Bayesian Networks in Philosophy of Science

Stephan Hartmann

Munich Center for Mathematical Philosophy
LMU Munich

Arizona State University (Zoom)

March 29, 2021

Motivation

- How I got interested in Bayesian networks:
  1. Problems of naturalized philosophy of science (limitations of the case study method, lack of a normative framework)
  2. Problems of traditional formal philosophy of science (too far away from real science, of not much use for science)
- What is so good about Bayesian networks in philosophy?
  They are simple, intuitive, come with a powerful mathematical machinery, are closely connected to statistics and cognitive science, embedded in a normative framework ("Bayesianism"), and allow for the analysis of less idealized epistemic scenarios,. . .
- Bayesian networks are a philosophical modeling tool which has many applications ("success stories").
- However, Bayesian networks are not only useful in periods of philosophical normal science. They also help in addressing a number of challenges to an established philosophical paradigm – Bayesianism.

Overview

1. Standard Bayesianism
2. Problem 1: Non-Empirical Evidence
3. Problem 2: New Types of Evidence
4. Problem 3: Genuinely New Evidence
5. Outlook

I. Standard Bayesianism

Bayesianism

- Bayesianism is a theory about the statics and dynamics of (partial) beliefs.
- Its starting point is the psychological truism that we believe different (contingent) propositions more or less strongly: we assign different degrees of belief to them.
- To make the concept "degree of belief" more precise, we need:
  (i) a calculus to combine different degrees of belief,
  (ii) an algorithm to update degrees of belief, and
  (iii) a normative justification for (i) and (ii).
- Bayesianism provides exactly this.
- Let's see how this works for the static and the dynamic part.

Bayesianism: Statics

- Bayesianism identifies degrees of belief with probabilities, i.e. the (rational) degrees of belief of an agent at a certain time have to satisfy the axioms of probability theory.
- What justifies this identification?
  1. Pragmatic arguments ("Dutch book arguments"): Show that if one has incoherent degrees of belief, i.e. degrees of belief which are not consistent with the probability calculus, then one will loose money in a betting scenario.
  2. Epistemic arguments: Identifying degrees of belief with probabilities makes sure that the expected inaccuracy of our beliefs is minimized ("Epistemic Utility Theory").

Bayesianism: Dynamics

- An agent considers a number of propositions. They are represented by an algebra $\mathcal{A}$ which comprises the propositional variables $A_1, \ldots, A_n$. A prior probability distribution P is defined accordingly.
- She then learns that $\mathrm{E} = \mathrm{A}_1$ is the case. This prompts her to switch from P to the posterior distribution P' which satisfies $P'(\mathrm{E}) = 1$. To make sure that her new degrees of belief are coherent, she applies

Bayes Theorem ("Conditionalization")

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

- If the evidence is not certain and the rigidity condition holds, then Conditionalization generalizes to Jeffrey Conditionalization.
- There are pragmatic and epistemic arguments that justify the use of Conditionalization. These arguments are, however, more controversial than in the static case.

- Confirmation theory is the most important (though by far not the only) application of Bayesianism in philosophy.
- Typical scenario: The evidence E is direct, i.e. it is a deductive or inductive consequence of the tested hypothesis H. In this case, the probabilistic relationship between the corresponding propositional variables H and E can be represented by the following Bayesian network (with a probability distribution P defined over it):

$$H \longrightarrow E$$

- We can then calculate $P(\mathrm{H})$ and $P'(\mathrm{H}) = P(\mathrm{H}|\mathrm{E})$ (if the evidence becomes certain) and test whether E confirms H or not (i.e. whether $P'(\mathrm{H}) > P(\mathrm{H})$ or not).
- One can also investigate more complicated scenarios. Note, though, that in all these cases the evidence is typically assumed to be direct.

Bayesianism and its Problems

- Bayesianism has many successful applications in philosophy and the sciences beyond confirmation theory.
- At the same time, Bayesianism faces a number of problems.
- Some of these problems are modeling challenges (such as the problem of old evidence); others may point to a better theory beyond Bayesianism (such as the possible failure to represent ignorance).
- I am not too concerned about these problems. Bayesianism should be treated like a scientific theory (and not like more), and all scientific theories have problems and face anomalies. Likewise, the successful application of scientific theories justifies (at least) their pragmatic use.
- Having said this, I will now identify three further problems and propose strategies to solve them.

1. Non-Empirical Evidence

Standard Bayesianism assumes that the evidence is direct in the sense that it is a deductive or inductive consequence of the theory under consideration. However, this may not always be the case. Some evidence may be non-empirical.

- There is a lively debate in the methodology of science whether there is non-empirical evidence for a scientific theory.
- Such confirmation would be helpful in many areas of fundamental physics where direct empirical evidence is scarce or non-existent. Here are two examples:
  1. The No Alternatives Argument: Does the observation that scientists have not yet found an alternative to string theory (despite a lot of effort and brain power) confirm the theory?
  2. Analogue Simulation: Is it possible to confirm a claim about an empirically unaccessible phenomenon (such as black hole Hawking radiation) by experimenting on a different physical system (such as a Bose Einstein condensate)?

1. Non-Empirical Evidence

- Note that in both cases the alleged evidence is indirect (and in this sense non-empirical).
- It would be fantastic to have such ways of indirect confirmation, but is it possible?
- And if it is in principle possible to have non-empirical confirmation, under which conditions is it possible?
- Note that theories such as HD confirmation or Popper's falsificationism dismiss this alleged evidence from the beginning.
- This inference is too quick: While it may well turn out that there is no non-empirical evidence, it should not be part of the basic assumptions of the confirmation (or corroboration) theory one uses. Such theories are not useful for understanding contemporary science.

- We will see that Bayesianism has the resources to analyze and reconstruct confirmation scenarios involving non-empirical evidence.

2. New Types of Evidence

Standard Bayesianism assumes that the evidence is propositional. However, there may be evidence which does not prompt a shift of the probability of one of the propositions in the algebra. Some evidence may be non-propositional.

- It seems to be too restrictive to only consider propositional evidence. There may be many other types of learning experiences that prompt us to change our beliefs. How can they be modeled in the Bayesian framework?
- Here are two examples:
  1. Structural Evidence: The agent may learn e.g. that the underlying causal network of a set of propositions is such and such.
  2. Indicative Conditionals: The agent may learn an indicative conditional of the form "If A, then C".

2. New Types of Evidence

- It is not clear whether these learning experiences can be modeled by conditioning on a proposition from the algebra.
- In the case of structural evidence one may not want to assume that the agent has meta-beliefs about the other propositions in her algebra. And even if she had, how should one update on them?
- In the case of indicative conditionals almost everything is controversial: Is the conditional propositional (or is there a way around Lewis' triviality results)? And if so, can a conditional be represented by a material conditional? We will see that things become especially complicated when the conditional is not strict.

- We will focus on learning from indicative conditionals and show that the distance-based approach to Bayesianism (which uses the diachronic norm Conservatism) has the resources to analyze such purported cases of non-propositional evidence.

3. Genuinely New Evidence

Standard Bayesianism assumes that the learned proposition is already on the agent's "radar". It is expected and has a prior probability. However, this may not always be the case. Some evidence may be genuinely new.

- It is not very plausible that agents have prior beliefs about all possible kinds of evidence they may learn.
- Here are three examples:
  1. Testimony: A bystander gives a testimony about a murder case. Here it is not plausible that the policeman had any prior expectations about this testimony.
  2. Argumentation: We are debating and you make a new argument which I didn't anticipate.
  3. Scientific Theory Change: A new theory shows up and it is not plausible that the agent expected it (let alone that she assigned a probability to it). It is also not plausible that the new theory emerged from the "catch all" of the old theory (which may work sometimes).

3. Genuinely New Evidence

- It is a serious restriction of the Bayesian framework that it requests that the algebra of propositions is kept fixed.
- At the same time, it is a big advantage of logical frameworks such as the AGM model of belief revision that they can deal with these cases.
- Our hope is to learn from these approaches.

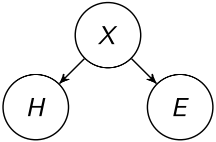
- We will see that an extension of the distance-based approach (which uses the diachronic norm Progressive Conservatism) has the resources to analyze these cases probabilistically.
- We will also extend the approach to modeling contraction, i.e. the elimination of propositions from the algebra. (Think, for example, of the phlogiston theory in the example of scientific theory change.)

I address the three problems in turn.

II. Problem 1: Non-Empirical Evidence

Non-Empirical Evidence

- Bayesianism has the resources to model such confirmation scenarios.
- There are, for example, interesting cases where the correlation between H and E is mediated by a "common cause" variable.



- To proceed, one has to find a variable X which (i) plays an active role in the reasoning of the agent and which (ii) plausibly screens off H from E.
- We will see that one can identify such variables for the analysis of the no alternatives argument and for the reasoning with analogue simulations.
- Other network structures are possible too.

Illustration: The No Alternatives Argument (NAA)

Scientists often argue like this:

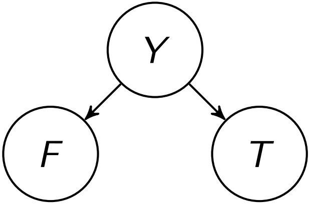
1. Hypothesis H satisfies several desirable conditions (it incorporates various principles, coheres with other theories,...)
2. Despite a lot of effort, the scientific community has not yet found an alternative to H.
3. Hence, we have one reason in support of H.

Questions:

- How good are NAAs?
- Under what conditions, if any, do they work?

The No Alternatives Argument: Variables

1. T has two values, viz. T: The hypothesis H is true, and $\neg$T: The hypothesis H is not true.
2. F also has two values, viz. F: The scientific community has not yet found an alternative to H that accounts for the data $\mathcal{D}$ and satisfies the constraints $\mathcal{C}$, and $\neg$F: The scientific community has found an alternative to H that accounts for $\mathcal{D}$ and satisfies $\mathcal{C}$.
3. Y has N values, viz. Y$_i$: There are exactly i hypotheses which explain $\mathcal{D}$ and fulfill $\mathcal{C}$. (H is one of them.)



The No Alternatives Argument: A Theorem

Theorem (Dawid, Hartmann and Sprenger 2015)

We set $P(\mathrm{Y_i}) =: y_i$, $P(\mathrm{F}|\mathrm{Y_i}) =: f_i$ and $P(\mathrm{T}|\mathrm{Y_i}) =: t_i$. If (a) f_i and t_i are monotonically decreasing in i, (b) $y_i < 1 \, \forall i$ and (c) there is at least one pair (i, j) with $j > i$ s.t. $y_i \, y_j > 0$, $f_i > f_j$ & $t_i > t_j$, then $P(\mathrm{T}|\mathrm{F}) > \mathrm{P}(\mathrm{T})$.

Discussion:

- The conclusion follows from three apparently plausible premises.
- Assumption (b) is the weakest. Note that the underdetermination thesis suggests that there are always infinitely many alternatives to a given theory that are equally in accordance with the data (whatever they are). So why should one not set $P(\mathrm{Y}_\infty) = 1$? A defender of the NAA has to answer this question (and Richard Dawid does so).

Upshot

- The case of analogue simulation can be analyzed accordingly (Dardashti et al. 2019).
- In general, the analysis of non-empirical evidence requires the specification of (i) other "active" variables (beyond H and E) and of (ii) a causal structure.
- Using this methodology, other alleged cases of non-empirical theory confirmation can be analyzed:
  1. Inference to the best explanation: 'H explains E best'; collider structure (Tesic, Eva, and Hartmann in prep.)
  2. The problem of old evidence: 'H explains E' and 'H's best competitor explains E' (Hartmann and Fitelson 2015, Eva and Hartmann 2019, Sprenger and Hartmann 2019).

III. Problem 2: New Types of Evidence

Learning Conditionals: An Example

The Ski Trip Example (Douven and Dietz 2011)

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,

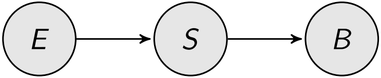
If Sue passed the exam, then her father will take her on a skiing vacation.

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

Question: How can this learning experience be modeled?

Using the Material Conditional

- Note that there are three propositional variables involved here:
  1. E: 'Sue passes the exam'
  2. S: 'Sue is invited on a ski trip'
  3. B: 'Sue buys a skiing outfit'
- 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 the material conditional $\mathrm{E} \supset \mathrm{S} \equiv \neg\mathrm{E} \vee \mathrm{S}$, one can show that the probability of E increases under plausible conditions.
- This becomes especially clear if one makes the additional assumption that $E \perp\!\!\!\perp B | S$ and if one uses the following chain structure:



Problems

- Representing the indicative conditional $\mathrm{A} \rightarrow \mathrm{C}$ by the material conditional $\mathrm{A} \supset \mathrm{C} \equiv \neg\mathrm{A} \vee \mathrm{C}$ has two problems:
  1. It cannot deal with non-extreme conditionals, e.g. when the conditional is not learned with certainty. Interestingly, (Jeffrey) conditionalizing on the material conditional leads to counter-intuitive consequences in these cases.
  2. It cannot deal with conditionals which are uttered by a partially reliable information source.

- The second problem is still (mostly) unsolved (see Collins et al. (2020), Hahn and Hartmann (2020)).
- To address the first problem, we introduce the distance-based approach.

The Distance-Based Approach

- Question: Which general principle should we use to specify our new probability distribution after learning a new item of information (propositional or otherwise)?
- Answer: The principle of Conservativity. That is, we request that the new probability distribution Q is as close as possible to the old probability distribution P, taking the learned information as a probabilistic constraint on Q into account.
- Examples:
  1. If we learn that the evidence E obtains, then the corresponding constraint is $Q(\mathrm{E}) = 1$.
  2. If we learn the indicative conditional "If A, then C", then the corresponding constraint is $Q(\mathrm{C}|\mathrm{A}) = 1$.
     Note that this does not require us to take a stance concerning the propositional status of an indicative conditional.
- Working out the proposal: To proceed, we need to choose a measure for the "distance" between two probability distributions.

Choosing a Distance Measure

f-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 Q are defined and let f be a convex function with $f(1) = 0$, $p_i := P(\mathrm{S}_\mathrm{i})$ and $q_i := Q(\mathrm{S}_\mathrm{i})$.Then

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

- The Kullback Leibler-divergence obtains for $f(t) = t \log t$. Other examples are the inverse KL-divergence, the χ^2-divergence, and the Hellinger distance.
- Note that f-divergencies are not necessarily symmetrical. They may also violate the triangle inequality.
- Question: Why should we focus on f-divergencies?

Jeffrey Conditionalization

Theorem (Diaconis and Zabell, 1982)

An agent considers the propositional variables H and E and has a probability distribution P defined over them. She then learns that $Q(\mathrm{E}) = e' < 1$. Minimizing an f-divergence between Q and P taking this constraint into account yields $Q(\mathrm{H}) = P(\mathrm{H}|\mathrm{E}) \cdot e' + P(\mathrm{H}|\neg\mathrm{E}) \cdot \overline{e'}$, i.e. Jeffrey Conditionalization.[a]

[a] Here and below we use the the shortcut $\overline{x} := 1 - x$.

- This is an important result that shows that all f-divergences are indistinguishable, at least when learning propositional evidence.
- We focus on f-divergences because we request that only those probabilistic divergences are admissible which yield Jeffrey conditionalization (two variables, rigidity holds).

Learning Indicative Conditionals

- If one learns the (strict) indicative conditional "If A, then C" from a perfectly reliable source, then the constraint on Q is $Q(\mathrm{C}|\mathrm{A}) = 1$.
- Minimizing an f-divergence between Q and P taking this constraint into account yields the same new probability distribution for all f-divergences and all scenarios discussed in the literature.
- This new probability distribution is identical with the one which one obtains by conditioning on the corresponding material conditional: $Q = P'$
- This is easy to see by noting that $Q(\mathrm{C}|\mathrm{A}) = 1$ iff $Q(\mathrm{A} \supset \mathrm{C}) = 1$ (if $Q(\mathrm{A}) > 0$).

Learning (non-strict) Indicative Conditionals

- Let us now consider non-strict indicative conditionals which are much more natural from a Bayesian point of view.
- In this case the constraint is $Q(\mathrm{C}|\mathrm{A}) < 1)$ and one finds that different f-divergencies yield different new probability distributions.
- We therefore have to "put our money" on one specific f-divergence. But on which?
- To proceed, we have the following three options:
  1. Accept the diachronic norm Minimizing Inaccuracy (as in Epistemic Utility Theory). Then the inverse KL-divergence is the unique probabilistic divergence.
  2. Identify other diachronic norms which (hopefully) restrict the class of admissible divergencies.
  3. Explore empirically which f-divergence is best (this may vary with the context).

Conditionalization and Its Limits

- As should be clear by now, I do not think that Conditionalization ("Bayes Theorem") is at the heart of Bayesianism.
- Conditionalization often yields the right results, but it should not be considered to be one of the defining elements of Bayesianism.
- This is because there are many other types of information one might learn, and the corresponding update can often not be modeled as an instance of Conditionalization.
- Even if we learn a proposition, there may be other relevant propositional variables involved which should, e.g., stay fixed across the update. Such additional constraints cannot be accounted for if the only principle we have available is Conditionalization.
- The distance-based approach, on the other hand, justifies (Jeffrey) Conditionalization (when it can be applied) and is much more general.

IV. Problem 3: Genuinely New Evidence

The Problem of the Algebra

- Genuine learning involves the addition of a new variable to the agent's algebra.
- It cannot be modeled as an application of Conditionalization as one can only condition on a proposition in one's algebra.
- It can also not be modeled by an application of the distance-based approach as this approach requires that the respective probability distributions are defined over the *same* algebra.
- Question: What can be done?

- N.B.: Genuine learning is standardly accounted for in models of belief revision such as the AGM model. This model introduces three basic operations: Belief expansion, contraction and revision and presents a number of postulates that these operations satisfy. This make sure that the new belief set is consistent.

Expansion

- The algebra $\mathcal{A}$ of an agent comprises the propositional variables $A_1, \ldots, A_n$.
- They are organized in a Bayesian network and the corresponding probability distribution is P.
- The agent then *decides* to add the propositional variable A_{n+1} to $\mathcal{A}$.
- Hence, the new ("expanded") algebra is $\mathcal{A}' = \mathcal{A} \cup A_{n+1}$.
- Question: What is her new probability distribution Q?

Expansion: The Procedure in Three Steps

1. (**Relate**) Relate A_{n+1} to $A_1, \ldots, A_n$:
   (i) Leave the network structure of $A_1, \ldots, A_n$ unchanged.
   (ii) Add arcs between A_{n+1} and $A_1, \ldots, A_n$.
   (iii) Identify as many constraints (involving A_{n+1}) on Q as possible (such as likelihoods).
2. (**Minimize**) Minimize an f-divergence between Q (restricted to $A_1, \ldots, A_n$) and P, taking the constraints into account.
3. (**Adjust**) If this does not fully specify Q, then one has at least two options:
   (i) adopt an imprecise probability distribution, or
   (ii) apply MaxEnt or a related method.

Progressive Conservatism

- The resulting diachronic norm is called Progressive Conservatism.
- Progressive, because the agent is encouraged to expand her algebra.
- Conservatism, because the agent makes sure that her old beliefs, i.e. her beliefs about the propositions in her old algebra,, change as little as possible.

- The proposed procedure leads to plausible results.
- For example, if one only specifies the marginal probability of the new proposition, then the probability distribution over the old variables does not change and the new variable is independent of the old ones.
- Another interesting example is novel confirmation.

Revision

- Expansion assumes that the network structure does not change.
- However, a new proposition may prompt you to make changes in the network structure.
- Example: You consider two positively correlated variables – *YellowFingers* and *HeartDisease*. You model the correlation by drawing an arc from one to the other. Then you learn a (genuinely) new variable – *Smoking* – which prompts you to change the network structure to a common cause network.
- The new network structure imposes a structural constraint on the new probability distribution.
- Then the Conservativity norm can be applied.
- It will be interesting to systematically study examples like this.

Contraction

Here the agent switches from an algebra of n propositional variables $(A_1, \ldots, A_n)$ to an algebra of $n-1$ propositional variables $(A_1, \ldots, A_{n-1})$.

Case 1: No change of the network structure of the old algebra (restricted to $A_1, \ldots, A_{n-1}$):

1. (**Eliminate**) Eliminate A_n and delete all arcs to and from A_n.
2. (**Minimize**) Minimize an f-divergence between Q and P (restricted to $A_1, \ldots, A_{n-1}$). This simply amount to marginalizing out A_n from P.

Case 2: No change of the network structure of the old algebra (restricted to $A_1, \ldots, A_{n-1}$):

1. (**Eliminate**) Eliminate A_n and delete all arcs to and from A_n.
2. (**Minimize**) Minimize an f-divergence between Q and P (restricted to $A_1, \ldots, A_{n-1}$). This may lead to a different result than the one which follows from marginalizing out A_n from P. (There may be interesting parallels to the principle of minimal change in the AGM model.)

VI. Outlook

Outlook

- Bayesian networks have many intriguing applications in philosophy of science (and in other parts of philosophy and science).
- These applications build on the fact that Bayesian networks provide a compact and intuitive representation of an agent's belief system.
- Bayesian networks are also an important tool for the development of philosophical theories. To show this, I focused on the Bayesian framework and identified strategies to solve three open problems:
  1. Non-empirical evidence: introduce new "active" variables.
  2. New types of evidence: introduce the distance-based approach with the Conservatism norm.
  3. Genuinely new evidence: introduce the new norm Progressive Conservatism.
- We have seen that Bayesianism is a powerful philosophical modeling framework that is open for empirical input (e.g. regarding which f-divergence is best). Bayesian networks make it even more powerful.

References

1. Bovens, L. and S. Hartmann (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.
2. Collins, P., K. Krzyzanowska, S. Hartmann, G. Wheeler and U. Hahn (2020). Conditionals and Testimony. *Cognitive Psychology* 122: 101329.
3. Dardashti, R., S. Hartmann, K. Thébault, and E. Winsberg (2019). Hawking Radiation and Analogue Experiments: A Bayesian Analysis. *Studies in History and Philosophy of Modern Physics* 67: 1–11.
4. Dawid, R., S. Hartmann and J. Sprenger (2015). The No Alternatives Argument. *The British Journal for the Philosophy of Science* 66(1): 213–234.
5. Eva, B. and S. Hartmann (2018). Bayesian Argumentation and the Value of Logical Validity. *Psychological Review* 125(5): 806–821.

References (cont'd)

6. Eva, B. and S. Hartmann (2020). On the Origins of Old Evidence. *Australasian Journal of Philosophy* 98(3): 481–494.
7. Eva, B., S. Hartmann and S. Rafiee Rad (2020). Learning from Conditionals. *Mind* 129(514): 461–508.
8. Hahn, U. and S. Hartmann (2020). A New Approach to Testimonial Conditionals, in *CogSci 2020 Proceedings*, Toronto, 981–986.
9. Hartmann, S. (2021). Bayes Nets and Rationality. To appear in M. Knauff and W. Spohn (eds.), *The Handbook of Rationality*. Boston, MA: MIT Press.
10. Sprenger, J. and S. Hartmann (2019). *Bayesian Philosophy of Science*. Oxford: Oxford University Press.
11. Stern, R. and S. Hartmann (2018). Two Sides of Modus Ponens. *The Journal of Philosophy* 115(11): 605–621.

Thanks...

...for your attention...

and to my collaborators Richard Dawid, Benjamin Eva, Ulrike Hahn, Soroush Rafiee Rad, Jan Sprenger, and Reuben Stern.