Weight Chapter Outline

9/1/2022

Contents

1 Introduction

Consider three different items of match evidence:[1] The suspects dog's fur matches the dog fur found in a carpet wrapped around one of the bodies. A hair found on one of the victims matches that of the suspect. Carpet fibers in a carpet used to wrap a victim's body matches carpet fibers in the suspect's house. What are the fact-finders to make of this evidence? To start with, some probabilistic evaluation thereof should be useful.

2 Precise and imprecise probabilism

Precise probabilism (PP) holds that a rational agent's uncertainty about a hypothesis H is to be represented as a single, precise probability measure. This is an elegant and simple theory. But representing our uncertainty about a proposition in terms of a single, precise probability runs into a number of difficulties. Precise probabilism fails to capture an important dimension of how our uncertainty connects with the evidence we have or have not obtained.

[1] These are stylized after the evidence in the notorious Wayne Williams case. Probabilities have been slightly but not unrealistically shifted to be closer to each other to make a conceptual point. The original probabilities were 1/8000 for the carpet evidence, 1/100 for the dog fur, and 29/1148 for Wayne Williams' hair. Also, in our example we assume carpets have been sampled, whereas the way the actual probability has been arrived at was by tracking down the carpet producer, investigating their sales record, assuming the carpets were evenly distributed in 10 southeastern states, assuming the average carpet is 12 feet long, and taking an estimate of total amount of carpet in the country for granted. We abstract from such considerations in this chapter.

Start with a case of complete lack of evidence. You hold a coin in your hands but have no evidence whatsoever about its bias. You are completely ignorant. You then start tossing the coin and observe the outcome of ten tosses, half of which turn out to be heads. This is some evidence for the real bias being around .5. How do you represent your stances before and after the observations? Precise probabilism has difficulties modeling the difference between the two situations. If you deploy the principle of insufficient evidence, you would start with $\mathsf{P}_0(H) = .5$ and end with $\mathsf{P}_1(H) = .5$, as if nothing changed. If you do not deploy the principle of insufficient evidence, what do you do?

Precise probabilism runs into difficulties even in cases that not depict complete lack of evidence. The following example—from the 1872 manuscript 'The Fixation of Belief' (W3 295) by C. S. Peirce—makes this clear:

> When we have drawn a thousand times, if about half [of the beans] have been white, we have great confidence in this result ... a confidence which would be entirely wanting if, instead of sampling the bag by 1000 drawings, we had done so by only two.

The difficulty for precise probabilism is this. Your best estimate of the probability of 'the next bean will be white' is .5 if half of the beans you have drawn randomly so far have been white, no matter whether you have drawn a thousand or only two of them. There is an intuitive difference between the two cases, but expressing one's uncertainty with a precise probability does not capture it.[2]

Both examples—Peirce's bean example and the earlier one about lack of evidence—suggest that PP is not appropriately responsive to evidence. It ends up assigning a probability of .5 to situations in which one's evidence is quite different: when no evidence whatsoever is available for or against a hypothesis; when there is minimal evidence of equipoise between hypotheses (say, after only 2 draws); when there is strong evidence of equipoise (say, after 1000 draws).[3]

What if we give up the assumption that probability assignments should be precise? Unlike PP, **imprecise probabilism** (IP) holds that an agent's credal stance towards a hypothesis H is to be represented by means of a *set of probability measures*, typically called a representor $\mathbb{P}$, rather than a single measure P. The representor should include all and only those probability measures which are compatible (in a sense to be specified) with the evidence. For instance, if an agent knows that the coin is fair, their credal state would be captured by the singleton set $\{\mathsf{P}\}$, where P is a probability measure which assigns .5 to H. If, on the other hand, the agent knows nothing about the coin's bias, their credal state would rather be represented by means of the set of all probabilistic measures, as none of them is excluded by the available evidence. Note that the set of probability measures does not represent admissible options that the agent could legitimately pick from the set. Rather, the agent's credal state is essentially imprecise and should be represented by means of the whole set of probability measures.[4]

Imprecise probabilism shares some similarities with what we might call **interval probabilism** due to [KYBURG 1961]. On interval probabilism, precise probabilities are replaced by intervals of probabilities. On imprecise probabilism, instead, precise probabilities are replaced by sets of probabilities. This makes imprecise probabilism more general since the probabilities of a proposition in the representor set do not have to form a closed interval. Both approaches, however, can model situations of complete lack of evidence by probability measures that assign values in the interval [0, 1].

REF Kyburg. Probability and the Logic of Rational Belief. Wesleyan University Press, Middletown Connecticut, 1961 and H. E. Kyburg and C. M. Teng. Uncertain Inference. Cambridge University Press, Cambridge, 2001.

M: can they model the Peirce's bean example, not clear to me. Seems we need to mention this.

As more evidence is gathered, the interval might widen or shrink (in Kyburg's approach) or some probability measures might be added to, or removed from, the representor set (in imprecise probabilism). Learning is modeled in a somewhat idiosyncratic way in Kyburg's interval probabilism by performing operations on intervals.[5] The advantage of imprecise probabilism, instead, is that it provides a straightforward picture of learning from evidence, that is a natural extension of the classical Bayesian approach. This makes imprecise probabilism much preferable to interval probabilism. When faced with

[2] Similar remarks can be found in Peirce's 1878 *Probability of Induction*. There, he also proposes to represent uncertainty by at least two numbers, the first depending on the inferred probability, and the second measuring the amount of knowledge obtained; as the latter, Peirce proposed to use some dispersion-related measure of error (but then suggested that an error of that estimate should also be estimated and so, so that ideally more numbers representing errors would be needed).

[3] Precise probabilism suffers from other difficulties. For example it has problems with formulating a sensible method of probabilistic opinion aggregation Stewart & Quintana (2018). A seemingly intuitive constraint is that if every member agrees that X and Y are probabilistically independent, the aggregated credence should respect this. But this is hard to achieve if we stick to PP (Dietrich & List, 2016). For instance, a *prima facie* obvious method of linear pooling does not respect this. Consider probabilistic measures p and q such that $p(X) = p(Y) = p(X|Y) = 1/3$ and $q(X) = q(Y) = q(X|Y) = 2/3$. On both measures, taken separately, X and Y are independent. Now take the average, $r = p/2 + q/2$. Then $r(X \cap Y) = 5/18 \neq r(X)r(Y) = 1/4$.

[4] For the development of IP see (Fraassen, 2006; Gärdenfors & Sahlin, 1982; Joyce, 2005; Kaplan, 1968; Keynes, 1921; Levi, 1974; Sturgeon, 2008; Walley, 1991), (Bradley, 2019) is a good source of literature.

[5] EXPLAIN

new evidence E between time t_0 and t_1, the representor set should be updated point-wise, running the standard Bayesian updating on each probability measure in the representor:

$$\mathbb{P}_{t_1} = \{\mathsf{P}_{t_1} | \exists\, \mathsf{P}_{t_0} \in \mathbb{P}_{t_0}\ \forall H \left[\mathsf{P}_{t_1}(H) = \mathsf{P}_{t_0}(H|E)\right]\}.$$

Unfortunately, because of this point-wise updating, imprecise probabilism runs into the problem of **belief inertia**. (Levi, 1980). Consider again Peirce bean's example. Say you start drawing beans knowing only that the true proportion of red beans is in the interval $(0,1)$. This models a situation of lack of evidence. As you draw beans from the urn and discover their color, you should be able to learn something about the proportion of colors in the urn. This is not so with imprecise probabilism, however. For suppose you draw two beans both of which are red. On imprecise probabilism, your initial credal state is to be modeled by the set of all possible probability measures over your algebra of propositions. Once you observe the two beans, each particular measure from your initial representor gets updated to a different one that assigns a higher probability to "red," but also each measure in your original representor can be obtained by updating some other measure in your original representor on the evidence (and the picture does not change if you continue with the remaining 2998 observations). Thus, if you are to update your representor point-wise, you will end up with the same representor set. Consequently, the edges of your resulting interval will remain the same. In the end, it is not clear how you are supposed to learn that the proportion of beans is such and such.[6]

M: open or closed interval?

M: Can this be more clear? Not sure I completely follow...

Some replies in defense of imprecise probabilism are available. One might insist that vacuous priors should not be used and that the framework gives the right results when the priors are non-vacuous. After all, you did not start with knowing truly nothing, then perhaps it is right to conclude that you will never learn anything. Another strategy is to say that, in a state of complete ignorance, a special updating rule should be deployed.[7] But no matter what we think about belief inertia, other problems plague imprecise probabilism.

One problem is that imprecise probabilism does not seem fine-grained enough. Consider a situation in which you are about to toss a coin whose bias is either .4 or .6. Say you have two coins and you know, for sure, that the probability of heads is .4. if you toss one coin and .6 if you toss the other coin. But you do not know which is which. You pick one of the two at random and toss it. You do not know the probability of heads on that toss, but you know it must be either .4 or .6. This situation can be easily represented by imprecise probabilism. The representator would contain two probability measures, one that assigns .4. and the other that assigns .6 to the hypothesis 'this coin lands heads.' But now suppose you have information that the lighter coin is more likely the one with the .4 bias. You have picked the lighter coin. So, upon tossing this coin, you should be more confident that the probability of heads is .4. rather than .6. Imprecise probabilism cannot represent this situation, at least not without moving to higher-order probabilities, in which case it is no longer clear whether the object-level imprecision performs any valuable task.

3 Higher order 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.

[6] Here's another example from (Rinard, 2013). Either all the marbles in the urn are green (H_1), or exactly one tenth of the marbles are green (H_2). Your initial credence $[0,1]$ in each. Then you learn that a marble drawn at random from the urn is green (E). After conditionalizing each function in your representor on this evidence, you end up with the the same spread of values for H_1 that you had before learning E, and no matter how many marbles are sampled from the urn and found to be green.

[7] (Elkin, 2017) suggests the rule of *credal set replacement* that recommends that upon receiving evidence the agent should drop measures rendered implausible, and add all non-extreme plausible probability measures. This however, is tricky: one needs a separate account of what makes a distribution plausible or not. Elkin admits that he has no solution to this: "But how do we determine what the set of plausible probability measures is relative to E? There is no precise rule that I am aware of for determining such set at this moment, but I might say that the set can sometimes be determined fairly easily" [p. 83] He goes on to a trivial example of learning that the coin is fair and dropping extreme probabilities. This is far from a general account. One also needs a principled account of why one should use a separate special update rule when starting with complete ignorance.

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.)

Not sure how far you want to engage with classical statistics; and I'm not sure what your suggested reply means, especially that now I'm thinking of this from the perspective of a potential critic such as Taroni

(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.

I'm thinking severed fingers, dogs and DNA, will cook up an example preparing for later critical comments about Taroni

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.

Right, but not with the diagnosis; let's use the legal examples to start with

4 Weight of evidence

Yeh, I think in the end this will be another chapter

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?)

Well, it's a bit funny as Joyce's weight uses precise chance hypotheses instead of IP, so hard to say

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.

Brilliant, I think I can start talking about conditional probabilities to begin with

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?

Yup, more or less

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).[8] 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...

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).

I think this will depend on how the probability of obtaining new evidence given guilt and given innocence are, I will keep thinking about this, we'll move to this once the earlier bits are done

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?

[8] 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).

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.

LR ratio and weight

6 Weight and accuracy

This section addresses the question, why care about weight?

Conclusion

Bradley, S. (2019). Imprecise Probabilities. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2019). https://plato.stanford.edu/archives/spr2019/entries/imprecise-probabilities/; Metaphysics Research Lab, Stanford University.

Dietrich, F., & List, C. (2016). Probabilistic opinion pooling. In A. Hajek & C. Hitchcock (Eds.), *Oxford handbook of philosophy and probability*. Oxford: Oxford University Press.

Elkin, L. (2017). *Imprecise probability in epistemology* (PhD thesis). Ludwig-Maximilians-Universität; Ludwig-Maximilians-Universität München.

Elkin, L., & Wheeler, G. (2018). Resolving peer disagreements through imprecise probabilities. *Noûs*, *52*(2), 260–278. https://doi.org/10.1111/nous.12143

Fraassen, B. C. V. (2006). Vague expectation value loss. *Philosophical Studies*, *127*(3), 483–491. https://doi.org/10.1007/s11098-004-7821-2

Gärdenfors, P., & Sahlin, N.-E. (1982). Unreliable probabilities, risk taking, and decision making. *Synthese*, *53*(3), 361–386. https://doi.org/10.1007/bf00486156

Joyce, J. M. (2005). How probabilities reflect evidence. *Philosophical Perspectives*, *19*(1), 153–178.

Kaplan, J. (1968). Decision theory and the fact-finding process. *Stanford Law Review*, *20*(6), 1065–1092.

Keynes, J. M. (1921). *A treatise on probability, 1921*. London: Macmillan.

Levi, I. (1974). On indeterminate probabilities. *The Journal of Philosophy*, *71*(13), 391. https://doi.org/10.2307/2025161

Levi, I. (1980). *The enterprise of knowledge: An essay on knowledge, credal probability, and chance*. MIT Press.

Rinard, S. (2013). Against radical credal imprecision. *Thought: A Journal of Philosophy*, *2*(1), 157–165. https://doi.org/10.1002/tht3.84

Stewart, R. T., & Quintana, I. O. (2018). Learning and pooling, pooling and learning. *Erkenntnis*, *83*(3), 1–21. https://doi.org/10.1007/s10670-017-9894-2

Sturgeon, S. (2008). Reason and the grain of belief. *Noûs*, *42*(1), 139–165. Retrieved from http://www.jstor.org/stable/25177157

Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. Chapman; Hall London.