Indifference and Imprecise Probabilities notes

PHIL 735 Week 11
Matt Kotzen and Jim Pryor

November 29, 2023

1 Motivating Imprecise Probabilities

- Two related hesitations about the standard Bayesian framework:
  - Is it plausible to suppose that agents have precise point-valued credences in every proposition in the relevant algebra? Or even in every proposition to which they have some doxastic attitude? What is your precise credence that it will snow in London on New Year's Day 2026 (Mahtani's "NYD")? What could make it the case that your credence in NYD is precisely .24223482734729034, rather than .24223482734729033 or .24223482734729035?
  - Even if it were plausible to suppose that creatures like us *could* have precise credences in propositions like NYD, why think that we always *should*? Often, our evidence that bears on a particular proposition is ambiguous or "fuzzy," so why think that a non-fuzzy ("sharp") credence is the epistemically rational response? Shouldn't our doxastic states "match" our evidence, leading to fuzzy credences in (at least some) situations where the evidence is fuzzy too? See Joyce 2005, Elga 2010.
- As we've seen earlier in the class (HW problem #41) and in the White paper, there are also worries about the standard ("sharp") Bayesian framework that are related to the Principle of Indifference. When we apply the POI, which parameter should we be indifferent over? In some cases, indifference over different parameters (all of which look equally natural) leads to contradictory results. One possibility: find a way to consistently fit the POI into the standard Bayesian framework. Another possibility: abandon the POI. A third possibility: abandon the standard Bayesian framework.
- Idea (Kyburg 1961 and Good 1962): instead of representing an agent's graded doxastic attitude as a real number in the [0,1] interval, perhaps we should instead represent it as a *subinterval* of the [0,1] interval

- So, for example, my doxastic attitude toward NYD might be the interval [.1,.4]
- And my doxastic attitude toward ¬NYD might correspondingly be [.6,.9]

- Notice that that's compatible with agents *sometimes* having point-valued credences; for example, we can represent a credence of precisely .7 as [.7,.7]

- If all the values in the interval assigned to A are higher than all the values in the interval assigned to B—say, if the agent's doxastic attitude toward A is [.6,.7] and their doxastic attitude toward B is [.4,.5]—then we can say that they are more confident in A than in B (and similarly for "less confident")

- If some of the values in A's interval are higher than some of the values in B's interval, but it's also the case that some of the values in B's interval are higher than some of the values in A's interval—say, if the agent's credence in A is [.6,.8] and their credence in B is [.5,.7], then perhaps we can just say that it's neither the case that they are more confident in A than in B, nor that they are more confident in B than in A.

- A thought: perhaps we can avoid the apparent inconsistency of the POI in cube-factory-style cases by assigning an extremely noncommittal interval like [0,1] to each of the relevant propositions (e.g., that the side-length > .5 in, that the volume is >.125 in^3)

- A question: Are doxastic attitudes like [.1,.2] ∪ [.3,.4] allowed? Should they be?

- A worry: Does this interval-valued approach really address the "two related hesitations" from above, or does it just leave us with two (or more) implausibly sharp credal boundaries, rather than one implausibly sharp credal value?

- Another worry: Does this approach enable us to make all of the distinctions that we want/need?
  - Suppose that I am certain that A entails B, and I would report myself as being more confident in B than I am in A by a difference of .1. (Perhaps B is logically equivalent to $A \vee X$, where I regard A and X to be mutually exclusive, and I assign a precise credence of .1 to X.)
  - And suppose that my doxastic attitude toward A is [.3,.5]. Accordingly, my doxastic attitude toward B is fuzzy too: [.4,.6]. On the approach under consideration, it's neither the case that I am more confident in A than in B, nor that I am more confident in B than in A. But that's odd: it's natural to describe me as being more confident (by a difference of .1) in B than in A.
  - Contrast that with a case where I regard A and B to be completely independent of each other, where my credence in A is [.3,.5], and where my credence in B is [.4,.6]. There, it's more plausible that it's neither the case that I am more confident in A than B, nor that I am more confident in B than A.

- Similarly, focusing only on intervals can "miss" the agent's judgments about probabilistic independence, probabilistic correlation, etc.

2 Representors

- Another idea: instead of representing an agent's doxastic state as a single function that assigns a real interval to each proposition, instead represent an agent's doxastic state as a *set* of point-valued probability functions, called the agent's "representor."
- Then, say that an agent's overall doxastic state displays certain properties iff those properties are shared by *every* function in the representor.
  - We do need to be careful about "certain properties" here. E.g., it better not be the case that these properties include the property of assigning a precise point-valued credence to each proposition; each function in the representor does that, but the representor itself does not.
- For example, if every probability function in the agent's representor assigns a value higher than .5 to A, then the agent's credence is higher than .5, even if there's no precise value higher than .5 that every function in her representor assigns to A.
- Similarly, if every function in the agent's representor assigns a value to B that is higher than the value that it assigns to A, then the agent is more confident in B than she is in A.
  - Notice that this can be the case even if some of functions in the representor assign a higher credence to A than some *other* functions assign to B. So, it looks like this approach has the resources to capture at least some distinctions that the interval-valued approach can't capture.
  - Of course, on this approach, we can still say things like "the agent's credence in A is in the range between .2 and .3, and the agent's credence in B is in the range between .3 and .4." But we have to bear in mind that, in making claims like that, we're not saying everything that there is to say about the agent's attitude toward A and B. The full story requires a more complete articulation of all of the probability functions in her representor; the fact that they all assign credences to A in the [.2,.3] range and credences to B in the [.3,.4] range is just one piece of information about what those functions encode regarding A and B.
- A very natural story to tell about updating: updating a representor on E is simply a matter of conditionalizing each function in the representor on E.

3 Questions and challenges regarding representors

- Gappy intervals
  - Let's return to the question from above about whether we should allow for agents to assign ranges like [.1,.2] ∪ [.3,.4] to propositions.
  - In the context of representors, we can think about whether we should accept what Weisberg 2009 calls the Interval Requirement: Suppose Pr_x and Pr_y are two distributions in the representor, and A is some proposition in the algebra. If $Pr_x(A) = x$ and $Pr_y(A) = y$, then for any real number z between x and y, there is a distribution Pr_z in the representor such that $Pr_z(A) = z$.
  - One (though not the only) way to ensure that the Interval Requirement is satisfied is to insist on Convexity: If a representor contains two distributions Pr_x and Pr_y, it also contains every linear combination of those two distributions. I.e., if Matt is in your representor, and Jim is in your representor, then every linear combination of (i.e., every "compromise between") Matt and Jim is in your representor too. (Convexity is logically stronger than the Interval Requirement.)
  - Importantly, every linear combination of two probability distributions is also a probability distribution, so there's no worry here that Convexity will insert distributions into the representor that violate the probability axioms.
  - One complication for Convexity: not every feature shared by both Pr_x and Pr_y is also shared by all linear combinations of Pr_x and Pr_y. We saw this in an earlier session: Jim and Matt might have different credences for A and B but agree that they're probabilistically independent; a 50-50 compromise between us will not necessarily also regard A and B to be probabilistically independent.

- Open vs. closed intervals and "belief inertia"
  - Consider a scenario in which you have just selected a coin from a bag, knowing only that the bag contains various coins, some of which may be biased to various unspecified degrees. You are going to toss the coin 25 times, and before you begin tossing the coin you contemplate claim HEADS25—i.e., the proposition that the coin will land heads on its 25^{th} toss.
  - Plausibly, your initial credence in HEADS25 should be very fuzzy
  - Should it be (0,1) or [0,1]?
  - The closed interval [0,1] is associated with a representor that includes a probability function that assigns 1 to HEADS25, and another probability function that assigns 0 to HEADS25, both of which violate Regularity. Also, a probability function that assigns 1 to the proposition that the coin is 100% tails-biased can't even be conditionalized on the coin landing heads (since that probability function assigns an expectedness of 0 to the coin landing heads).

- So maybe the open interval (0,1) is the way to go.
- But: suppose that you draw the first coin, and it's heads. As Mahtani explains (pp. 116-17), even if every function in your representor increases the probability it assigns to HEADS25 in response, your posterior credence in HEADS25 will remain (0,1). "Thus your credence in HEADS25 will not shift from the range (0,1) no matter how much evidence you amass in favor of HEADS25."

- Dilation
  - White presents the following puzzle: Consider some claim X that you're maximally fuzzy about—your credence in X is [0,1]. You know the following: I know whether X is true, and I took a fair coin and covered up the heads side with a statement of the truth regarding X and I covered up the tails side with a statement of what's false regarding X—i.e., if X is true, then I wrote X on the heads side and $\neg X$ on the tails side, and if X is false, then I wrote $\neg X$ on the heads side and $\neg X$ on the tails side. Your credence at t_0 (before tossing the coin) in HEADS is .5, since you know that the coin is fair. I toss the coin. At t_1, you see the coin land X-side up. So, at t_1, you know that HEADS is true iff X is true, so you assign the same credence to HEADS and to X.
  - Should your credence at t_1 in both HEADS and in X be [0,1]—i.e., should your credence in HEADS "dilate" to [0,1]? That seems weird. Your credence at t_0 in HEADS was .5, all you've learned is that the coin landed X-side up, and you have no information at all about whether X is true.
  - Should your credence at t_1 in both HEADS and in X be .5—i.e., should your credence in X "sharpen" to .5? That seems weird too. Your credence in X was [0,1], and regardless of whether the coin had landed X-side up or $\neg X$-side up, you'd have equally good reasons to sharpen your credence in X to .5. So, by the Reflection Principle, it seems as though you were in a position to sharpen your credence in X to .5 at t_0, without having to go through the whole coin procedure.
  - Conclusion: credences should be sharp; your doxastic state is best modeled by a single probability function, not by a set of containing many probability functions.

- Decision Theory
  - How do we do decision theory on a representor approach?
  - When act A_1 has a higher expected value than act A_2 on *every* probability function in an agent's representor, it's fairly uncontroversial that he should prefer A_1 to A_2.

- But things get trickier when A_1 has a higher expected value than A_2 on some functions in the agent's representor, and yet A_2 has a higher expected value than A_1 on some other functions in the agent's representor.
- Permissive choice rules say: the agent may rationally perform any action that is recommended by at least one of the functions in her representor.
- Maximin says: for each contemplated action, calculate the expected utility according to each function in the representor, and find the lowest expected utility for that action. Choose the action that has the highest minimum expected utility.
- Elga 2010: it's an adequacy constraint on decision theories that they not permit agents to pass up guaranteed gains (at least where that's the only relevant difference between two contemplated options). Suppose that I offer to sell you a ticket for \$.65 that pays \$1 if A is true. I then offer to sell you a ticket for \$.30 that pays \$1 if A is false. Purchasing both tickets guarantees you a net gain of \$.05 in every possible world. So, according to Elga, no adequate decision theory should permit you to decline both bets.
- Exercise: Does a permissive choice rule satisfy this adequacy constraint? Does Maximin?