Learning Causal Conditionals

Stephan Hartmann

Munich Center for Mathematical Philosophy

LMU Munich

Bayes Forum

Garching, March 2016

Motivation

- Bayesian approaches are becoming more and more important in statistics.
- Bayesian approaches are also becoming more and more important in (mathematical) philosophy.
- The goal of this talk is to motivate the philosophers' interest in Bayesianism and to discuss one specific example in more detail.
- The example studies how we should change our beliefs in the light of evidence of the form "If A, B".

Overview

I. Bayesianism in Philosophy

Confirmation Theory

- What does it mean that a piece of evidence E confirms (or supports) a theory or hypothesis H?
- To address this question, philosophers developed deductivist and inductivist accounts.
  - Deductivist accounts: the hypothetico-dedcutive model, Popper's falsificationsim
  - Inductivist accounts: Bayesianism
- The problems of deductivist accounts lad to the present popularity of Bayesianism.

Deductivist Accounts I: The Hypothetico-Deductive Model

- According to the hypothetico-deductive model, a theory or hypothesis H is confirmed by a piece of evidence E iff E is predicted by H (i.e. if E is a deductive consequence of H) and if E is observed.
- The model has a number of well-known problems, e.g.
  1. The Tacking Problem: If E confirms H, then it also confirms $H \wedge X$. Note that X can be a completely irrelevant proposition. This is counter-intuitive.
  2. Degrees of confirmation: Some evidence confirms a theory or hypothesis stronger than other evidence. However, according to the hypothetico-deductive model, we can only make the qualitative inference that E confirms H (or not).
  3. Kinds of Evidence: Theories can only be confirmed in the light of observed deductive consequences of the theory in question. However, there may be other kinds of evidence.

Deductivist Accounts II: Popper's Falsificationism

- According to naive falsificationism, a theory or hypothesis H is corroborated if an empirically testable prediction of H obtains. Otherwise it is falsified and should be rejected and replaced by an alternative theory.
- Corroboration is only concerned with the past performance of the hypothesis in question. It does not tell us anything about the future. (This is a consequence of Popper's strong anti-inductivism.)
- The three problems of the HD model are also a problems for falsificationism.
- N.B.: More sophisticated versions of falsificationism have the same problems as naive falsificationism, and so I won't discuss them here.

Inductivist Accounts: Bayesian Confirmation Theory

- According to Bayesian Confirmation Theory, a theory or hypothesis H is confirmed by a piece of evidence E iff the observation of E raises the (subjective) probability of H.
- Scientists attach a degree of belief (= a probability) to a theory or hypothesis and change ("update") it in the light of new evidence.
- What evidence? A deductive or inductive consequence of the hypothesis, a testimony from a partially reliable source,. . .
- How should one update? Conditionalization: The posterior probability of H (i.e. $P'(\mathrm{H})$) follows from the prior probability of H (i.e. $P(\mathrm{H})$), the likelihood of the evidence (i.e. $P(\mathrm{E}|\mathrm{H})$) and the expectancy of the evidence $i.e. P(\mathrm{E})$:

Bayes' Theorem

$$P'(\mathrm{H}) := P(\mathrm{H}|\mathrm{E}) = \frac{P(\mathrm{E}|\mathrm{H}) \cdot P(\mathrm{H})}{P(\mathrm{E})}$$

Jeffrey Conditionalization

If the evidence is uncertain, then Jeffrey Conditionalization has to be used:

Jeffrey Conditionalization

$$P'(\mathrm{H}) := P(\mathrm{H}|\mathrm{E})\, P'(\mathrm{E}) + P(\mathrm{H}|\neg\mathrm{E})\, P'(\neg\mathrm{E})$$

- Note that this rule follows from the law of total probability

  $$P'(\mathrm{H}) := P'(\mathrm{H}|\mathrm{E})\, P'(\mathrm{E}) + P'(\mathrm{H}|\neg\mathrm{E})\, P'(\neg\mathrm{E})$$

  under the assumption ("Rigidity") that $P'(\mathrm{H}|\mathrm{E_i}) = P'(\mathrm{H}|\mathrm{E_i})$ for all elements of the partition of E.
- Bayes' Thorem is a special case of Jeffrey Conditionalization (when the evidence is certain, i.e. when $P'(\mathrm{E}) = 1$).

Justifications Bayes' Theorem

1. Pragmatic: Dutch Book arguments
2. Epistemic: Minimization of the inaccuracy of beliefs
3. "Distance" minimization: Kullback-Leibler divergence, Hellinger,. . .

The Kullback-Leibler Divergence

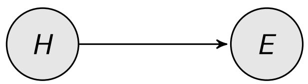
- Let $\mathrm{S}_1, \ldots, \mathrm{S}_\mathrm{n}$ be the possible values of a random variable S over which probability distributions P and P' are defined.
- The Kullback-Leibler divergence between P' and P is then given by

$$D_{KL}(P'||P) := \sum_{i=1}^{n} P'(\mathrm{S_i}) \log \frac{P'(\mathrm{S_i})}{P(\mathrm{S_i})} .$$

- Note that the KL divergence is not symmetrical. So it is not a distance.
- Note also that if the old distribution P is the uniform distribution, then minimizing the Kullback-Leibler divergence amounts to maximizing the entropy $Ent(P) := -\sum_{i=1}^{n} P(\mathrm{S_i}) \log P(\mathrm{S_i})$.

Conditionalization

- We introduce the binary propositional variables H and E:
  H: "The hypothesis holds", and ¬H: "The hypothesis does not hold".
  E: "The evidence obtains", and ¬E: "The evidence does not obtain".
- The probabilistic relation between H and E can be represented in a Bayesian Network:



- We set $P(\mathrm{H}) = h$ and $P(\mathrm{E}|\mathrm{H}) = p$, $P(\mathrm{E}|\neg\mathrm{H}) = q$.

Conditionalization

- Calculate the prior distribution over H and E. ($\overline{x} := 1 - x$)

$$P(\mathrm{H},\mathrm{E}) = h\,p \quad , \quad P(\mathrm{H},\neg\mathrm{E}) = h\,\overline{p}$$
$$P(\neg\mathrm{H},\mathrm{E}) = \overline{h}\,q \quad , \quad P(\neg\mathrm{H},\neg\mathrm{E}) = \overline{h}\,\overline{q}\,.$$

- Next, we learn that E obtains, i.e. $P'(\mathrm{E}) = 1$.
- We assume that the network stays the same as before. Hence

$$P'(\mathrm{H},\mathrm{E}) = h'\,p' \quad , \quad P'(\mathrm{H},\neg\mathrm{E}) = h'\,\overline{p'}$$
$$P'(\neg\mathrm{H},\mathrm{E}) = \overline{h'}\,q' \quad , \quad P'(\neg\mathrm{H},\neg\mathrm{E}) = \overline{h'}\,\overline{q'}\,.$$

- From $P'(\mathrm{E}) = h'\,p' + \overline{h'}\,q' = 1$, we conclude that $p' = q' = 1$.
- Minimize the KL divergence: $P'(\mathrm{H}) = P(\mathrm{H}|\mathrm{E})$
- Note that Jeffrey conditionalization obtains if one learns E with $P'(\mathrm{E}) =: e' < 1$.

Other Distances

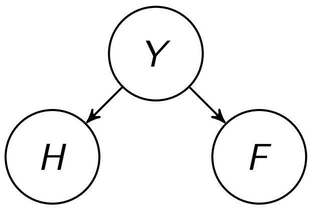
The Hellinger distance between P' and P is then given by

$$D_H(P'||P) := \sum_{i=1}^{n} \left(\sqrt{P'(\mathrm{S_i})} - \sqrt{P(\mathrm{S_i})}\right)^2 .$$

yields the same results (i.e. Bayes' Theorem and Jeffrey Condtionalization).

Different Kinds of Evidence

- Sometimes the evidence is not a deductive or inductive consequence of the hypothesis in question.
- Example 1: The No Alternatives Argument
- Example 2: Analog Simulation
- These cases can be modeled and analyzed in the Bayesian framework. Hence, there is no need to aim for a new methodology of science, as George Ellis and Joe Silk suggested in a recent article in *Nature*.



II. Learning Conditionals: Four Challenges and a Recipe

Yet Another Type of Evidence: Conditionals

- Sometimes we learn indicative conditionals of the form "if A, B".
- Here we ask: How should we change our beliefs in the light of this evidence?
- I will discuss several examples below that show that learning a conditional sometimes makes us change our beliefs, and sometimes not. But how should a rational agent change her beliefs in the light of this evidence?
- There are several proposals discussed in the literature, but in a recent survey, Igor Douven (2012) concludes that a proper general account of probabilistic belief updating by learning (probabilistic) conditional information is still to be formulated.
- My goal is to provide such an account (at least for causal conditionals).

Conditionalization and the Material Conditional

- How should we change our beliefs in the light of this evidence?
- If we want to use Bayesian conditionalization, then we have to formally represent the conditional.
- Perhaps naturally, we use the material conditional and identify $\mathrm{A} \rightarrow \mathrm{B}$ with $\neg\mathrm{A} \vee \mathrm{B}$. Popper and Miller (1983) have shown that then

  $$P^*(\mathrm{A}) := P(\mathrm{A}|\mathrm{A} \rightarrow \mathrm{B}) < P(\mathrm{A})$$

  if $P(\mathrm{A}) < 1$ and $P(\mathrm{B}|\mathrm{A}) < 1$.
- This leads to counterintuitive consequences as, e.g., the sundowners example below demonstrates.
- However, if we do not use the material conditional, then we cannot express the conditional in Boolean terms, and hence we cannot apply conditionalization.
- Question: What can be done?

Stalnaker's Thesis

Stalnaker proposed to identify the probability of a conditional with the conditional probability:

Stalnaker's Thesis

$$P(\mathrm{A} \rightarrow \mathrm{B}) = P(\mathrm{B}|\mathrm{A})$$

- This thesis, which Stalnaker found trivial, has been criticized, most famously perhaps by Lewis who came up with various triviality results.
- Note, however, that Stalnaker's thesis (even if it were true) cannot be applied directly to learning a conditional via Bayes' Theorem. It simply does not tell us how to do this.

My General Recipe

- One way to proceed is to use the conditional probability assignment as a constraint on the new probability distribution P'. Apart from satisfying the constraint, P' has to be as close as possible to the old distribution P, i.e. we want to change our beliefs conservatively.
- Technically, this is done by minimizing the Kullback-Leibler divergence between the posterior and the prior distribution.
- While this might sound like a reasonable (and practicable) proposal, van Fraassen and others have confronted it with counterexamples, most famously the Judy Benjamin example.
- In this talk I want to show that the proposed procedure works if one additionally makes sure that the causal structure of the problem at hand is properly taken into account.

Challenge 1: The Ski Trip Example

Harry sees his friend Sue buying a skiing outfit. This surprises him a bit, because he did not know of any plans of hers to go on a skiing trip. He knows that she recently had an important exam and thinks it unlikely that she passed. Then he meets Tom, his best friend and also a friend of Sue, who is just on his way to Sue to hear whether she passed the exam, and who tells him,

If Sue passed the exam, then her father will take her on a skiing vacation.

Challenge 1: The Ski Trip Example

Harry sees his friend Sue buying a skiing outfit. This surprises him a bit, because he did not know of any plans of hers to go on a skiing trip. He knows that she recently had an important exam and thinks it unlikely that she passed. Then he meets Tom, his best friend and also a friend of Sue, who is just on his way to Sue to hear whether she passed the exam, and who tells him,

If Sue passed the exam, then her father will take her on a skiing vacation.

Recalling his earlier observation, Harry now comes to find it more likely that Sue passed the exam.

Ref.: Douven and Dietz (2011)

Challenge 2: The Driving Test Example

Betty knows that Kevin, the son of her neighbors, was to take his driving test yesterday. She has no idea whether or not Kevin is a good driver; she deems it about as likely as not that Kevin passed the test. Betty notices that her neighbors have started to spade their garden. Then her mother, who is friends with Kevin's parents, calls her and tells her the following:

If Kevin passed the driving test, then his parents will throw a garden party.

Challenge 2: The Driving Test Example

Betty knows that Kevin, the son of her neighbors, was to take his driving test yesterday. She has no idea whether or not Kevin is a good driver; she deems it about as likely as not that Kevin passed the test. Betty notices that her neighbors have started to spade their garden. Then her mother, who is friends with Kevin's parents, calls her and tells her the following:

If Kevin passed the driving test, then his parents will throw a garden party.

Betty figures that, given the spading that has just begun, it is doubtful (even if not wholly excluded) that a party can be held in the garden of Kevin's parents in the near future. As a result, Betty lowers her degree of belief for Kevin having passed the driving test.

Ref.: Douven (2011)

Challenge 3: The Sundowners Example

Sarah and her sister Marian have arranged to go for sundowners at the Westcliff hotel tomorrow. Sarah feels there is some chance that it will rain, but thinks they can always enjoy the view from inside. To make sure, Marian consults the staff at the Westcliff hotel and finds out that the inside area will be occupied by a wedding party. So she tells Sarah:

If it rains tomorrow, then we cannot have sundowners at the Westcliff.

Challenge 3: The Sundowners Example

Sarah and her sister Marian have arranged to go for sundowners at the Westcliff hotel tomorrow. Sarah feels there is some chance that it will rain, but thinks they can always enjoy the view from inside. To make sure, Marian consults the staff at the Westcliff hotel and finds out that the inside area will be occupied by a wedding party. So she tells Sarah:

If it rains tomorrow, then we cannot have sundowners at the Westcliff.

Upon learning this conditional, Sarah sets her probability for sundowners and rain to 0, but she does not adapt her probability for rain.

Ref.: Douven and Romeijn (2011)

Challenge 4: Judy Benjamin Problem

A soldier is dropped with her platoon in a territory that is divided in two parts, the Red Territory (R) and the Blue Territory (¬R) where each territory is also divided in two parts, Second Company (S) and Headquarters Company (¬S), forming four sections of almost equal size. The platoon is dropped somewhere in the middle so she finds it equally likely to be in one section as in any of the others, i.e. $P(\mathrm{R}, \mathrm{S}) = \mathrm{P}(\mathrm{R}, \neg\mathrm{S}) = \mathrm{P}(\neg\mathrm{R}, \mathrm{S}) = \mathrm{P}(\neg\mathrm{R}, \neg\mathrm{S}) = 1/4$. Then they receive a radio message:

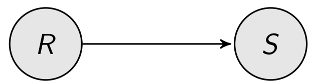
I can not be sure where you are. If you are in Red Territory the odds are 3:1 that you are in the Secondary Company.

How should Judy Benjamin update her belief function based on this communication?

Ref.: van Fraassen (1981)

Judy Benjamin Problem

- We introduce two binary propositional variables. The variable R has the values R: "Judy lands in Red Territory", and ¬R: "Judy lands in Blue Territory". The variable S has the values S: "Judy lands in Second Company", and ¬S: "Judy lands in Headquarters".
- The probabilistic relation between the variables:



- Learning: $P'(\mathrm{S}|\mathrm{R}) = k \neq 1/2$
- Assume that the network does not change. Then minimizing the Kullback-Leibler divergence yields $P'(\mathrm{R}) < P(\mathrm{R})$, which is not intuitive.

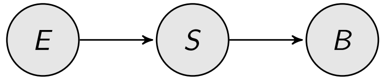
III. The Challenges Met

The Ski Trip Example

We define three variables:

- E: Sue has passed the exam.
- S: Sue is invited to a ski vacation.
- B: Sue buys a ski outfit.

The causal structure is given as follows:



Additionally, we set $P(\mathrm{E}) = e$ and

$$P(\mathrm{S}|\mathrm{E}) = p_1 \quad , \quad P(\mathrm{S}|\neg\mathrm{E}) = q_1$$
$$P(\mathrm{B}|\mathrm{S}) = p_2 \quad , \quad P(\mathrm{B}|\neg\mathrm{S}) = q_2.$$

Note that the story suggests that $p_1 > q_1$ and $p_2 > q_2$.

The Ski Trip Example

- Learning: $P'(\mathrm{B}) = 1$ and $P'(\mathrm{S}|\mathrm{E}) = 1$.
- Again, the causal structure does not change.

  Theorem: *Consider the Bayesian Network above with the prior probability distribution. Let*

  $$k_0 := \frac{p_1\, p_2}{q_1\, p_2 + \overline{q_1}\, q_2}.$$

  We furthermore assume that (i) the posterior probability distribution P' is defined over the same Bayesian Network, (ii) the learned information is modeled as constraints on P', and (iii) P' minimizes the Kullback-Leibler divergence to P. Then $P'(\mathrm{E}) > P(\mathrm{E})$, iff $k_0 > 1$.

- The same result obtains for the material conditional.

The Ski Trip Example: Assessing k_0

1. Harry thought that it is unlikely that Sue passed the exam, hence e is small.
2. Harry is surprised that Sue bought a skiing outfit, hence

   $$P(\mathrm{B}) = e\,(p_1\, p_2 + \overline{p_1}\, q_2) + \overline{e}\,(q_1\, p_2 + \overline{q_1}\, q_2)$$

   is small.
3. As e is small, we conclude that $q_1\, p_2 + \overline{q_1}\, q_2 := \epsilon$ is small.
4. p_2 is fairly large (≈ 1), because Harry did not know of Sue's plans to go skiing, perhaps he even did not know that she is a skier. And so it is very likely that she has to buy a skiing outfit to go on the skiing trip.
5. At the same time, q_2 will be very small as there is no reason for Harry to expect Sue to buy such an outfit in this case.
6. p_1 may not be very large, but the previous considerations suggest that $p_1 \gg \epsilon$.

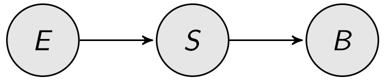
The Ski Trip Example: Assessing k_0

We conclude that

$$k_0 := \frac{p_1\, p_2}{q_1\, p_2 + \overline{q_1}\, q_2} = \frac{p_1}{\epsilon} \cdot p_2$$

will typically be greater than 1. Hence, $P'(\mathrm{E}) > P(\mathrm{E})$.

Using the Material Conditional



- It turns out that the Theorem also obtains if one keeps the causal structure fixed and uses the material conditional ($E \rightarrow S \equiv \neg E \vee S$), i.e. if one calculates $P^*(\mathrm{E}) = P(\mathrm{E}|\neg \mathrm{E} \vee \mathrm{S}, \mathrm{B})$.

No Causal Structure

- What if no causal structure is imposed?
- We computed this case, i.e. we considered only the three variables B, E and S and modeled the learning of $P'(\mathrm{B}) = 1$ and $P'(\mathrm{S}|\mathrm{E}) = 1$ in the usual way.
- Minimizing the KL divergence then leads to $P'(\mathrm{E}) < P(\mathrm{E})$, i.e. to the wrong result.

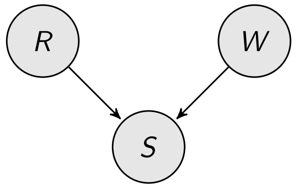
N.B. The driving test example has the same causal structure as the ski trip example and the calculation proceeds accordingly. We do not consider it here.

The Sundowners Example

We define three propositional variables.

- R: It is going to rain.
- W: The inside area is occupied by a wedding party.
- S: Sara and Marian enjoy the sundowners.

The causal structure can be represented as follows:



The Sundowners Example

- Next, we define a probability distribution over this network and let

$$\begin{array}{rcl} P(\mathrm{R}) = r & , & P(\mathrm{P}) = p \\ P(\mathrm{S}|\mathrm{R},\mathrm{W}) = e_1 & , & P(\mathrm{S}|\mathrm{R},\neg\mathrm{W}) = e_2 \\ P(\mathrm{S}|\neg\mathrm{R},\mathrm{W}) = e_3 & , & P(\mathrm{S}|\neg\mathrm{R},\neg\mathrm{W}) = e_4 \end{array}$$

- We assume that $e_1 = 0$.
- With this, we calculate the prior distribution P over the variables R, S, and W.

The Sundowners Example

- Next we learn R → ¬S.
- If we use the material conditional and update ob ¬R ∨ S, then we obtain

$$\begin{aligned} P^*(\mathrm{R}) &= P(\mathrm{R}|\neg\mathrm{R}\vee\neg\mathrm{S}) = \frac{P(\mathrm{R}\wedge(\neg\mathrm{R}\vee\neg\mathrm{S}))}{P(\neg\mathrm{R}\vee\neg\mathrm{S})} \\ &= \frac{P(\mathrm{R}\wedge\neg\mathrm{S})}{P(\neg\mathrm{R}\vee\neg\mathrm{S})} \\ &= \frac{P(\mathrm{R}) - P(\mathrm{R},\mathrm{S})}{1 - P(\mathrm{R},\mathrm{S})}. \end{aligned}$$

- This is counter-intuitive as the probability of R should not change also if $P(\mathrm{R},\mathrm{S}) > 0$.

The Sundowners Example

- Next we learn R → ¬S, i.e. $P'(\neg\mathrm{S}|\mathrm{R}) = 1$ or $P'(\mathrm{S}|\mathrm{R}) = 0 = \mathrm{P}'(\mathrm{R},\mathrm{S})$.
- This is a constraint on the posterior distribution P'.
- Then the following theorem holds.

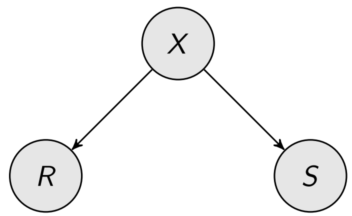
  Theorem: *Consider the Bayesian Network depicted above with the prior probability distribution P. We furthermore assume that (i) the posterior probability distribution P' is defined over the same Bayesian Network, (ii) the learned conditional is modeled as a constraint on P', and (iii) P' minimizes the Kullback-Leibler divergence to P. Then $P'(\mathrm{R}) = P(\mathrm{R})$.*

- Note that using the material conditional yields $P^*(\mathrm{R}) < P(\mathrm{R})$.

The Judy Benjamin Example

We define:

- R: The platoon is dropped in the Red Territory.
- S: The platoon is dropped in the Secondary Company.
- X: Wind comes from a certain direction (or any other cause that comes to mind).



The Judy Benjamin Example

- Learning: $P'(\mathrm{S}|\mathrm{R}) = k \neq 1/2$
- Then the following theorem holds:

Theorem: *Consider the Bayesian Network above with a suitable prior probability distribution P. We furthermore assume that (i) P' is defined over the same Bayesian Network, (ii) the learned information is modeled as a constraint on P', and (iii) P' minimizes the Kullback-Leibler divergence to P. Then $P'(\mathrm{R}) = P(\mathrm{R})$.*

No Causal Structure

- Perhaps it is enough to only take the existence of a third variable X into account, without imposing a causal structure. Let us compute this case!
- We find that imposing $P'(\mathrm{S}|\mathrm{R}) = k \neq 1/2$ as a constraint on the posterior distribution and minimizing the KL divergence leads to $P'(\mathrm{R}) < P(\mathrm{R})$, i.e. to the wrong result.

Problem

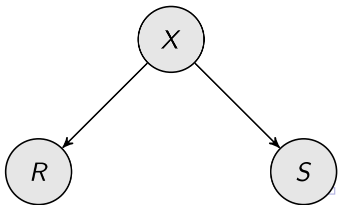
- Intuitively, one would expect that the following posterior distribution:

$$P'(\mathrm{R},\mathrm{S}) = k/2 \quad , \quad P'(\neg\mathrm{R},\mathrm{S}) = 1/4$$
$$P'(\mathrm{R},\neg\mathrm{S}) = \overline{k}/2 \quad , \quad P'(\neg\mathrm{R},\neg\mathrm{S}) = 1/4$$

- However, one obtains

$$P'(\mathrm{R},\mathrm{S}) = k/2 \quad , \quad P'(\neg\mathrm{R},\mathrm{S}) = k/2$$
$$P'(\mathrm{R},\neg\mathrm{S}) = \overline{k}/2 \quad , \quad P'(\neg\mathrm{R},\neg\mathrm{S}) = \overline{k}/2.$$

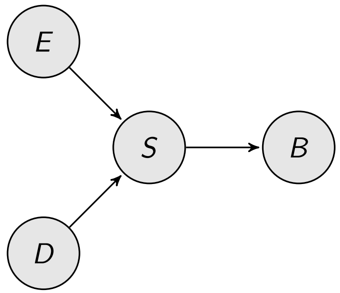
- Note that the learned conditional is not along the causal chain. So our proposal only applies to causal conditionals and another recipe has to be found for non-causal conditionals.



IV. Disabling Conditions

The Ski Trip Example Revisited

- A disabling condition D could obtain.
- Then the modified network looks as follows.



The Ski Trip Example Revisited

- Learning: $P'(\mathrm{S}|\mathrm{E}, \neg\mathrm{D}) = 1$ and, as before, $P'(\mathrm{B}) = 1$.
- Then the following theorem holds:

Theorem: *Consider the Bayesian Network in Figure 7 with a prior probability distribution. Let*

$$k_d := \frac{p_1\, p_2}{q_1\, p_2 + (\overline{q_1} - d)\, q_2}.$$

We furthermore assume that (i) the posterior probability distribution P' is defined over the same Bayesian Network, (ii) the learned information is modeled as constraints on P', and (iii) P' minimizes the Kullback-Leibler divergence to P. Then $P'(\mathrm{E}) > P(\mathrm{E})$, iff $k_d > 1$. Moreover, if $k_d > 1$ and $p_2 > q_2$, then $P'(\mathrm{D}) < P(\mathrm{D})$.

V. Conclusions

Summary

1. We have proposed a unified account for the learning of indicative causal conditionals in a Bayesian framework.
2. Intuitively correct results obtain if we
   - (i) represent the causal relations between all relevant variables in a Bayesian Network,
   - (ii) consider the learned conditional as a conditional probability constraint on the posterior distribution, and
   - (iii) determine the posterior distribution by minimizing the KL divergence between the posterior and the prior distribution.

Summary

3. The proposed account goes beyond standard Bayesianism where learning is modeled as conditionalization (or Jeffrey conditionalization). This turned out to be necessary as the standard account fails in the considered cases.
4. Moreover, conditionalization follows from our proposed procedure if a piece of evidence E is learned ($P'(\mathrm{E}) = 1$ or < 1 for Jeffrey conditionalization).

Why Accepting the Proposed Account

Worry: Our account misses a top-down (axiomatic, Dutch book etc.) justification. However:

1. It is non-trivial that our account provides the right answers in all considered cases.
2. In some cases it also forced us to reconsider our intuitions.
3. There is no other account that achieves this.
4. Not fixing the correct causal structure leads to wrong results.

Upshot: Taken together, these points provide a strong justification for the proposed account.

Open Questions

1. Using other "distances" (e.g. the Helling distance). Which distance is right will to some extent be an empirical question.
2. Formulating an account that also works for non-causal conditionals (as in the Judy Benjamin example).
3. Formulating an account that also works for nested conditionals such as A $\rightarrow$ (B $\rightarrow$ C)
4. ...

Thanks for your attention!

The talk is based on joint work with
Soroush Rafiee Rad (ILLC Amsterdam).