Separability Assumptions in Scoring-Rule-Based Arguments for Probabilism

Lara Buchak and Branden Fitelson
UC Berkeley
buchak@berkeley.edu branden@fitelson.org
May 30, 2009

Separability in Decision Theory

- We'll start with a classic debate in decision theory: whether the utility function should be separable across states.
- In decision theory, an agent is deciding how to value a gamble that results in different outcomes in different states. Each outcome gets a utility value for the agent.
- We are interested in how to aggregate the values of the outcomes in different states; that is, how to evaluate a gamble. The standard answer to this question is that we should maximize expected utility.
- A worry about this is that it doesn't take into account how actual agents value *risk:* for example, if two gambles yield the same average monetary payoff, then agents tend to prefer the gamble with less variance.
- Some of this can be accounted for by allowing that the utility function diminishes marginally.
- However, the Allais paradox shows that diminishing marginal utility functions cannot fully account for typical attitudes towards risk.

Separability in Decision Theory

- In the Allais Paradox, agents are presented with a choice between A and B and a choice between C and D, where the gambles are as follows:

A: $5,000,000 with probability 0.1, $0 otherwise.
B: $1,000,000 with probability 0.11, $0 otherwise.

C: $1,000,000 with probability 0.89, $5,000,000 with probability 0.1, $0 with probability 0.01.
D: $1,000,000 with probability 1.

- People tend to choose A over B, and D over C, but there are no utility values we can assign to $0, $1M, and $5M such that these choices maximize expected utility.
- There are two potential responses to the Allais paradox on behalf of standard decision theory. The first claims that the standard preferences are simply *irrational*, and the second claims that we've *misdescribed* the outcomes in some way; e.g. "$0" in gamble C is really "$0 and regret" or "$1M" in gamble D is really "$1M and the pleasant feeling of certainty before the gamble is taken."
- However, if we think that the Allais preferences are rational and correctly described, then what this paradox shows is that rational people care about *global properties* of gambles, e.g. the minimum, the variance, etc.
- The axiom of Savage's decision theory that explicitly rules this out (and that the Allais preferences clearly violate) is the Sure-Thing Principle.

Sure-Thing Principle (STP): For all X and Y (and for all E): If E and ~E are mutually exclusive and exhaustive events (sets of states), where E is not the null event, then, for all Z and W, I prefer {X if E, Z if ~E} to {Y if E, Z if ~E} iff I prefer {X if E, W if ~E} to {Y if E, W if ~E}.

Event	E	~E		Event	E	~E
Gamble A	X	Z		Gamble C	X	W
Gamble B	Y	Z		Gamble D	Y	W

- Note that E ranges over *events*, rather than states (and therefore that X, Y, Z, and W might themselves be gambles). So the Sure-Thing Principle is a kind of event-wise dominance, rather than state-wise dominance.
- Many people mischaracterize the question of whether to accept STP as the question of whether possible worlds should affect the value of the actual world: for example, whether Z or W should affect the value of X.
- But the real question is how to aggregate values across possible worlds. STP entails that only local properties (i.e. what happens in a particular state) can make a difference to the value of a gamble, while those who reject STP think that global properties, such as what happens in the worst state, can make a difference to the value of a gamble.
- For example, it might be that pairing X with Z rather than with W affects the variance more than pairing Y with X rather than W. Or it might be that gamble A and gamble C have very different minimums, but gamble B and gamble D do not. Again, the question is whether these properties should be allowed to make a difference.

Scoring Rule Arguments for Probabilism

- The decision to adopt a particular credence function (or forecast) f over events is thought of as the decision to take a particular gamble that gives you an "epistemic utility" in each state of the world. Epistemic utility is supposed to be the utility of a having a particular forecast in a particular state of the world, if the only thing we value is how close our forecast is to the truth.
- Epistemic utility is usually conceived of as a penalty based on the difference between the forecast for a proposition (or propositions) and the truth value of that proposition (or those propositions) in that particular state of the world, and I will not question this assumption.
- This penalty is defined as a scoring rule: s(1, f(E)) is the score for my forecast for E in worlds in which X is true, s(0, f(E)) is the score for my forecast for E in worlds in which E is false.
- We can calculate the expected epistemic utility of a forecast, relative to one's current forecast. A scoring rule is **proper** if **whenever one's current forecast is coherent,** it assigns a **lower (better) expected epistemic utility** to one's current forecast than to any other forecast. In other words, if one's current forecast is a probability function, epistemic utility considerations do not give one a reason to adopt a different forecast instead.
- Scoring Rule Arguments for probabilism use the assumption that scoring rules should be proper to argue that there is something wrong with being incoherent. For example, take this theorem:

Theorem (Predd et al): Assume we have a set of states; events (subsets of the state set) about which an agent makes a forecast; and a proper scoring rule S. Then for each forecast f, if f is coherent then f is not weakly dominated over the set of states, relative to S, by any other forecast; and if f is incoherent, then f is strongly dominated over the set of states, relative to S, by some coherent forecast.

Separability in Scoring Rule Arguments

- The assumption that scoring rules should be proper only makes sense if the epistemic utility of adopting a particular forecast is a separable function of the epistemic utility of that forecast in each state (in particular, if the epistemic utility of a gamble is its expected epistemic utility).
- But it doesn't *follow* from the fact that we care only about truth that the epistemic utility of a forecast should be a separable function of the epistemic utility of the forecast in each possible state of the world.
- On the contrary, caring only about truth leaves open that we might aggregate the epistemic utilities in the individual states in a variety of ways, just as caring only about consequences leaves open how to value gambles.
- For example, in the epistemic case, we might care about not being too far away from the truth in any one state of the world. So when given the choice between moving one forecast closer to the truth by 0.2 and another closer to the truth by 0.2, we might prefer that the belief antecedently farther from the truth gets moved. Or, when given the choice between two forecasts that are the same except one is closer to the truth by 0.2 in one world and the other is closer by 0.2 in another (equiprobable) world, we might prefer the one whose maximum value from the truth in these two worlds is lower.
- That is, all things equal, we might prefer that there is no world in which we are *really* wrong about what's true, in which our forecast is *way* off.

Separability in Scoring Rule Arguments

- The intuition that we might prefer to get incrementally closer to the truth in states in which we are far away from the truth initially seems to be accounted for by allowing that the scoring rule increases marginally (or the penalty diminishes marginally as we get closer to the truth). For example, this is true of the Brier score:

  S(0, 0.2) = 0.04. S(0, 0.4) = 0.16. S(0, 0.6) = 0.36.

  The difference in epistemic utility for a change of 0.2 units is greater the farther away from the truth we are.
- This is analogous to the way diminishing marginal utilities are supposed to handle risk aversion.
- Both responses say that only local properties matter. They claim that when we have the intuition that a particular global property matters (e.g. variance; or maximum distance from the truth), we are actually responding only to particular local properties (e.g. distance of some particular outcome from $0; distance of some particular belief from the truth).
- To see what this response rules out, consider an example:
- Two forecasts, f and g, have the same expected Brier score. Forecast f is not more than 0.6 "units" away from the truth in any state of the world, but it's not closer than 0.2 units away either. However, forecast g can be very good or very bad, depending on what state of the world we're in: in some states, it is exactly right, and in some states, it is 0.9 units from the truth. Caring about *expected* epistemic utility – and, more generally, requiring that the value of forecasts be separable across states of the world – rules out that a rational epistemic agent can prefer f because he will not be "really wrong" come what may, and it rules out that a rational epistemic agent can prefer g because he has some possibility of being "spot on."

Intuitions about Aggregating Epistemic Utility

Setup: You know X is true. You're deciding to among gambles that will give you a particular degree of belief in X depending on which state obtains.

We could cash this out in many ways, some of them fanciful, some of them more realistic:

- You are getting into a fictional "degree of belief" machine. Based on a chance mechanism, the machine will cause you to undergo a process which will result in your having a particular brain state.
- You are deciding whether to take a particular drug. It has a good chance of enhancing your memory – and thus allowing you to believe X to degree 1 tomorrow – but it also has a non-negligible chance of making your memory worse, so you will believe X to degree 0.5 tomorrow. If you don't take the drug, your memory will just fade a bit, and you will believe X to degree 0.7 tomorrow.
- You are choosing among graduate schools, all of which will indoctrinate you to a certain degree. Having different advisors will lead to different credences in X, and your advisor is determined by a lottery. At school A, there is an advisor with whom you will believe X to 0.9 and one with whom you will believe X to 0.1. At school B, all of the advisors will lead you to have a credence of 0.6 in X.
- X is something that is hard to remember, so you want to store it in your memory using a mnemonic device. Different mnemonic devices will cause you to have different degrees of certainty in X, but some are harder to remember. So, for example, you might decide between a mnemonic that will give you a credence of 0.9 in X if you remember it correctly, but a credence of 0.1 in X if you remember incorrectly (and you have, say, a 60% chance of remembering it correctly), and a mnemonic that you are sure you will remember correctly but will only give you a 0.6 credence in X.

An Epistemic Allais paradox?

- Again, you know X is true. Now you have to decide between some pairs of gambles.

- Caring only about epistemic considerations, which would you rather have a:

A: 10% chance of believing X to degree 0.9, believe X to degree 0.01 otherwise.
B: 11% chance of believing X to degree 0.8, believe X to degree 0.01 otherwise.

- Caring only about epistemic considerations, which would you rather have a:

C: 89% chance of believing X to degree 0.8, 10% chance of believing X to degree 0.9, 1% chance of believing X to degree 0.01.
D: 100% chance of believing X to degree 0.8.

- If this is analogous to the Allais case, most people will prefer A to B and D to C. However, if an agent is an expected epistemic utility maximizer, no epistemic utility function (scoring rule) can make A preferable to B and D preferable to C.

- If we do have these preferences, then arguments involving propriety miss something of value: namely, the value of **global epistemic properties** in adopting a particular credence function.

Possible Responses

- We might suggest analogues of the responses that are given for the decision theory version of the Allais paradox.
- One response is to simply claim that an agent with these preferences is being irrational in having these preferences.
- This response is plausible in the decision theory case only if we think that the sure-thing principle (or maximizing expected utility) is somehow *constitutive* of rationality. It's hard to argue from some other principle of rationality to the claim that we should obey the sure-thing principle.
- But if we assume that STP is constitutive of rationality, we might as well adopt representation theorem arguments for probabilism; it is unclear that scoring-rule arguments add anything here.
- And remember that scoring rule arguments are not supposed to assume probabilism. But is hard to see why we should care about expected value unless we are already probabilists: why maximize a weighted sum if that weighted sum is not an average?
- Reply? Strictly speaking, the assumption of propriety is just that probabilists care about expectation. But still, why should we require that non-probabilists use scoring rules that are proper?

Possible Responses

- Another response, again an analogue of a response sometimes given for the decision theory version, is to claim that we've misdescribed the outcomes.
- For example, we might say that "Believing X to degree 0.01" is not the same outcome in gamble A as it is in gamble C, because in gamble C that outcome is really "Believing X to degree 0.01 and *regret*." Or "Believing X to degree 0.8" is not the same outcome in gamble B as it is in gamble D, because in gamble D that outcome is really "Believing X to degree 0.8 and the pleasant feeling of certainty before you take the gamble."
- However, all we care about is epistemic utility. And feelings cannot have epistemic value. "Believing X to degree 0.01" and "Believing X to degree 0.01 and *regret*" are not different outcomes from the point of view in which only epistemic considerations are relevant. So, contrary to what the response suggests, these two outcome must have exactly the same epistemic utility!
- Perhaps we could redescribe outcomes to take account of global epistemic properties? For example, the outcome in gamble D is really "believing X to degree 0.2 and not having a chance of believing X to degree lower than 0.2."
- However, if outcomes are this finely described, then it is no longer clear that the requirement that a scoring rule is proper will entail that non-probabilistic credences will be dominated in every world!

Final Thought: Separability in New Joyce Theorem?

- In Joyce's framework, we look at a *partition* of the state space, rather than a set of propositions, to see if we can say anything against an agent's forecast.
- Given some minimal assumptions (not including propriety), Joyce shows that if an agent is incoherent on a partition, then the agent's forecast is dominated on every member of the partition.
- If an agent's forecast is non-probabilistic, then he is incoherent on *some* partition of the relevant state space, and therefore his forecast will be dominated on every member of that partition.
- But note that the agent won't necessarily be incoherent on the most fine-grained partition. So the relevant notion of dominance must be a kind of event-wise dominance, rather than state-wise dominance. That is, the force of the theorem depends on the assumption that if a forecast f is dominated *on a particular partition* of the space by another forecast g, then adopting g is better for the agent than adopting f.
- Accepting event-wise dominance amounts to accepting separability.
- So it seems as if separability might be necessary in this theorem as well?