Title: Argumentation, Conditionals, and the Use of Information Theoretic Concepts in Bayesianism

Date: Apr 13, 2018 10:00 AM

URL: http://pirsa.org/18040121

Abstract: In this talk, I show how information theoretic concepts can be used to extend the scope of traditional Bayesianism. I will focus on the learning of indicative conditionals (â€œIf A, then Bâ€•) and a Bayesian account of argumentation. We will see that there are also interesting connections to research done in the psychology of reasoning. The talk is partly based on the paper â€œBayesian Argumentation and the Value of Logical Validityâ€• (with Ben Eva, forthcoming in Psychological Review, http://philsci-archive.pitt.edu/14491/).

Argumentation, Conditionals, and the Use of Information Theoretic Concepts in Bayesianism

Stephan Hartmann

Munich Center for Mathematical Philosophy

Algorithmic Information, Induction and Observers in Physics
Perimeter Institute, April 2018

Motivation

- Bayesianism is the leading theory of uncertain reasoning with many applications in artificial intelligence, philosophy, statistics, and other sciences. (Like established disciplines such as mechanics and electrodynamics, it has even been quantized.)
- I am interested in developing the Bayesian framework and study applications to old and new philosophical problems.
- In particular, I am interested in understanding how the learning of evidence which may not be propositional can be modeled. In this case (Jeffrey) Conditionalization cannot be applied and new ways of dealing with the problem have to be found.
- I propose such a method (which is inspired by information theory) and relate it to discussions in cognitive science.
- The proposal also raises some challenges to information theory, but more about this later.
- I also hope that there are applications in physics.

Argumentation

- Argumentation plays an important role in science as well as in every day life.
- While some arguments are good and convincing, others are misleading or simply do not work.
- It is therefore desirable to have a general theory that helps us to distinguish good arguments from bad arguments.
- So far there is no such theory although it is widely acknowledged that classical logic is too limited. But there is no consensus about how to proceed.
- I will show how the research program on Bayesian Argumentation, which was initiated by the cognitive scientists Ulrike Hahn and Mike Oaksford, can be further developed.

Overview

1. Bayesian Argumentation 1.0
2. Learning (Indicative) Conditionals
3. The Distance-Based Approach to Bayesianism
4. Bayesian Argumentation 2.0
5. Outlook

What is an Argument?

- An argument is a set of statements ("premises") to support another statement ("conclusion").
- An argument is the better, the more the premises support the conclusion.
- But what are the principles for assessing arguments? The following factors play a role here:
  1. The structure of the argument: is the argument logically valid?
  2. The content of the argument: are the premises plausible? How plausible is the conclusion initially?
  3. The source that makes the argument: is the source reliable?

Argument Schemes

- Affirming the Consequent (AC)

$$\frac{\begin{array}{c} C \\ A \rightarrow C \end{array}}{A}$$

- Denying the Antecedent (DA)

$$\frac{\begin{array}{c} \neg A \\ A \rightarrow C \end{array}}{\neg C}$$

Good and Bad Arguments

- AC and DA are often called fallacies. They are not logically valid.
- However, logically invalid arguments can nevertheless be good arguments in the sense that they can raise the probability of the conclusion.
- This suggests that we study argumentation from a probabilistic point of view.
- To do so, we need a framework theory which allows us to model the process of argumentation.
- We choose Bayesianism because (i) it is the simplest and (ii) most developed one. Besides, (iii) it has a normative foundation and (iv) comes with the powerful machinery of Bayesian Networks which will help us to represent beliefs and to effectively compute the change of belief that a good argument induces.

Argument Schemes

- Affirming the Consequent (AC)

$$\frac{\begin{array}{c} C \\ A \to C \end{array}}{A}$$

- Denying the Antecedent (DA)

$$\frac{\begin{array}{c} \neg A \\ A \to C \end{array}}{\neg C}$$

Good and Bad Arguments

- AC and DA are often called fallacies. They are not logically valid.
- However, logically invalid arguments can nevertheless be good arguments in the sense that they can raise the probability of the conclusion.
- This suggests that we study argumentation from a probabilistic point of view.
- To do so, we need a framework theory which allows us to model the process of argumentation.
- We choose Bayesianism because (i) it is the simplest and (ii) most developed one. Besides, (iii) it has a normative foundation and (iv) comes with the powerful machinery of Bayesian Networks which will help us to represent beliefs and to effectively compute the change of belief that a good argument induces.

Bayesian Argumentation: The Idea

- Consider an agent who entertains the propositions A and B. Let P be a prior probability distribution defined over the corresponding propositional variables A and B.
- Someone else then makes a MP argument (I work with these simple schemes for illustrative purposes) and states:

$$\begin{array}{c} \mathsf{A} \\ \mathsf{A} \rightarrow \mathsf{C} \\ \hline \mathsf{C} \end{array}$$

- Here the agent learns the premises A and A $\rightarrow$ C and updates her (partial) beliefs accordingly.
- Representing the conditional A $\rightarrow$ C by the material conditional $\neg \mathrm{A} \vee \mathrm{C}$ and conditionalizing on both propositions yields

$$P^*(\mathrm{C}) := P(\mathrm{C}|\mathrm{A}, \neg \mathrm{A} \vee \mathrm{C}) = 1.$$

- Slogan: "Argumentation is learning"

The Value of Logical Validity

- If the propositions $\phi_1, \phi_2 \ldots, \phi_n$ are 'prior premises' and ϕ is a logical consequence of them, then the probability of ϕ increases if the probability of at least one of the premises increases (and no probability decreases).
- This is a special case of the Uncertainty Sum Rules found by Adams.
- Define the uncertainty of a proposition as $u(\phi) := 1 - P(\phi)$.

1. Statics

If $\phi_1, \phi_2 \ldots, \phi_n$ are 'prior premises' and ϕ is a logical consequence of $\phi_1, \phi_2 \ldots, \phi_n$, then $u(\phi) \leq u(\phi_1) + \cdots + u(\phi_n)$.

2. Dynamics

If $\phi_1, \phi_2 \ldots, \phi_n$ are 'prior premises', ι is a new premise, and ϕ is a logical consequence of $\phi_1, \phi_2 \ldots, \phi_n$ and ι, then $u(\phi) \leq u(\phi_1|\iota) + \cdots + u(\phi_n|\iota)$.

Bayesian Argumentation: Generalization

This ideas presented so far can be generalized:

1. An agent represents her beliefs as a Bayesian Network with a prior probability distribution P defined over it.
2. She then learns the premises of an argument from another agent and updates her beliefs using (Jeffrey) conditonalization (leaving the network structure unchanged).
3. The procedure can also be applied to other argument schemes, e.g. to new argument schemes used in science.

Bayesian Argumentation: Two Problems

1. What if the conditional is non-extreme, i.e. if the conditional is not learnt with certainty? Can it still be represented as a material conditional?
2. Does the information gathering process matter? If so, how can it be modeled?

In this talk, I focus on the first problem. In joint work with Peter Collins, Ulrike Hahn, Karolina Krzyzanowska and Greg Wheeler, we address the second problem.

II. Learning (Indicative) Conditionals

Representing Indicative Conditionals

- An agent considers two propositions, A and C, and has a prior probability distribution P defined over the corresponding propositional variables.
- She then learns the (ordinary language) indicative conditional "If A, then C" (which we denote by $\mathrm{A} \rightarrow \mathrm{C}$).
- Questions:
  1. How should she update her beliefs?
  2. Do conditionals have a probability at all?
  3. Are indicative conditionals propositions?
- To start with, let us consider

Stalnaker's Thesis

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

- Using this thesis, David Lewis proofed a famous triviality result.

Lewis' Triviality Result

- The calculation proceeds as follows:

$$\begin{aligned} P(\mathrm{A} \to \mathrm{C}) &= P(\mathrm{A} \to \mathrm{C}|\mathrm{C}) \cdot P(\mathrm{C}) + P(\mathrm{A} \to \mathrm{C}|\neg\mathrm{C}) \cdot P(\neg\mathrm{C}) \\ &= P(\mathrm{C}|\mathrm{A},\mathrm{C}) \cdot P(\mathrm{C}) + P(\mathrm{C}|\mathrm{A},\neg\mathrm{C}) \cdot P(\neg\mathrm{C}) \\ &= 1 \cdot P(\mathrm{C}) + 0 \cdot P(\neg\mathrm{C}) \\ &= P(\mathrm{C}) \end{aligned}$$

- Here we have used the Law of Total Probability (line 1), Stalnaker's Thesis (line 2) and the assumption that $P(\mathrm{A},\mathrm{C}), P(\mathrm{A},\neg\mathrm{C}) > 0$.
- From this, people concluded that we should not assign probabilities to conditionals and that conditionals are not propositions. This is a radical conclusion that has many unwanted consequences. It should only be given up if there is no way out.
- To proceed, let us explore how we can model the learning of a conditional (which does not necessarily presuppose that we assign a probability to it) by representing the conditional by the material conditional $\mathrm{A} \supset \mathrm{C} = \neg\mathrm{A} \vee \mathrm{C}$.

Representing Indicative Conditionals

- An agent considers two propositions, A and C, and has a prior probability distribution P defined over the corresponding propositional variables.
- She then learns the (ordinary language) indicative conditional "If A, then C" (which we denote by $\mathrm{A} \rightarrow \mathrm{C}$).
- Questions:
  1. How should she update her beliefs?
  2. Do conditionals have a probability at all?
  3. Are indicative conditionals propositions?

- To start with, let us consider

Stalnaker's Thesis

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

- Using this thesis, David Lewis proofed a famous triviality result.

Lewis' Triviality Result

- The calculation proceeds as follows:

$$\begin{aligned} P(\mathrm{A} \to \mathrm{C}) &= P(\mathrm{A} \to \mathrm{C}|\mathrm{C}) \cdot P(\mathrm{C}) + P(\mathrm{A} \to \mathrm{C}|\neg\mathrm{C}) \cdot P(\neg\mathrm{C}) \\ &= P(\mathrm{C}|\mathrm{A},\mathrm{C}) \cdot P(\mathrm{C}) + P(\mathrm{C}|\mathrm{A},\neg\mathrm{C}) \cdot P(\neg\mathrm{C}) \\ &= 1 \cdot P(\mathrm{C}) + 0 \cdot P(\neg\mathrm{C}) \\ &= P(\mathrm{C}) \end{aligned}$$

- Here we have used the Law of Total Probability (line 1), Stalnaker's Thesis (line 2) and the assumption that $P(\mathrm{A},\mathrm{C}), P(\mathrm{A},\neg\mathrm{C}) > 0$.
- From this, people concluded that we should not assign probabilities to conditionals and that conditionals are not propositions. This is a radical conclusion that has many unwanted consequences. It should only be given up if there is no way out.
- To proceed, let us explore how we can model the learning of a conditional (which does not necessarily presuppose that we assign a probability to it) by representing the conditional by the material conditional $\mathrm{A} \supset \mathrm{C} = \neg\mathrm{A} \vee \mathrm{C}$.

Representing Indicative Conditionals

- An agent considers two propositions, A and C, and has a prior probability distribution P defined over the corresponding propositional variables.
- She then learns the (ordinary language) indicative conditional "If A, then C" (which we denote by $\mathrm{A} \to \mathrm{C}$).
- Questions:
  1. How should she update her beliefs?
  2. Do conditionals have a probability at all?
  3. Are indicative conditionals propositions?
- To start with, let us consider

Stalnaker's Thesis

$P(\mathrm{A} \to \mathrm{C}) = P(\mathrm{C}|\mathrm{A})$.

- Using this thesis, David Lewis proofed a famous triviality result.

Lewis' Triviality Result

- The calculation proceeds as follows:

$$\begin{aligned} P(\mathrm{A} \to \mathrm{C}) &= P(\mathrm{A} \to \mathrm{C}|\mathrm{C}) \cdot P(\mathrm{C}) + P(\mathrm{A} \to \mathrm{C}|\neg\mathrm{C}) \cdot P(\neg\mathrm{C}) \\ &= P(\mathrm{C}|\mathrm{A},\mathrm{C}) \cdot P(\mathrm{C}) + P(\mathrm{C}|\mathrm{A},\neg\mathrm{C}) \cdot P(\neg\mathrm{C}) \\ &= 1 \cdot P(\mathrm{C}) + 0 \cdot P(\neg\mathrm{C}) \\ &= P(\mathrm{C}) \end{aligned}$$

- Here we have used the Law of Total Probability (line 1), Stalnaker's Thesis (line 2) and the assumption that $P(\mathrm{A},\mathrm{C}), P(\mathrm{A},\neg\mathrm{C}) > 0$.
- From this, people concluded that we should not assign probabilities to conditionals and that conditionals are not propositions. This is a radical conclusion that has many unwanted consequences. It should only be given up if there is no way out.
- To proceed, let us explore how we can model the learning of a conditional (which does not necessarily presuppose that we assign a probability to it) by representing the conditional by the material conditional $\mathrm{A} \supset \mathrm{C} = \neg\mathrm{A} \vee \mathrm{C}$.

Learning Indicative Conditionals

- An agent considers two propositions, A and C, and has a prior probability distribution P defined over the corresponding propositional variables.
- She then learns the indicative conditional "If A, then C".
- Representing the conditional by the material conditional $\neg \mathrm{A} \vee \mathrm{C}$, once can then show that
  1. $P^*(\mathrm{A}) := P(\mathrm{A}|\neg \mathrm{A} \vee \mathrm{C})$ decreases, and that
  2. $P^*(\mathrm{C}) := P(\mathrm{C}|\neg \mathrm{A} \vee \mathrm{C})$ increases,

  if the probability distribution P is not extreme.
- These results are plausible: Once we learn the conditional, the antecedent becomes more informative (and therefore less likely).
- Note that all this only holds for two propositions and without any additional constraints.

Alleged Problems: Douven's Counterexamples

- Igor Douven presented a number of counterexamples against conditonalizing on the material conditional.
- Here is one of them (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.

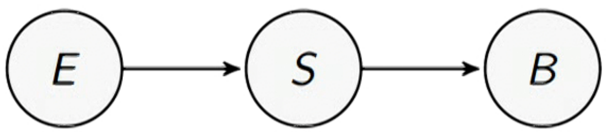
Learning Indicative Conditionals

- An agent considers two propositions, A and C, and has a prior probability distribution P defined over the corresponding propositional variables.
- She then learns the indicative conditional "If A, then C".
- Representing the conditional by the material conditional $\neg\mathrm{A} \vee \mathrm{C}$, once can then show that
  1. $P^*(\mathrm{A}) := P(\mathrm{A}|\neg\mathrm{A} \vee \mathrm{C})$ decreases, and that
  2. $P^*(\mathrm{C}) := P(\mathrm{C}|\neg\mathrm{A} \vee \mathrm{C})$ increases,

  if the probability distribution P is not extreme.
- These results are plausible: Once we learn the conditional, the antecedent becomes more informative (and therefore less likely).
- Note that all this only holds for two propositions and without any additional constraints.

Alleged Problems: Douven's Counterexamples

- Note that there are three variables involved here: E (with the values "Sue passes the exam" etc.), S ("Sue is invited on a ski trip" etc.), and B ("Sue buys a skiing outfit" etc.).
- Harry has prior beliefs about these propositions and then learns two items of information: (i) B and (ii) "If E, then S".
- Conditionalizing on B and $\neg \mathrm{E} \vee \mathrm{S}$, one can then show that the probability of E increases. This becomes especially clear if one additionally makes the assumption that $E \perp\!\!\!\perp B | S$.



- All other "counterexamples" (Sundowners, Driving Test, etc.) can be dealt with analogously.

Problems (as before)

1. What if the conditional is non-extreme, i.e. if the conditional is not learnt with certainty (as, e.g., in the Judy Benjamin problem)?
2. Does the information gathering process matter? If so, how can it be modeled?

To address the first problem, we develop the distance-based approach to Bayesianism.

III. The Distance-Based Approach to Bayesianism

Repetition: Conditionalization ("Bayes' Rule")

- There are various rules that specify how a Bayesian agent should change her beliefs in a rational way if she learns a proposition. Most prominently, there is

Conditionalization

$$P^*(\mathrm{H}) := P(\mathrm{H}|\mathrm{E}) = \frac{P(\mathrm{E}|\mathrm{H}) \cdot P(\mathrm{H})}{P(\mathrm{E})}.$$

- Conditionalization applies if the evidence becomes certain (i.e. if $P^*(\mathrm{E}) = 1$).

Repetition: Jeffrey Conditionalization

Jeffrey Conditionalization

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

- Jeffrey Conditionalization applies if the evidence remains uncertain, e.g. if it increases from $P(\mathrm{E}) = .4$ to $P^*(\mathrm{E}) = .9$.
- Jeffrey Conditionalization follows from the Law of Total Probability,

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

if the following condition holds:

Rigidity Condition

$$P'(\mathrm{H}|\mathrm{E_i}) = P(\mathrm{H}|\mathrm{E_i})$$

for all elements of a partition of E.

The Distance-Based Approach

Question: What can be done if the learned evidence is not propositional?

- An agent holds beliefs, represented by a probability distribution P.
- She then learns some new information which poses a constraint on the new probability distribution Q.
- Q is then found by minimizing a "distance" between Q and P, i.e. we assume that the agents wants to change her beliefs in a conservative way.
- This approach requires the specification of a distance measure and we will see that there are several candidate measures (such as the Kullback-Leibler divergence).
- We request that any admissible distance measure should imply (Jeffrey) Conditionalization.
- It can then be shown that this approach can be applied to the learning of other kinds of evidence (such as indicative conditionals, structural evidence).

The Kullback-Leibler Divergence

Kullback-Leibler Divergence between Q and P

Let $\mathrm{S}_1, \ldots, \mathrm{S}_n$ be the possible values of a random variable S over which probability distributions P and Q are defined and let $p_i := P(\mathrm{S_i})$ and $q_i := Q(\mathrm{S_i})$. Then:

$$D_{KL}(Q||P) := \sum_{i=1}^{n} q_i \cdot \log\left(\frac{q_i}{p_i}\right) .$$

Note that the KL divergence is not symmetrical and that it may not satisfy the triangle inequality. So it is not a distance measure in the mathematical sense of the term.

The f-divergence

The f-divergence is a generalization of the KL divergence:

f-Divergence (Csiszár 1967)

Let f be a convex function with $f(1) = 0$ and $p_i := P(\mathrm{S_i})$ and $q_i := Q(\mathrm{S_i})$. Then

$$D_f(P'||P) := \sum_{i=1}^{n} p_i \cdot f\left(\frac{q_i}{p_i}\right).$$

- Note that the KL divergence results for $f(t) = t \log t$.
- Other measures, such as the inverse KL divergence ($f(t) = -\log t$), the Hellinger distance and the χ^2-divergence are also f-divergences.
- Important: All f-divergences imply (Jeffrey) Conditionalization.

Learning (Indicative) Conditionals

- If one learns the indicative conditional "If A, then C" from a perfectly reliable source, then the constraint on Q is $Q(\mathrm{C}|\mathrm{A}) = 1$.
- Note that $Q(\mathrm{C}|\mathrm{A}) = 1$ iff $Q(\neg\mathrm{A} \vee \mathrm{C}) = 1$ (if $Q(\mathrm{A}) > 0$) because

$$Q(\neg\mathrm{A} \vee \mathrm{C}) = Q(\mathrm{A}) + Q(\neg\mathrm{A})\, Q(\mathrm{C}|\mathrm{A}).$$

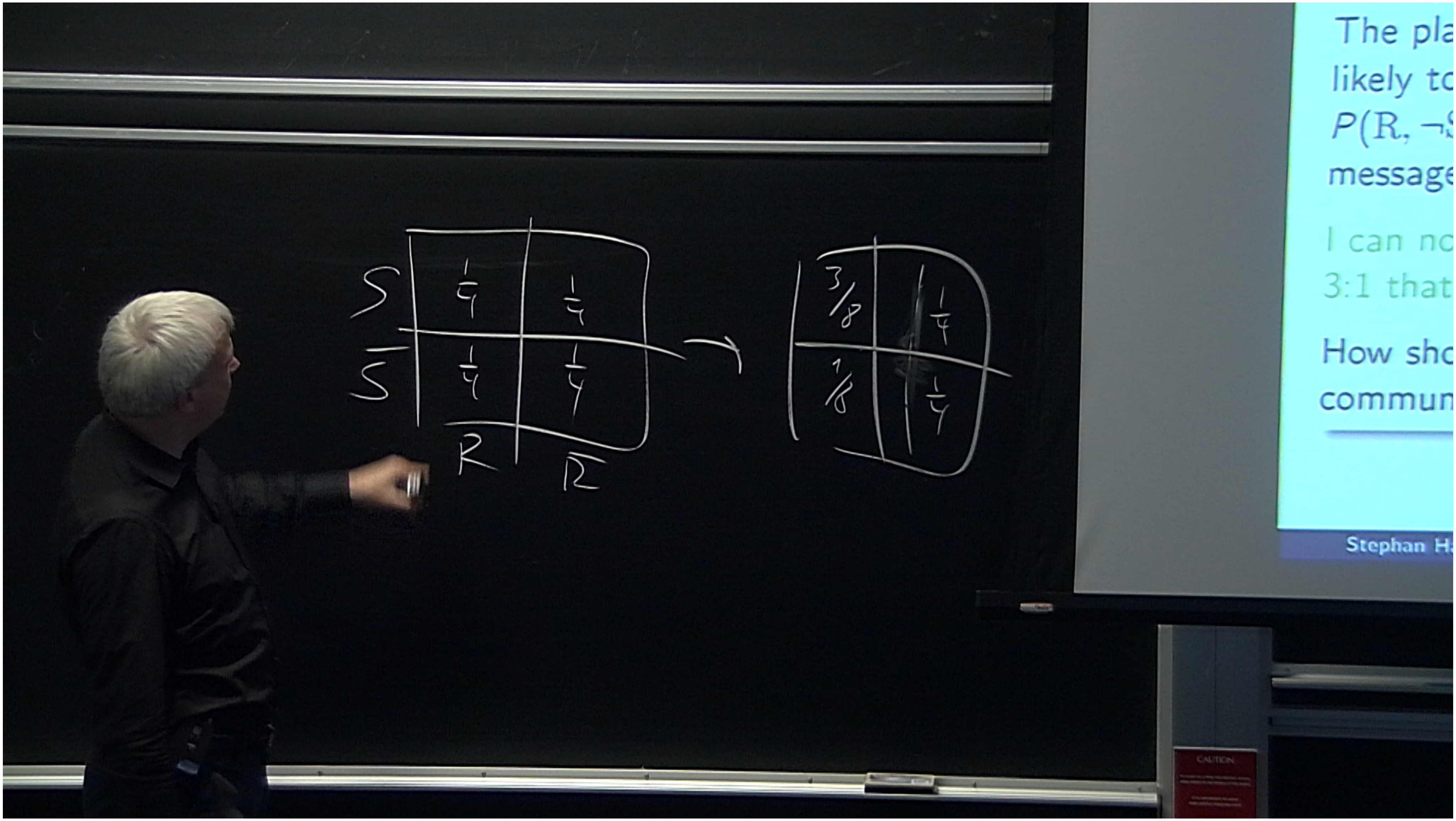
- It is therefore not surprising that one gets the same results for conditonalizing on the material conditional and for minimizing an f-divergence if the learned conditional is strict.
- However, if the learned conditional is not strict, then one gets different results for both procedures (and for different divergencies).
- If one additionally accepts the norm Minimizing Inaccuracy (as in Epistemic Utility Theory), then the KL-divergence is the only divergence left (logarithmic scoring rule).

Illustration: The 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. $P(\mathrm{R}, \mathrm{S}) = P(\mathrm{R}, \neg\mathrm{S}) = P(\neg\mathrm{R}, \mathrm{S}) = P(\neg\mathrm{R}, \neg\mathrm{S}) = 1/4$. 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? (van Fraassen 1981)



S
$\frac{1}{4}$
$\frac{1}{4}$
$\bar{S}$
$\frac{1}{4}$
$\frac{1}{4}$
R
$\bar{R}$
3/8
$\frac{1}{4}$
1/8
$\frac{1}{4}$
The pla
likely to
P(R, ¬S
message
I can no
3:1 that
How sho
commun
Stephan H

Illustration: The Judy Benjamin Problem

- We translate the learnt conditional into the following constraint on the new distribution Q: $Q(\mathrm{S}|\mathrm{R}) = k$, where $k = n/(n+1)$ if the odds are $n:1$. (In the present case, $k = 3/4$.)
- Minimizing the KL divergence then yields $Q(\mathrm{R}) < 1/2$, which contradicts the intuition many people have.
- It has been noted, however, that as $n \to \infty$, $Q(\mathrm{R}, \neg\mathrm{S}) \to 0$. Hence, of the four original quarters only three remain. Learning the conditional excludes one of four possibilities. Using the Principle of Indifference, it therefore seems rational to assign a probability of $1/3$ to each of the remaining quarters. Hence, in the limit, $Q(\mathrm{R}) = 1/3$. This limiting value is reached as n increases and it is therefore reasonable to have $Q(\mathrm{R}) < 1/2$ for $n = 3$.

Illustration: The Judy Benjamin Problem

A Theorem

Minimizing an f-divergence between Q and P for the Judy Benjamin Problem yields $q_1 = k\,\delta, q_2 = \overline{k}\,\delta$ and $q_3 = q_4 = \overline{\delta}/2$ with $\delta \in (0,1)$ The value of δ depends on the respective f-divergence: $\delta_{IKL} = 1/2$,

$$\delta_{KL} = \frac{1}{1 + 2\,k^k\,\overline{k}^{\overline{k}}}, \quad , \quad \delta_H = \frac{1 + 2\sqrt{k\,\overline{k}}}{3 + 2\sqrt{k\,\overline{k}}} \quad , \quad \delta_{\chi^2} = \frac{1}{3 - 4\,k\,\overline{k}}.$$

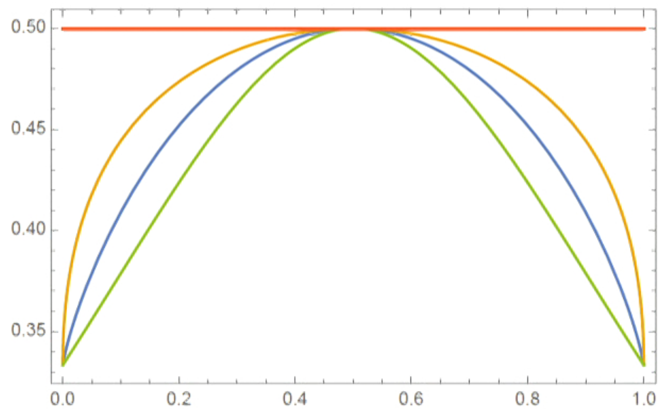


Illustration: The Judy Benjamin Problem

- But what about other f-divergencies? The plot shows that all f-divergencies apart from the IKL divergence yield $Q(\mathrm{R}) < 1/2$.
- Only the IKL divergence yields $Q(\mathrm{R}) = 1/2$.
- However, one might want to argue that we do not only learn (i) $Q(\mathrm{S}|\mathrm{R}) = k$, but also (ii) $Q(\mathrm{R}) = P(\mathrm{R})$. If one translates the learned information in these two constraints on Q, then all $f-$divergencies yield $Q(\mathrm{R}) = 1/2$.
- To conclude, it may not always be uncontroversial which constraints on Q follow from the information which is provided.

IV. Bayesian Argumentation 2.0

The Main Idea: "Argumentation is Learning"

1. An agent entertains beliefs about a set of propositions, represented by a Bayesian Network with a prior probability distribution P defined over it.
2. She then learns the premises of an argument from another agent. The network structure is not changed as a result of the learning experience. (There could, however, be information that may result in a change of the network structure.)
3. This poses constraints on the new probability distribution Q.
4. The full distribution Q is then determined by minimizing an f-divergence (such as the KL-divergence) between Q and P.
5. As a result, the probability of the conclusion changes.

Illustration 1: MP, Certain Premises

$$\frac{\begin{array}{c}\mathsf{A} \\ \mathsf{A} \to \mathsf{B}\end{array}}{\mathsf{B}}$$

- The agent has beliefs about the propositions A and B. These beliefs are represented by a probability function P.
- The agent learns from a perfectly reliable information source that
  - A
  - A $\to$ B.
- The learned information puts constraints on the new distribution Q:
  - A: $Q(\mathrm{A}) = 1$
  - A $\to$ B: $Q(\mathrm{B}|\mathrm{A}) = 1$.
- In this case, the full probability distribution Q results from the constraints an no minimization is neccessary. We obtain, as expected, $Q(\mathrm{B}) = 1$.

Illustration 1: DA, Certain Premises

$$\frac{\begin{array}{c}\neg A \\ A \rightarrow B\end{array}}{\neg B}$$

- The agent has beliefs about the propositions A and B. These beliefs are represented by a probability function P.
- The agent learns from a perfectly reliable information source that
  - $\neg A$
  - $A \rightarrow B$.
- The learned information puts constraints on the new distribution Q:
  - $\neg A$: $Q(\neg A) = 1$
  - $A \rightarrow B$: $Q(B|A) = 1$.
- The agent then minimizes an f-divergence between Q and P and obtains in accordance with other approaches that $Q(\neg B) = P(\neg B|\neg A)$.

Illustration 2: MP, Uncertain Minor Premise

$$\frac{\begin{array}{c}\text{A}\\ \text{A} \rightarrow \text{B}\end{array}}{\text{B}}$$

- The agent has beliefs about the propositions A and B. These beliefs are represented by a probability function P.
- The agent learns that A $\rightarrow$ B.
- Furthermore, the agent learns from a partially reliable information source that A.
- The learned information puts constraints on the new distribution Q:
  - A: $Q(\mathrm{A}) > P(\mathrm{A})$
  - A $\rightarrow$ B: $Q(\mathrm{B}|\mathrm{A}) = 1$.
- The agent then minimizes an f-divergence between Q and P and obtains that $Q(\mathrm{B}) > P(\mathrm{B})$.

Illustration 3: Uncertain Mayor Premise

$$\frac{\begin{array}{c}\text{A} \\ \text{A} \rightarrow \text{B}\end{array}}{\text{B}}$$

- The agent has beliefs about the propositions A and B. These beliefs are represented by a probability function P.
- The agent learns that A and that A $\rightarrow$ B.
- The learned information puts constraints on the new distribution Q:
  - A: $Q(\mathrm{A}) = P(\mathrm{A})$
  - A $\rightarrow$ B: $Q(\mathrm{B}|\mathrm{A}) > P(\mathrm{B}|\mathrm{A})$.
- The agent then minimizes an f-divergence between Q and P and obtains that $Q(\neg\mathrm{B}) > P(\neg\mathrm{B})$.
- A similar result obtains for MT, but not for the invalid schemes AC and DA.

A Conjecture

Generalizing from our findings, the following conjecture emerges:

(i) If the underlying argument pattern is logically valid, then the probability of the conclusion (following the procedure described above) always increases if the probability of at least one of the premises increases (and the probability of no other premise goes down).

(ii) For arguments which are not logically valid, this is not the case.

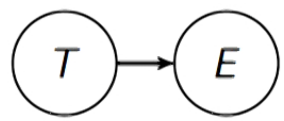
Proving a refinement of this generalization of Adams' sum rules is a major challenge.

Upshot

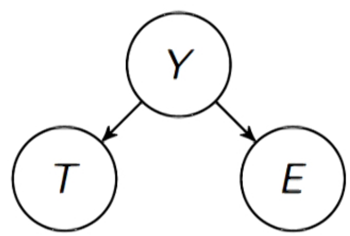
- Argumentation is a dialogical endeavor: One agent (agent 1) wants to convince another agent (agent 2) of a conclusion.
- Agent 1 does this in an indirect way by making the premises of the argument more likely. As a result (and to make sure that the beliefs remain coherent), agent 2 changes her beliefs also about the conclusion.
- The problem is that agent 1 does not know the prior probability distribution of agent 2.
- However, if agent 1 uses a logically valid argument pattern such as MP or MT, then she can be certain that agent 2 will raise the probability of the conclusion. This does not hold for AC and DA.

Argumentation Beyond MP and MT

- Scientific argumentation (as well as ordinary argumentation) is much more than MP and MT.
- An indeed, scientists come up with new argument schemes which they use to convince each other.
- The corresponding arguments typically involve a fair amount of uncertainty about the truth of the premises.
- What is more, they are often indirect in a specific sense.
- Direct confirmation:



- Indirect confirmation (involving a "common cause"):



V. Outlook

Outlook

- We have sketched a unified Bayesian theory of argumentation.
- The key idea is that argumentation is learning.
- I have offered a conjecture which answers the question why logical validity is a valuable ingredient to argumentation even if the premises are uncertain.
- Along the way, I have sketched the Distance-Based Approach to Bayesian and argued that it has many applications in the psychology of reasoning and beyond.
- In future work, we plan to (i) find other normative constraints that single out subclasses of f-divergencies and (ii) conduct experiments and see how well different f-divergencies do empirically.
- I also hope for further applications in information theory and physics.

References

1. Collins, P., K. Krzyzanowska, S. Hartmann, G. Wheeler and U. Hahn: Conditionals and Testimony. Under review.
2. Dawid, R., S. Hartmann and J. Sprenger: The No Alternatives Argument. *The British Journal for the Philosophy of Science* 66(1): 213–234 (2015).
3. Dardashti, R., S. Hartmann, K. Thébault and E. Winsberg: Hawking Radiation and Analogue Experiments: A Bayesian Analysis. Under review. (http://philsci-archive.pitt.edu/14234/)
4. Eva, B. and S. Hartmann: Bayesian Argumentation and the Value of Logical Validity. Forthcoming in *Psychological Review*. (http://philsci-archive.pitt.edu/14491/)
5. Eva, B., S. Hartmann and S. Rafiee Rad: Learning from Conditionals. Under review.
6. Sprenger, J. and S. Hartmann: *Bayesian Philosophy of Science*. To appear with Oxford University Press.

Thanks. . .

. . . for your attention!