Second-order Probabilism: Expressive Power and Accuracy

Rafal Urbaniak and Marcello Di Bello

2024-04-15

Table of contents

DISCLAIMER: This is a draft of work in progress, please do not cite or distribute without permission.

Abstract.

need to write one when done

1 Introduction

As rational agents, we form beliefs about a variety of propositions on the basis of the evidence available to us. But believing a proposition is not an all-or-nothing affair; it is a matter of degrees. We are uncertain, to a greater or lesser extent, about the truth of many propositions since the evidence we possess about them is often fallible. To represent this uncertainty, it is natural to use a probability measure that assigns to each proposition a value between 0 and 1 (also called a degree of belief or credence). This approach—known as *precise probabilism*—models an agent's state of uncertainty (or credal state) with a single probability measure: each proposition is assigned one probability value (a sharp degree of belief). The problem is that a sharp probability measure is not expressive enough to distinguish between intuitively different states of uncertainty rational agents may find themselves in (§**??**). To avoid this problem, a *set* of probability measures, rather than a single one, can be used to represent the uncertainty of a rational agent. This approach is known as *imprecise probabilism*. It outperforms precise probabilism in some respects, but also runs into problems of its own (§**??**).

To make progress, this paper argues that the uncertainty of a rational agent is to be represented neither by a single probability measure nor a set of measures. Rather, it is to be represented by a higher-order probability measure, more specifically, a probability distribution over multiple probability measures. Call this view *higher-order probabilism*. We show that higher-order probabilism addresses all the problems and philosophical puzzles that plague both precise and imprecise probabilism (§**??** and §**??**).

Moreover, Bayesian probabilistic programming already provides a fairly reliable implementation framework of this approach.

add structure description

2 Precise probabilism

Think about including synergy

Precise probabilism (PP) holds that a rational agent's uncertainty about a proposition is to be represented as a single, precise probability measure. Bayesian updating regulates how the prior probability measure should change in light of new evidence that the agent learns. The updating can be iterated multiple times for multiple pieces of evidence considered successively. This is an elegant and simple theory with many powerful applications. Unfortunately, representing our uncertainty about a proposition in terms of a single, precise probability measure runs into a number of difficulties.

Precise probabilism fails to capture an important dimension of how our fallible beliefs reflect the evidence we have (or have not) obtained. A couple of stylized examples featuring coin tosses should make the point clear.

> **No evidence v. fair coin** You are about to toss a coin, but have no evidence about its bias. You are completely ignorant. Compare this to the situation in which you know, based on overwhelming evidence, that the coin is fair.

On precise probabilism, both scenarios are represented by assigning a probability of .5 to the outcome *heads*. If you are completely ignorant, the principle of insufficient evidence suggests that you assign .5 to both outcomes. Similarly, if you know for sure the coin is fair, assigning .5 seems the best way to quantify the uncertainty about the outcome. The agent's evidence in the two scenarios is quite different, but precise probabilities fail to capture this difference.

> **Learning from ignorance** You toss a coin with unknown bias. You toss it 10 times and observe *heads* 5 times. Suppose you toss it further and observe 50

heads in 100 tosses.

Since the coin initially had unknown bias, you should presumably assign a probability of .5 to both outcomes if you stick with precise probabilism. After the 10 tosses, you again assess the probability to be .5. You must have learned something, but whatever that is, it is not modeled by precise probabilities. When you toss the coin 100 times and observe 50 heads, you learn something new as well. But your precise probability assessment will again be .5.

These examples suggest that precise probabilism is not appropriately responsive to evidence. Representing an agent's uncertainty by a precise probability measure can fail to track what an agent has learned from new evidence. Precise probabilism assigns the same probability in situations in which one's evidence is quite different: when no evidence is available about a coin's bias; when there is little evidence that the coin is fair (say, after only 10 tosses); and when there is strong evidence that the coin is fair (say, after 100 tosses). In fact, analogous problems also arise for evidence that the coin is not fair. Suppose the rational agent starts with a weak belief that the coin is .6 biased towards heads. They can strengthen that belief by tossing the coin repeatedly and observing, say, 60 heads in 100 tosses. But this improvement in their evidence is not mirrored in the .6 probability they are supposed to assign to *heads*.[1]

add reference about sweetening

These problems generalize beyond cases of coin tossing. It is one thing not to know much about whether a proposition is true, for example, whether an individual is guilty of a crime. It is another thing to have strong evidence that favors a hypothesis and equally strong evidence that favors its negation, for example, strong evidence favoring the guilt hypothesis and equally strong evidence favoring the hypothesis of innocence. Despite this difference, precise probabilism would reccomend that a probability of .5 be assigned to both hypotheses in either case. Here, too, precise probabilism fails to be appropriately responsive to the evidence. In addition, evidence can accumulate in a way that does not require changing our initial probability assignments. It could be that, at first, one has stronger evidence for A than for B. So the probability assigned to A should be greater than the probability assigned to B. As more evidence accumulates, suppose the overall evidence still favors A over B to the same degree. No change in the probabilities is thus required. But something has changed about the agent's state of uncertainty towards A and B: the quantity of evidence on which the agent can make their assessment that A is more probable than B is larger. And yet, this change is not reflected in the precise probabilities assigned to these propositions.

3 Imprecise probabilism

What if we give up the assumption that probability assignments should be precise? Imprecise probabilism (IP) holds that a rational agent's credal stance towards a hypothesis is to be represented by 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 with the evidence (more on this point later).[2] It is easy to see that modeling an

[1] Here is another problem for precise probabilism. Imagine a rational agent who does not know the bias of the coin. For precise probabilism, this state of uncertainty should be represented by a .5 probability assignment to the *heads*. Next, the agent learns that the bias towards heads, whatever the bias is, has been slightly increased, say by .001. The addition of this new information is called *sweetening* in the philosophical literature. This sweetening should now make the agent bet on heads: if the probability of *heads* was initially .5, it must be ever so slightly above .5 after sweetening. But, intuitively, the new information should leave the agent equally undecided about betting on heads or tails. After sweetening, the agent still does not know much about the actual bias of the coin.

[2] For the development of imprecise probabilism, see Keynes (1921); Levi (1974); Gärdenfors & Sahlin (1982); Kaplan (1968); Joyce (2005); Fraassen (2006); Sturgeon (2008); Walley (1991). Bradley (2019) is a good source of further references. Imprecise probabilism shares some similarities with what we might call **interval probabilism** (Kyburg, 1961; Kyburg Jr & Teng, 2001). On interval probabilism, precise probabilities are

agent's credal state by sets of probability measure avoids some of the shortcomings of precise probabilism. For instance, if an agent knows that the coin is fair, their credal state would be represented by the singleton set $\{\mathrm{P}\}$, where P is a probability measure that assigns $.5$ to *heads*. If, on the other hand, the agent knows nothing about the coin's bias, their credal state would be represented by the set of all probabilistic measures, since 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. Rather, the agent's credal state is essentially imprecise and should be represented by means of the entire set of probability measures.

So far as good. But, just as precise probabilism fails to be appropriately evidence-responsive in certain scenarios, imprecise probabilism runs in similar difficulties in other scenarios.

> **Even v. uneven bias:** You have two coins and you know, for sure, that the probability of getting 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. Contrast this with an uneven case. You have four coins and you know that three of them have bias .4 and one of them has bias .6. You pick a coin at random and plan to toss it. You should be three times more confident that the probability of getting heads is .4. rather than .6.

The first situation can be easily represented by imprecise probabilism. The representor would contain two probability measures, one that assigns .4. and the other that assigns .6 to the hypothesis 'this coin lands heads'. But imprecise probabilism cannot represent the second situation. Since the probability measures in the set are all compatible with the agent's evidence, no probability measure can be assigned a greater (higher-order) probability than any other.[3]

These examples show that imprecise probabilism is not expressive enough to model the scenario of uneven bias. Defenders of imprecise probabilism could concede this point but prefer their account for reasons of simplicity. They could also point out that imprecise probabilism models scenarios that precise probabilism cannot model, for example, a state of complete lack of evidence. In this respect, imprecise probabilism outperforms precise probabilism in expressive power, but also retains theoretical simplicity. Unfortunately, it is questionable whether imprecise probabilism actually outperforms precise probabilism *all things considered*. As we will now seee, imprecise probabilism suffers from a number of shortcomings that do not affect precise probabilism.

The first problem has not received extensive discussion in the literature, but it is fundamental. Recall that, for imprecise probabilism, an agent's state of uncertainty is represented by those probability measures that are *compatible* with the agent's evidence. The question is, how should the notion of compatibility be understood here? Perhaps we can think of compatibility as the fact that the agent's evidence is consistent with the probability measure in question. But mere consistency wouldn't get the agent very far in excluding probability measures, as too many probability measures are strictly speaking still consistent with most observations and data. Admittedly, there will be clear-cut cases: if you see the outcome of a coin toss to be

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.

[3]Other scenarios can be constructed in which imprecise probabilism fails to capture distinctive intuitions about evidence and uncertainty; see, for example, (Rinard, 2013). Suppose you know of two urns, GREEN and MYSTERY. You are certain GREEN contains only green marbles, but have no information about MYSTERY. A marble will be drawn at random from each. You should be certain that the marble drawn from GREEN will be green (G), and you should be more confident about this than about the proposition that the marble from MYSTERY will be green (M). In line with how lack of information is to be represented on IP, for each $r \in [0, 1]$ your representor contains a P with $\mathrm{P}(M) = r$. But then, it also contains one with $\mathrm{P}(M) = 1$. This means that it is not the case that for any probability measure P in your representor, $\mathrm{P}(G) > \mathrm{P}(M)$, that is, it is not the case that RA is more confident of G than of M. This is highly counter-intuitive.

heads, you reject the measure with $\mathrm{P}(H) = 0$, and similarly for tails. Another class of cases might arise while randomly drawing objects from a finite set where the true frequencies or objective chances are already known, because the finite set has been inspected. But such clear-cut cases aside, what else? In the end, evidence will often be consistent with a probability measure.[4]

A second, well-known problem for imprecise probabilism is belief inertia. Precise probabilism offers an elegant model of learning from evidence: Bayesian updating. Imprecise probabilism, at least *prima facie*, offers an equally elegant model of learning from evidence, richer and more nuanced. It is a natural extension of the classical Bayesian approach that uses precise probabilities. When faced with 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:

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

The hope is that, if we start with a range of probabilities that is not extremely wide, point-wise learning will behave appropriately. For instance, if we start with a prior probability of *heads* equal to .4 or .6, then those measures should be updated to something closer to .5 once we learn that a given coin has already been tossed ten times with the observed number of heads equal 5 (call this evidence E). This would mean that if the initial range of values was $[.4, .6]$ the posterior range of values should be narrower.

Unfortunately, this narrowing of the range of values becomes impossisble whenever the starting point is complete lack of knowledge, as imprecise probabilism runs into the problem of belief inertia (Levi, 1980). This problem arises in situations in which no amount of evidence could lead the agent to change their belief state, according to a given modeling strategy. Consider a situation in which you start tossing a coin knowing nothing about its bias. The range of possibilities is $[0, 1]$. After a few tosses, if you observed at least one tail and one heads, you can exclude the measures assigning 0 or 1 to *heads*. But what else have you learned? If you are to update your representor set point-wise, you will end up with the same representor set. For any sequence of outcomes that you can obtain and any probability value in $[0, 1]$, there will exist a probability measure (conditional on the oucomes) that assigns that probability to *heads*. Consequently, the edges of your resulting interval will remain the same. In the end, it is not clear how you are supposed to learn anything if you start from complete ignorance.[5]

Some downplay the problem of belief inertia. After all, if you started with knowing truly nothing, then it is right to conclude that you will never learn anything. Joyce (2010) writes:

> You cannot learn anything in cases of pronounced ignorance simply because a prerequisite for learning is to have prior views about how potential data should

[4] Probability measures can be inconsistent with evidential constraints that agents believe to be true. Mathematically, non-trivial evidential constraints are easy to model (Bradley, 2012). They can take the form, for example, of the *evidence of chances* $\{\mathrm{P}(X) = x\}$ or $\mathrm{P}(X) \in [x, y]$, or *structural constraints* such as "X and Y are independent" or "X is more likely than Y." These constraints are something that an agent can come to accept outright, but only if offered such information by an expert whom the agent completely defers to. But, besides these idealized cases, it is unclear how an agent could come to accept such structural constraints upon observation. There will usually be some degree of uncertainty about the acceptability of these constraints.

[5] Here's another example of inertia, coming 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). Suppose your initial credence about these two hypothesis is complete uncertainty with interval. Next, suppose you learn that a marble drawn at random from the urn is green (E). After using this evidence to condition each probability measure in your representor (which initially contains all possible probability measures over the relevant space) on this evidence, you end up with the same spread of values for H_1 that you had before learning E. This holds no matter how many marbles are sampled from the urn and found to be green. This is counterintuitive: if you continue drawing green marbles, even if you started with complete uncertainty, you should become more inclided towards the hypothesis that all marbles are green.

alter your beliefs (p. 291)

The upshot is that vacuous priors should not be used and that imprecise probabilism gives the right results when the priors are non-vacuous. Another strategy is to say that, in a state of complete ignorance, a special updating rule should be deployed.[6]

add reference Joyce, J. M. (2010). A Defence of Imprecise Credences in Inference and Decision Making, *Philosophical Perspectives* 24, pp. 281–323.

Finally, imprecise probabilism faces a third, deeper problem that does not arise for precise probabilism. As it turns out, it is impossible to define proper scoring rules for measuring the accuracy of a representor set of probability measures. Workable *scoring rules* exist for measuring the accuracy of a single, precise probability measure, such as the Brier score. These rules measure the distance between a rational agent's probability measure (also called credence function) and the actual value. A requirement of scoring rules is that they be *proper*: any rational agent will score their own probability measure to be more accurate than any other. After all, if an agent thought a different probability measure was more accurate, they should switch to it. Proper scoring rules are then used to formulate accuracy-based arguments for precise probabilism. These arguments show (roughly) that, if your precise measure follows the axioms of probability theory, no other measure is going to be more accurate than yours whatever the facts are. Can the same be done for imprecise probabilism? It cannot. Impossibility theorems demonstrate that no proper scoring rules are available for representor sets. So, as many have noted, the prospects for an accuracy-based argument for imprecise probabilism look dim (Campbell-Moore, 2020; Mayo-Wilson & Wheeler, 2016; Schoenfield, 2017; Seidenfeld, Schervish, & Kadane, 2012). Moreover, as shown by Schoenfield (2017), if an accuracy measure satisfies certain plausible formal constraints, it will never strictly recommend an imprecise stance, as for any imprecise stance there will be a precise one with at least the same accuracy.

4 Higher-order probabilism

Let us take stock. Imprecise probabilism is more expressive than precise probabilism. It can model the difference between a state in which there is no evidence about a proposition (or its negation) and a state in which the evidence for and against a proposition is in equipoise. But imprecise probabilism has its own expressive limitations: it cannot model the case of uneven bias. In addition, imprecise probabilism faces difficulties that do not affect precise probabilism: the notion of compatibility between a probability measure and the evidence is too permissive; belief inertia trivializes Bayesian updating; and no proper scoring rules exist for imprecise probabilism. In this section, we show that higher-order probabilism overcomes the expressive limitations of imprecise probabilism without falling prey to any such difficulties.

Proponents of imprecise probabilism already hinted to the need of relying on higher order-probabilities. For instance, Bradley compares the measures in a representor to committee members, each voting on a particular issue, say the true chance or bias of a coin. As they acquire more evidence, the committee members will often converge on a chance hypothesis.

> …the committee members are "bunching up". Whatever measure you put over the set of probability functions—whatever "second order probability" you use—the "mass" of this measure gets more and more concentrated around the true chance hypothesis. (Bradley, 2012, p. 157)

But such bunching up cannot be modeled by imprecise probabilism alone: a probability distri-

[6] 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 from a principled account of why one should use a separate special update rule when starting with complete ignorance.