

Causal Statistical Decision Theory|Why subjunctive probability?

David Johnston

February 18, 2020

I have too many adjectives here. I think I should either go with *consequence maps* and *stochastic consequence spaces* or *subjunctive probabilities* and *subjunctive probability spaces*. The latter is more precise, but also harder to parse before you read the definition.

1 Probability spaces and subjunctive probability spaces

For our causal analysis we work with *subjunctive probability spaces*, a generalisation of the more familiar probability spaces. The subjunctive mood is used to describe hypothetical or supposed states of the world, and subjunctive probability spaces are models that ask for some hypothetical state and give us back a probability space. Subjunctive probability spaces are also different to *conditional probability spaces* Rényi (1956) as *hypothesising* or *supposing* (that is, those things described by the subjunctive mood) are different to *conditioning*, which is more closely analogous to *focussing your attention*.

We use subjunctive probability spaces because *supposition* is a core part of decision problems, one we cannot get away from. In contrast, modelling supposition with conditional probability (which would be necessary if we were to insist on using probability spaces) adds additional structure to our models which isn't clearly warranted and is potentially confusing.

Any decision problem must involve the comparison of different decisions. This comparison takes the form

- *Suppose* I choose the first decision - then the consequence would be X
- *Suppose* I choose the second decision - then the consequence would be Y
- Etc.

If we prefer X to Y , then perhaps we should choose the first decision. We may be ambivalent as to what type of thing X and Y represent, how we ought to determine what values X and Y take or how we ought to determine what

Does this definition need to be here? I didn't know what this word meant before I looked it up

is preferable, but it is hard to do away with supposing that different decisions were taken and that these decisions come with their own consequences and still have what could reasonably be considered a decision problem. This process of supposition implicitly invokes a “subjunctive function” - we provide a hypothetical decision, and we are given a consequence of that hypothetical. Of particular interest are models that, for each hypothetical choice, return a probability distribution over space Ω . Such models are called *supposition functions* by Joyce (2000), but we will call them *consequence maps* in this work. We consider this particular type of subjunctive function - from possible decisions to probability distributions - to be attractive because we consider probability to be a sound choice for modelling uncertainty and stochasticity.

Formally, a consequence map \mathcal{C} is a stochastic function or *Markov kernel* from $D \rightarrow \Delta(\Omega)$, where $\Delta(\Omega)$ represents the set of all probability distributions on Ω . One might recall that, given two random variables $Y : \Omega \rightarrow Y$ and $X : \Omega \rightarrow X$ on some probability space $(\mathbb{P}, \Omega, \mathcal{F})$, the probability of Y conditional on X is a Markov kernel $X \rightarrow \Delta(\mathcal{Y})$. We might then be tempted to consider the consequence map to be a conditional probability on the expanded sample space $\Omega \times D$. However, there are good reasons to keep the concepts of consequence maps and conditional probability separate. Firstly, we have argued that a consequence map is required for a well posed decision problem, but we have not shown that this consequence map must represent some conditional probability, so this would be adding an unnecessary assumption. Secondly, it is not clear that consequence maps ought to behave like conditional probabilities. Consider an expanded set of options:

1. Suppose I choose the first decision d_1 - then the consequence would be P_1
2. Suppose I choose the second decision d_2 - then the consequence would be P_2
3. Suppose I choose either the first or second decision d_1 or d_2 - then the consequence would be ???
4. Suppose I choose d_1 with probability $0 \leq q \leq 1$ and d_2 otherwise - then the consequence would be ???

If we regard the consequence map as a conditional probability then it would be natural to consider the result of the third option to be a unique probability distribution obtained by conditioning on $\{d_1, d_2\}$, equal to $\alpha P_1 + (1 - \alpha) P_2$ for some fixed $0 \leq \alpha \leq 1$. However, there is no obvious reason that these should be mixed in some fixed proportion α - it seems more appropriate to me to say that if d_1 or d_2 is chosen then the result will be P_1 or P_2 .

On the other hand, the result of the fourth option seems like it should be given by $q P_1 + (1 - q) P_2$, at least for most ordinary problems. If we were confronted with a mind reader who could tell the difference between us having chosen d_1 and us having randomised between d_1 and d_2 but come up with d_1 anyway then we might have reason to revise this assumption, but we will generally proceed

under the assumption that such mind readers are absent. For any ambient probability measure, we can choose q not to be equal to α (as defined in the previous paragraph), and so the result will not be equal to the ambient measure conditioned on $\{d_1, d_2\}$.

In general, choosing the consequence map to be a conditional probability requires that we understand there to exist some joint distribution over our decision D and all other random variables in the problem. However, when we adopt hypotheses about what decisions we might make, we throw away key parts of this joint distribution (for example, we throw away the marginal distribution of D) and replace it with whatever we want to suppose instead. Instead of adopting a full joint distribution and then throwing away some parts to get hold of the consequence map, as we would if we were working with a probability space, a subjunctive probability space only supplies those parts which we intend to keep - i.e. in only supplies the consequence map.

References

- James M. Joyce. Why We Still Need the Logic of Decision. *Philosophy of Science*, 67:S1–S13, 2000. ISSN 0031-8248. URL www.jstor.org/stable/188653.
- A. Rényi. On Conditional Probability Spaces Generated by a Dimensionally Ordered Set of Measures. *Theory of Probability & Its Applications*, 1(1): 55–64, January 1956. ISSN 0040-585X. doi: 10.1137/1101005. URL <https://epubs.siam.org/doi/abs/10.1137/1101005>.

Appendix: