Weight Chapter Outline

9/1/2022

1 Introduction

The probability we assign to a hypothesis expresses our epistemic uncertainty about the hypothesis. But another uncertainty is at play. It is the uncertainty about our assessment of the probability of the hypothesis. Consider a familiar example from the courtroom. A criminal defendant is standing trial. Traces that were found at the crime scene match the defendant's genetic profile. An expert testifies that the probability that a random person would coincidentally match is as low as 1 in 10 billion. This extremely low probability can serve, together with other prior information, to assess the probability that the defendant was actually the source. Say the probability of the source hypothesis is assessed to be 95%. This number expresses our uncertainty about the hypothesis. The problem is that the 95% probability rests, among other things, on the random match probability of 1 in 10 billion. This figure could itself be questioned. So, besides the uncertainty about the source hypothesis, there is uncertainty about the 95% probability assigned to the sources hypothesis.

Such higher order uncertainty is quite ubiquitous. For suppose we test positive for a certain disease. The test isn't infallible. So, based on the test result, the probability we should assign to the hypothesis that we are actually positive is somewhere below 100%, say it is assessed to be 75%. This number expresses our uncertainty about the hypothesis that we are positive. The 75% figure is the output of applying Bayes' theorem, which however requires as inputs the base rate of the disease and the error rates of the test. These numbers are themselves uncertain. So the 75% probability that we are positive carries its own uncertainty.

Numbers carry with them an aura of objectivity. Their presentation by an expert might suggest they are not up for debate. But that would be a mistake. A skilled defense lawyer will inquiry about the credentials of the expert, the source of the data, their reliability. At one extreme, the numbers presented in a courtroom can be made up. In the infamous People v. Collins, the prosecutor made up the frequencies of certain identifying characteristics: interracial couple, driving a yellow convertible, women with ponytail, man with mustache, etc. But the numbers presented in a courtroom can also rest on well-established practices of data collection and genetic modeling, as in the case of random match probabilities for DNA evidence. All in all, second-order uncertainty comes in degrees, just like first-order uncertainty about a hypothesis of interest. We need a theory of both uncertainties.

This chapter articulates a formal framework for modeling first-order uncertainty about hypotheses as well as uncertainty about the first-order uncertainty. The framework fits naturally with Bayesian statistics and can be implemented thorough Bayesian networks. We develop this framework as an improvement of imprecise probabilism in Bayesian epistemology. As the name suggests, imprecise probabilism posits that agents can be imprecise about the probabilities they assign. Instead of single probability measures, they can entertain multiple probability measures. The higher order framework we develop brings imprecise probabilism one step further. Imprecise probabilism assumes that the multiple probability measures which agents can entertain are all equally likely. The higher order approach relaxes this assumption. The possible probability measures can themselves be more or less probable.

The two-tiered framework for modeling uncertainty will then serve to articulate a theory of evidential weight. (Need to add why a theory fo weight of evidence is important, etc.)

Side comment: I think for expository clarity, we should emphasize two contributions here. One is the contribution to higher order uncertainty. This is an improvement on imprecise probabilism. This I think is already an important contribution that fares better than imprecise probabilism. The second contribution is the theory of weight that is built out of the higher order uncertainty framework. The two issues seems separate to me. One could be interested in question of higher order uncertainty and imprecise probabilism without being interested in questions of weight.

2 Precise and imprecise probabilism

3 Higher oreder probabilism

These two sections will describe the three main frameworks of probabilism: precise, imprecise, and higher order probabilism. They should do two things. First, they should motivate why the move to imprecise probabilism is warranted but also show the limits of imprecise probabilism. Second, they should outline what higher order probabilism looks like and why it overcomes the problems of imprecise probabilism.

3.1 Motivating examples

Some of the motivating examples are from **section 1**:

(a1) Use example by Peirce (sampling 100 v. sampling 1000 times, with equal sample proportion). Balance of evidence for/against a certain hypothesis seems the same, but the informational basis (weight) is wider in the second case. (Brief comment –perhaps in a footnote– that this difference is modeled in standard statistics as the SE of the sample proportion– the SE decreases as the sample size increases. This is not what we are after though since we are not simply trying to estimate a parameter value such as population proportion, but assess the probability that a hypothesis is true.)

(a2) Give an example like Peirce's using varying sample sizes to estimate relative proportion of some identifying feature in match evidence, handwriting, genetic profile, fiber, etc. DNA evidence or a simpler form of match evidence (see e.g. the Georgia v. Wayne Williams case) should do here. This can connect back to DNA evidence example in the introduction.

3.2 Addditional motivating examples

(b) Another intuitive example is, one thing is absolute ignorance about an event (or suspension of judgment because of lack of knowledge), and another thing is having equal information for/against. Assigning in both cases a sharp probability of .5 fails to capture the intuitive difference. In one case the informational basis is very limited, or non-existent, while in the other the informal basis is wider, even though the balance might seem the same.

(c) A similar problem is raised by the "negation problem" by Cohen (little evidence in favor of H, so Pr(H) is low, cannot mean there is a lot of evidence in favor of not-H, Pr(not-H) is high).

Comment: What is now in **sections 8**, **9** and **10** should go in these two sections. I would remove all the stuff about Joyce (since this is about weight), but basically all the materials we need are in those sections. Also, what is now in **section 12** ("Higher order probability and weight in BNs") should be part of these two sections.To talk about "higher ordar Bayesian networks" there is not need to introduce all the stuff about weight. In fact, a reader interested in Bayesian networks might want to learn about higher order Bayesian networks even though they are not interested in weight.

Possible addition: To give the reader an intuitive picture, we could provide three Bayesian networks for the diagnostic example from the introduction section. The first network has the shape with $D \rightarrow T$, and is just how one would do things in the standard way with sharp probabilities. The second network contains multiple probability measures about (a) the prior probability of D and second-order uncertainty about the error rates that go into the conditional table for $P(T|D)$. This is the network using imprecise probabilism. This is also something like "sensitivity analysis." Finally, the third network contains distributions over multiple probability distributions—the higher order approach. The same could be done for the DNA match example. This comparison would convey succinctly the first major contribution of the chapter. The three Bayesian networks will also nicely connect with the two motivating examples right at the start of the chapter. You use the Sally Clark case as an illustration. This goes even one step further, but it might be good to have a simple illustration even with a simple match-source Bayesian network.

4 Weight of evidence

The chapter now turns to the weight of evidence and its formalization.

Comment: It is conceptually important to separate the discussion about precise, imprecise, and higher order probabilism (previous sections) from weight (this section). Weight of evidence is one way in which higher order probabilism can be put to use. It can be confusing to run the discussion of higher order probabilism together with weight of evidence. Higher order probabilism can still make perfect sense even if no theory of weight cen be worked out.

4.1 Motivating examples

This section should start with illustrative examples of the weight/balance distinction and why "balance alone" isn't enough to model the evidential uncertainty relative to a hypothesis of interest. These examples should be chosen carefully. We can use legal and non-legal examples. The driving intuition is given by Keynes with the weight/balance distinction. Some of the examples we saw earlier in talking about imprecise probabilism can be mentioned here again, such as (a1) and (a2), and perhaps also (b) and (c).

Upshot is that uncertainty cannot be captured by balance of evidence alone. There is a further dimension to uncertainty. So we need a theory that can accommodate this further level of uncertainty. This theory is essentially the higher order probabilism introduced before.

4.2 Desiderata

Here we can discuss monotonicity, completeness, strong increase, etc (see current **section 1**). We can list the intuitive properties (based on the example we presented in both philosophy and law) that any theory of weight (and perhaps also of completeness/resilience, but on these notions, see later) should be able to capture. We should try to keep these requirements as simple as possible and leave complications to footnotes.

4.3 Formal characterization of weight

Higher order probabilism is then put to use to deliver a theory of weight. What is now in **section 11** ("Weight of a distribution") and **sections 13** and **14** (" Weight of evidence" and "Weights in Bayesian Networks") forms the bulk of the theory.

We should also demonstrate that the proposed theory of weight does meet the intuitive desiderata and can handle the motivating examples. To better appreciset the novelty of the proposal, It might be interesting to raise the following questions:

- q1 what does a theory of weight based on precise probabilism look like? (maybe it consists of something like Skyrms' resilience or Kaye's completeness, the problem being that these are not measures of weight, but of something else, more on these later)
- q2 what does a theory of weight based on imprecise probabilism look like? (is Joyce's theory essentially an attempt to use imprecise probabilism to construct a theory of weight?)
- q3 what does a theory of weight based on higher order probabilism look like?

Here we are defending a theory fo weight based on higher order probabilism, but it is interesting to contrast it with a theory of weight based on the other version of legal probabilism. Here we can also show why Joyce's theory of weight does not work (either in the main text or a footnote).

Comment: The current exposition in chapter 11, 13 and 14, however, is complicated—perhaps overly so. The move from "weight of a distribution" to "weight of evidence" is not intuitive and can confuse the reader. Is there a simpler story to be told here? I think so. See below.

Suggestion: There seems to be a nice symmetry. Start with precise probabilism. We can use sharp probability theory to offer a theory of the value of the evidence (i.e. likelihood ratio). Actually, I think that the likelihood ratio model the idea of balance of the evidence. What Keynes distinction weight/balance shows is that likelihood ratio are not, by themselves, enough to model the value of the evidence. The straightforward move here seems to just have **higher order likelihood ratios**. Wouldn't higher order likelihood ratio be essentially your formal model of the weight of the evidence? Your measure of weight tracks the difference between (the weight of the) prior distribution (and the weight of the) posterior distribution. But higher order likelihood ratios essentially do the same thing, just like precise likelihood ratios track the difference between prior and posterior. Is this right?

Comment: If weight if measured by higher order likelihood ratios, then this can be seen as a generalization of thoughts that many other had – say that the absolute value of the likelihood ratio is a measure of weight (Nance, Glenn Shafer) or that likelihood ratio must be a measure of weight (Good; see current **section 4**). So I think using "“higher order likelihood ratio” could be a more appealing way to sell the idea of weight of evidence since most people are already familiar with likelihood ratios.

4.4 Limits of our contribution

Work by Nance of Dahlman suggests that “weight” should play a role in the standard of proof. We do not take a position on that. Weight could be regulated by legal rules at the level of rules of decision, rule of evidence, admissibility, sanctions at the appellate level. All that matters to us is that, in general, legal decision-making is sensitive to these further levels of uncertainty (quantity, completeness, resilience), but whether this should be codified at the level of the standard of proof or somewhere else, we are not going to take a stance on that.

4.5 Objection

Ronald Allen or Bart Verheij might object as follows. Precise probabilism is bad because we do not always have the numbers we need to plug into the Bayesian network. Imprecise probabilism partly addresses this problem by allowing for a range instead of precise numbers. How does higher order probabilism help address the practical objection that we often we do not have the numbers we need to plug into the Bayesian network?

5 Completeness (and resilience?)

Next the chapter turns to notions related to the weight of evidence, such as completeness (and perhaps resilience as well). See current **sections 5** and **6**.

5.1 Motivating example

Give an example using completeness of evidence (pick one or more court cases). The court case we can use is Porter v. City of San Francisco (see file with Marcello’s notes).[1] The jury is given an instruction that a call recording is missing, but no instruction whether the call should be assumed to be favorable or not.

What is the jury supposed to do with this information? If the call could contain information that is favorable or not, shouldn’t the jury simply ignore the fact that the call recording is missing (Hamer’s claim)? Modelling with Bayesian network might turn out useful. Cite also David Kaye on the issue of completeness. His claim is that when evidence is known to be missing, then this information should simply be added as part of the evidence, which is precisely what the court in Porter does. But again, once we add the fact that the evidence is missing what is the evidentiary significance of that? What is te jury supposed to do with that? Does Pr(H) go up, down or stays the same? Kaye does not say…

[1] This is a wrongful death case in which victim was committed to a hospital facility, but escaped and then died under unclear circumstances. So the nurses and other hospital workers—actually, the city of San Francisco—are accused of contributing to this person’s death. Need to check exact accusation—this is not a criminal case. A phone call was made to social services shortly after the person disappeared, but its content was erased from hospital records. Court agrees that content of phone call would be helpful to understand what happened and to assess the credibility of hospital’s workers ("The Okupnik call is the only contemporaneous record of what information was reported to the SFSD about Nuriddin’s disappearance, and could contain facts not otherwise known about her disappearance and CCSF’s response. Additionally, the call is relevant to a jury’s assessment of Okupnik’s credibility"). The court thought that the hospital should have kept records of that call. But court did not think the hospital acted in bad faith or intentionally, so it did NOT issue an "adverse inference instruction" (=the missing evidence was favorable to the party that should have preserved it, but failed to do it).

5.2 Bayesian network model

Comment I am thinking that incompleteness is modeled by adding an evidence node to a Bayesian network but without setting a precise value for that node, and then see if the updated network yields a different probability than the previous network without the missing evidence node. The missing evidence node could be added in different places and this might changes things. In the Porter case the missing evidence seems to affect the credibility of the other evidence in the case, we would have a network like this: $H \rightarrow E \leftarrow C$, where C is the missing evidence node and E is the available evidence nose. My hunch is that (see also our paper on reverse Bayesianiam and unanticipated possibilities) the addition of this credibility node will affect the probability of the hypothesis (thus proving Hamer wrong).

5.3 Expected weight model

Question: If what I say above in the comment is correct, then a question arises, do we need higher order probabilism to model completeness?

Possible answer: We can use expected weight (see current **section 14**). If the expected weight of an additional item of evidence is null, that would mean that its addition (not matter the value the added evidence would take) cannot change the probability of the hypothesis. If the expected weight is different from zero (pace Hamer who thinks the expected weight is always null), then the evidence can change the probability of the hypothesis.

6 Weight and accuracy

This section addresses the question, why care about weight?

7 Conclusion