H if and only if $L > 1$.
Solution

5.5 Final Remarks (5:40)

Our discussion illustrates that we always have to provide a probability model when we analyze problems like the Monty Hall Problem or other confirmation-theoretical scenarios. When we have made assumptions about what the relevant variables are, we have to specify which dependency relations hold between them; And we have to fix the prior probabilities and the likelihoods. Once this is done, the Bayesian machinery allows us to draw conclusions which will eventually help us to make good decisions.

Confirmation theory is one of the major research areas of the mathematical philosopher. This lecture focused on the most popular theory of confirmation: Bayesian confirmation theory. According to this theory, confirmation is probability raising. It assumes that agents assign probabilities to hypotheses and that they change these probability assignments when faced with relevant evidence. A hypothesis is confirmed by the evidence if its probability increases. It is disconfirmed if the probability decreases.

We focused on the Monty Hall problem, but this simple but powerful idea sheds light on many problems in epistemology and the philosophy of science. To close, let me provide you with a few more examples.

In the philosophy of science, Bayesian confirmation theory is used to account for the variety of evidence thesis. According to this thesis, more varied evidence confirms a hypothesis better than less varied evidence. What is more varied evidence? The idea is this: Let's assume that you want to test the Newtonian mechanics. You can do this, for example, by dropping a rock from a hight of 1 meter and measure how long it takes the rock to fall down. Then compare your measurement with the Newtonian prediction and, if all goes well, the measurement confirms the theory. To get more confirmation, you can repeat the experiment again and again and again. However, a better strategy to confirm the theory seems to be to invoke different evidence - Say experiments involving a pendulum, or data about the orbits of the various planets. If these tests also confirm the theory, then their combination will speak much more in favor of the theory than the simple repetition of the same experiment. This is the variety of evidence thesis. Bayesian confirmation theory allows us to model different testing scenarios and to explore under which conditions the thesis holds.

In epistemology, Bayesian confirmation theory is used, for instance, to study whether and how we learn from testimony. We often rely on the testimony of others - In the court room, as well as in science. But how shall we take testimonial evidence into account? This requires us to carefully model the various information sources, their dependencies and their reliabilities. To do so, Bayesian statisticians and computer scientists have developed helpful tools. Most importantly, the theory of Bayesian networks provides powerful tools to model scenarios involving a large number of variables and to represent their dependencies and independencies in an intuitive graphical way. Mathematical philosophy is, after all, an interdisciplinary endeavor, and it is important to look around which potentially helpful tools other sciences can provide us with.

Bayesian confirmation theory is also of relevance for empirical Psychology. Here psychologist study how humans actually reason and which putative reasoning fallacies occur. To assess actual human reasoning it is pertinent to properly model the cognitive state of the agent and her reasoning process. Not surprisingly, Bayesian confirmation theory is of much help here as an enormous recent literature documents. In the exercises, we discuss some of them and leave it to you find other applications of Bayesian confirmation theory.

There are many more paradoxes involving probabilities and probabilistic fallacies. Some of them are discussed in one of the books that Hannes mentioned already at the end of Lecture 1:

Clark, M., *Paradoxes from A to Z*, second edition, London: Routledge, 2007.

Here are four excellent relevant entries from the Stanford Encyclopedia of Philosophy:
http://plato.stanford.edu/entries/epistemology-bayesian/,
http://plato.stanford.edu/entries/confirmation/,
http://plato.stanford.edu/entries/bayes-theorem/.

There is also an excellent relevant entry in the Internet Encyclopedia of Philosophy:
http://www.iep.utm.edu/conf-ind/.

Here are some books about probability theory and its history:

Hacking, I., *The Taming of Chance*, Cambridge: Cambridge University Press, 1990.

Hacking, I., *An Introduction to Probability and Inductive Logic*, Cambridge: Cambridge University Press, 2000.

Hacking, I., *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability Induction and Statistical Inference*, second edition, Cambridge: Cambridge University Press, 2006.

Haigh, J., *Probability: A Very Short Introduction*, Oxford: Oxford University Press, 2012.

Skyrms, B., *Choice and Chance: An Introduction to Inductive Logic*, Wadsworth Publishing, 1999.

And here are some relevant books on Bayesian reasoning:

Bovens, L. and Hartmann, S., *Bayesian Epistemology*, Oxford: Oxford University Press, 2003.

Earman, J., *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*, Cambridge, Mass.: The MIT Press, 1992.

Howson, C. and Urbach, P., *Scientific Reasoning: The Bayesian Approach: The Bayesian Method*, third edition, Open Court Publishing, 2006.

Jeffrey, R., *Subjective Probability: The Real Thing*, Cambridge: Cambridge University Press, 2004.

Finally, here are some relevant references from the psychology of reasoning:

Gigerenzer, G., *Rationality for Mortals: How People Cope with Uncertainty*, Oxford: Oxford University Press, 2010.

Hahn, U., and Oaksford, M., "The Rationality of Informal Argumentation: A Bayesian Approach to Reasoning Fallacies." *Psychological Review*, 114 (2007), 704-732.

Kahneman, D., *Thinking, Fast and Slow*, Farrar, Straus and Giroux, 2011.

Oaksford, M. and Chater, N., *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*, Oxford: Oxford University Press, 2007.

Appendix A

Quiz Solutions Week 5: Confirmation

Quiz 40:

Before we proceed, let us reflect upon the assignment of initial probabilities for H_1, H_2 and H_3. Our argument for assigning the same probability to H_1, H_2 and H_3 was something like this: If there is "no known reason for predicating for our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability." (J. M. Keynes, A Treatise on Probability, p. 42) Keynes called this rule the Principle of Indifference (ibid., p. 41). Unfortunately, this principle may lead to contradictory conclusions. Consider the following case. If there is no known reason for saying that my friend's new sports car is yellow rather than red, then, relative to this knowledge, it is equally probable that the sports car is yellow and that the sports car is red. We conclude that the probability of "The sports car is yellow" is 1/2 and that the probability of "The sports car is red" is 1/2. Similarly, we conclude that the probability of each of the propositions "The sports car is yellow" and "The sports car is orange" is 1/2. Show that this leads to a contradiction.

SOLUTION Quiz 40:

The three alternatives – "The sports car is yellow," "The sports car is orange", and "The sports car is red" are mutually exclusive alternatives. Hence, the probability of their disjunction cannot add to more than one. However,

$P(\text{Yellow OR Red OR Orange}) = P(\text{Yellow}) + P(\text{Red}) + P(\text{Orange}) = 3 \cdot \frac{1}{2} = \frac{3}{2} > 1.$ Contradiction! (Here we have used short-hand notation for the three propositions.)

Back to quiz

Quiz 41:

To familiarize ourselves with Bayesian reasoning, consider the following example.

You cheated in a casino by smuggling in a loaded die. When the croupier finds out, you run away and escape with your sports car. As it happens, there are only red and yellow sports cars in the city. 90% of the sports cars in the city are red and 10% are yellow. The croupier identified the car as yellow. The police tested the reliability of the croupier under the same circumstances that existed in the night of the run away and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time. What should the police conclude about the probability that you escaped with a yellow sports car knowing that the croupier identified it as yellow?

Before you calculate, guess the result.

SOLUTION Quiz 41:

We introduce the following variables:

R: The car is red.

Y: The car is yellow.

$RepR$: The croupier identifies the car as red.

$RepY$: The croupier identifies the car as yellow.

Next, we assign probabilities based on the given information:

$P(R) = .9$

$P(Y) = .1$

$P(RepR|R) = .8$

$P(RepY|Y) = .8$

From this information, we want to calculate $P(Y|RepY)$. Using Bayes Rule, we get

$P(Y|RepY) = \frac{P(RepY|Y)P(Y)}{P(RepY)} = \frac{.8 \cdot .1}{P(RepY)} = \frac{.08}{P(RepY)}$

Next, we can calculate

$P(RepY) = P(RepY|Y)P(Y) + P(RepY|\neg Y)P(\neg Y)$.

With $P(\neg Y) = P(R) = .9$

and

$P(RepY|\neg Y) = P(\neg RepR|R) = 1 - P(RepR|R) = .2$, we obtain

$P(RepY) = .8 \cdot .1 + .2 \cdot .9 = .26.$

Hence,

$P(Y|RepY) = \frac{.08}{.26} = \frac{8}{26} = .31.$

Many people would have expected a much higher probability, given that the witness is fairly reliable. And yet, given the report of the croupier, it is twice as likely that the car is red than that the car is yellow. The reason is that there are so much more red cars than yellow cars in the city. The base rate (i.e the prior probability) that the car is red is much higher than the base rate that the car is yellow. Neglecting the base rate in one's reasoning means to commit the base rate fallacy. It turns out that many people commit this fallacy.

For more on this, see

http://en.wikipedia.org/wiki/Base_rate_fallacy

http://plato.stanford.edu/entries/bayes-theorem/, especially Example 2 in the supplement:

http://plato.stanford.edu/entries/bayes-theorem/supplement

I also recommend the following article, written by two psychologists:

Gigerenzer, G. and U. Hoffrage, "How to Improve Bayesian Reasoning without Instruction: Frequency Formats", Psychological Review 102(4), 684-704.

Back to quiz

Quiz 42:

The simulation assumes that probabilities are limiting relative frequencies. In the long run, i.e. the more games we run, our numerical experiments will approximate the true probability. Note, though, that the identification of probabilities with relatives frequencies is not without problems. Have a look, for example, at

http://plato.stanford.edu/entries/probability-interpret/

which discuses various versions of the frequency interpretation and their problems (along with other interpretations of probability and their problems).

Here we consider a position called finite frequentism, according to which the probability of an attribute A in a finite reference class B is the relative frequency of actual occurrences of A within B. Here is what this means: Let us assume that a coin is tossed ten times with the following result (the "reference class"): $T, H, T, H, T, T, T, H, T, H$. Simple counting reveals that the attribute H occurred 4 times, and hence the relative frequency is $4/10 = .4$. Show that, in this example, $P(H) + P(T) = 1$. Next, show that this equation also holds if the coin is tossed N times (with n occurrences of H).

SOLUTION Quiz 42:

Note $P(T) = 6/10 = .6$, and hence $P(H) + P(T) = 1$. If the coin is tossed N times and H occurs n times, then (according to finite frequentism) $P(H) = n/N$. T occurs $N - n$ times, hence (according to finite frequentism) $P(T) = \frac{(N-n)}{N}$. We now calculate

$P(H) + P(T) = \frac{n}{N} + \frac{(N-n)}{N} = 1$.

Back to quiz

Quiz 43:

(1): Show that E confirms H if and only if E disconfirms $\neg$H.

(2): Show that E confirms H if and only if $\neg$E disconfirms H.

(3): Le H be a hypothesis and E be a piece of evidence. The likelihood ratio is defined as follows:

$$L := \frac{P(\mathrm{E}|\mathrm{H})}{P(\mathrm{E}|\neg\mathrm{H})}.$$

Show that E confirms H if and only if $L > 1$.

SOLUTION Quiz 43:

Solution (1): E confirms H if and only if $P(\mathrm{H}|\mathrm{E}) > P(\mathrm{H})$. The latter expression is equivalent to $1 - P(\mathrm{H}|\mathrm{E}) < 1 - P(\mathrm{H})$, which is equivalent to $P(\neg\mathrm{H}|\mathrm{E}) < P(\neg\mathrm{H})$, which means that E disconfirms $\neg$H.

Solution (2): $\neg$E disconfirms H if and only if $P(\mathrm{H}|\neg\mathrm{E}) < P(\mathrm{H})$. The latter is equivalent to $P(\mathrm{H}, \neg\mathrm{E}) < P(\neg\mathrm{E})P(\mathrm{H})$, which is equivalent to $P(\mathrm{H}) - P(\mathrm{H}, \mathrm{E}) < P(\mathrm{H}) - P(\mathrm{E})P(\mathrm{H})$. This is equivalent to $P(\mathrm{H}, \mathrm{E}) > P(\mathrm{E})P(\mathrm{H})$, which in turn is equivalent to $P(\mathrm{H}|\mathrm{E}) > P(\mathrm{H})$, which means that E confirms H.

Solution (3): E confirms H if and only if $P(\mathrm{H}|\mathrm{E}) > P(\mathrm{H})$. The latter expression is equivalent to $\frac{P(\mathrm{E}|\mathrm{H})P(\mathrm{H})}{P(\mathrm{E}|\mathrm{H})P(\mathrm{H}) + P(\mathrm{E}|\neg\mathrm{H})P(\neg\mathrm{H})} > P(\mathrm{H})$. Using the definition of the likelihood ratio, this is equivalent to $\frac{L \cdot P(\mathrm{H})}{L \cdot P(\mathrm{H}) + P(\neg\mathrm{H})} > P(\mathrm{H})$. Multiplying both sides of the inequality with $L \cdot P(\mathrm{H}) + P(\neg\mathrm{H})$, we arrive at the equivalent inequality $L \cdot P(\mathrm{H}) > L \cdot P(\mathrm{H})^2 + P(\mathrm{H})P(\neg\mathrm{H})$, which is equivalent to $L > 1$ (make this last step explicit!).

Likelihood ratios play an important role in confirmation theory, inductive logic and statistics. If you want to learn more about them, have a look at

http://plato.stanford.edu/entries/logic-inductive/

and chapter 1 of Elliott Sober's book Evidence and Evolution: The Logic Behind the Science, Cambridge: Cambridge University Press, 2008.

Back to quiz

List of Figures