Probability and triviality

Day 4 // ESSLLI, Conditionals and Information Sensitivity // David Boylan and Matt Mandelkern

1 Probability Primer

A probability measure Pr is a function from σ-algebra over W to $[0, 1]$ s.t.

1. (Normality) $Pr(\top) = 1$
2. (Additivity) If $A \cap B = \emptyset$ then $Pr(A \vee B) = Pr(A) + Pr(B)$

$\top$ is the top element of the algebra, W itself. Intuitively though the top is the tautology.

Where Ω is a set of possible worlds, we can treat the elements of the algebra as propositions, and generally can assume the algebra is the powerset algebra.

propositions are called 'events' by statisticians

Lots of basic rules fall out of this definition:

We write ab for $a \wedge b$. I write $>$ for $>_i$.

- (Negation.) $Pr(\neg A) = 1 - Pr(A)$
- (Equivalence.) If A is equivalent to B, then $Pr(A) = Pr(B)$
- (General additivity.) $Pr(A \vee B) = Pr(A) + Pr(B) - Pr(A \wedge B)$
- (Decomposition.) $Pr(A) = Pr(A \wedge B) + Pr(A \wedge \neg B)$

On the orthodox Bayesian picture, probability measures can be used to represent *degrees of confidence*;

- $Pr(A) > Pr(B)$ may represent that I am more confident in A than I am in B.
- Here the probability axioms are taken to be norms of rationality.

Highly idealised norms, of course; recall Normality.

Conditional probability of B given A, $Pr(B|A)$:

- On the subjective interpretation of probability, intuitively the probability of B on the supposition that A.
  - E.g. suppose that the first card drawn in a pack is the ace of spades. What is the probability that the next card is an ace?
- Obeys the *Ratio Formula*

  Ratio Formula. If $Pr(A) > 0$ then $Pr(C|A) = \frac{Pr(A \wedge C)}{Pr(A)}$

  Some take this to be *definitional* of conditional probability; others simply take it to be a norm of rationality (at least for sub. probability).
- Given the Ratio Formula, easy to show that $Pr(\cdot|A)$ is itself a probability measure, when defined.

Ratio Formula gives us some further rules:

- (Conjunction.) $Pr(A \wedge B) = Pr(A) \times Pr(B|A)$

Trivial, given Ratio Formula.

- (Law of Total Probability.) $Pr(A) = Pr(A|B)Pr(B) + Pr(A|\neg B)Pr(\neg B)$

Combine Decomposition and Conjunction.

Bayesian picture also includes a norm of *belief revision*.

- Suppose I have seen the first card drawn is the ace of spades. What should my *new* probability be that the second will be an ace as well?
  - Natural answer: the same as my *old* conditional probability!
- Norm of conditionalisation:
  - If Pr is my prior probability function, then my posterior probability function after getting evidence E, Pr_E, should be my old conditional probability function $Pr(\cdot|E)$.
  - i.e. $Pr_E = Pr(\cdot|E)$

2 Why Probability?

Simple case:

- Suppose I am $\frac{1}{2}$ confident this fair coin was tossed at 12pm. What's the probability that *if this fair coin was flipped at 12pm, it landed heads*?
  - Intuitively $\frac{1}{2}$; i.e. $Pr(H|F)$.

Just as we might want a theory of conditionals to predict judgements of truth or validity, we might want it to predict judgements like this about probability.

- More data!
- And many alternatives to our baseline semantics do badly in even this extremely simple case.

Case 1: materialism.

- Remember that $A \supset C$ is equivalent to $\neg A \vee C$.
- So $Pr(\mathit{flipped} > \mathit{heads})$ should be $Pr(\neg \mathit{flipped}) + Pr(\mathit{heads})$

Note that $Pr(\neg \mathit{tossed} \wedge \mathit{heads})$ will be 0 here.

- Can calulate $Pr(\mathit{heads})$ to be $\frac{1}{4}$ given LoTP.

$$Pr(H) = \underbrace{Pr(H|F)Pr(F)}_{1/4} + \underbrace{Pr(H|\neg F)Pr(\neg F)}_{0}$$

- So according to materialism, $Pr(F > H) = \frac{3}{4}$!

Case 2: It's also prima facie inconsistent with strictism.

Though see Rothschild 2013 for an interesting strictist approach. The key here is to pick the right modality for $\Box$.

- Recall on strict view that $F > H$ is equivalent to $\Box(F \supset H)$

- $\Box(F \supset H)$ is equivalent to $\neg\Diamond(F \wedge \neg H)$.
- But $Pr(\Diamond(F \wedge \neg H))$ seems like it should be 1. After all, maybe the coin was tossed and landed tails!
  - So $\Box(F \supset H)$ should be 0!

Motivation for CEM:

- Each of the following get 0.5:

(1) If I flip the coin, it will land heads.

(2) If I flip the coin, it won't land heads.

- Plausibly the two are exclusive here; and so the following should get probability 1:

(3) If I flip the coin, it will land heads $\vee$ if I flip the coin, it won't land heads.

3 The Thesis

Generalisation: seems like your probability in a conditional should just be the corresponding conditional probability.

- *The Thesis*: For all $a, c, Pr : Pr(a > c) = Pr(c|a)$ if $Pr(a) > 0$
  - Say that *Pr yields the Thesis* for a conditional operator $>$ iff *Pr* and $>$ satisfy the above equation.

Due to Adams 1975; Stalnaker 1970. Sometimes called 'Adams' Thesis', 'Stalnaker's Thesis', or 'The Equation'.

- Could be thought of as an extra rational constraint analogous to Normality and Additivity; though some, particularly "non-factualists", think

Further cases:

- Very plausible in many cases, e.g. what's the probability of:

(4) If the die lands on an even number, it will land on 2.

  - Natural answer: $\frac{1}{3}$, the conditional probability.

- Note that these are not judgments about 'probability' conditionals:

(5) There is $\frac{1}{6}$ probability that, if the die is rolled, it will land on 2.

(5) is true. (4) may or may not be true, but it has probability $\frac{1}{6}$. Both facts need to be accounted for. Our focus is on the latter.

4 Triviality 1

(Lewis, 1976)

Assumptions:

1. All rational probability functions yield the Thesis.
2. The class of rational probability functions is closed under conditionalisation.
   - That is, if Pr is rational, then so too is Pr_E.
3. $0 < Pr(c) < 1$ and $Pr(ac) > 0$ and $Pr(a\neg c) > 0$,

Now we have:

$$Pr(a > c) = Pr(a > c|c)Pr(c) + Pr(a > c|\neg c)Pr(\neg c)$$

by the law of total probability. $Pr(b|a) = \frac{Pr(ab)}{Pr(a)}$

$$Pr(a > c|c) = Pr_c(a > c) = Pr_c(c|a) = 1$$

by *The Thesis*, where Pr_c is obtained from Pr by conditioning on c

$$Pr(a > c|\neg c) = Pr_{\neg c}(a > c) = Pr_{\neg c}(c|a) = 0$$

by *The Thesis*, where $Pr_{\neg c}$ is obtained from Pr by conditioning on $\neg c$

$$Pr(a > c) = \underbrace{Pr(a > c|c)}_{1} Pr(c) + \underbrace{Pr(a > c|\neg c)}_{0} Pr(\neg c) = Pr(c)$$

But it's implausible that the probability of a conditional equals the probability of its consequent:

(6) If the car crashes, the airbag will go off.

5 Triviality 2

Assumptions:

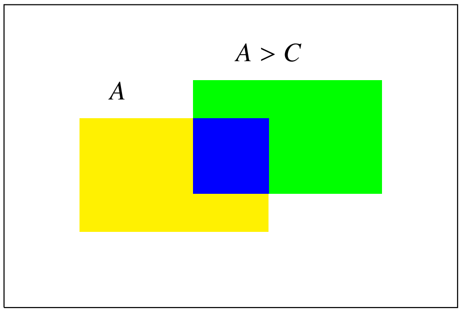
- All rational probability functions yield the Thesis.
- For some rational Pr, $0 < Pr(c) < 1$ and $Pr(ac) > 0$ and $Pr(a\neg c) > 0$,
- For that same Pr, $Pr_{a\vee(\neg a\wedge a>c)}$ is rational.

The argument informally:

- Since Pr is rational, $Pr(a > c) = Pr(c|a)$
- Now consider $Pr_{a\vee(\neg a\wedge a>c)}$.
  - $Pr_{a\vee(\neg a\wedge a>c)}(c|a)$ cannot change: it must be $Pr(c|a)$.
  - $Pr_{a\vee(\neg a\wedge a>c)}(a > c)$ *must* change: since we discard worlds where $a > c$ is false, but not worlds where it's true.

Thanks to Snow for this lovely picture of logical space.

Logical Space



- So the Thesis does not in fact hold on $Pr_{a\vee(\neg a\wedge a>c)}$. Contradiction.

The argument formally:

1. $Pr_{a\vee(\neg a\wedge a>c)}(a>c) = Pr(a>c|a)Pr(a|a\vee(\neg a\wedge a>c)) +$
$$Pr(a>c|\neg a\wedge a>c)Pr(\neg a\wedge a>c|a\vee(\neg a\wedge a>c))$$

2. $Pr(a>c|a) = \frac{Pr(a\wedge a>c)}{Pr(a)} = \frac{Pr(ac)}{Pr(a)} = Pr(c|a)$ by strong centering

$$Pr(a>c|\neg a\wedge a>c) = 1, \text{ so}$$

$Pr_{a\vee(\neg a\wedge a>c)}(a>c) = Pr(c|a)Pr(a|a\vee(\neg a\wedge a>c)) + Pr(\neg a\wedge a>c|a\vee(\neg a\wedge a>c)) \geq$

$Pr(c|a)Pr(a|a\vee(\neg a\wedge a>c)) + Pr(c|a)Pr(\neg a\wedge a>c|a\vee(\neg a\wedge a>c)) = Pr(c|a)$.

But $Pr_{a\vee(\neg a\wedge a>c)}(c|a) = Pr(c|a)$.

So if we start with a Thesis-friendly probability measure, we can easily get to a Thesis-unfriendly one, just by conditionalizing on $a\vee(\neg a\wedge a>c)$.

6 Triviality 3

(Hajek and Hall, 1994)

Assumptions:

1. As before.
2. For some Pr, both Pr and $Pr(\cdot \mid A)$ are rational;
3. For that same Pr, there is some B s.t. $0 < Pr(A\wedge B) < Pr(A) < 1$

Letting $C = (\neg A\vee B) > \neg A$, we thus have

$$(i)\ Pr(C) = Pr(\neg A \mid \neg A\vee B)$$
$$(ii)\ Pr(C \mid A) = Pr_A(\neg A \mid \neg A\vee B) = 0$$

By (ii), $Pr(C) \leq Pr(\neg A)$, so by (i),

$$Pr(\neg A \mid \neg A \vee B) \leq Pr(\neg A).$$

But

$$Pr(\neg A | \neg A \vee B) = \frac{Pr(\neg A \wedge (\neg A \vee B))}{Pr(\neg A \vee B)} = \frac{Pr(\neg A)}{Pr(\neg A \vee B)}.$$

So we have

$$\frac{Pr(\neg A)}{Pr(\neg A \vee B)} \leq Pr(\neg A)$$

and hence

$$Pr(\neg A \vee B) = 1$$

and thus

$$Pr(A \wedge \neg B) = 0.$$

But this is inconsistent with the stipulation that $Pr(A \wedge B) < Pr(A)$.

7 Triviality 4

Bradley notes that The Thesis entails

Preservation: $Pr(c) = 0 \wedge Pr(a) > 0 \rightarrow Pr(a > c) = 0.$

Assumptions:

1. Same as before.
2. If A is consistent, then there is some reasonable probability function such that $Pr(A) = 1$

Argument:

- From 1 we get that *Preservation* holds for all probability measures.
- Then either a or $a > c$ must entail c.
  - Otherwise, there will be parts of the state-space where $(a > c) \wedge \overline{c}$ holds and likewise parts where $a\overline{c}$ holds.
  - Assign the disjunction of these states probability 1 and we'll have a counterexample to *Preservation*.

8 What to say?

Two options:

1. Reject the claim that rational probability functions are closed under conditionalisation.
   - Radical — this is a core part of orthodox Bayesianism.

- Notice as well we must deny it in some very simple cases: there will be cases where Pr is rational, but $Pr(\cdot|A)$ is not!
- Not immediately obvious how it helps with Preservation argument.

2. Restrict the range of measures which yield the Thesis somehow.

First reason for option 2: Thesis imposes a strong constraint. (Ellis, 1978)

- Note that The Thesis holding is equivalent to a conditional being probabilistically independent of its antecedent, given Strong Centering:

$$Pr(a > c) = Pr(c|a) \equiv$$
$$Pr(a > c) = \frac{Pr(ac)}{Pr(a)} \equiv$$
$$Pr(a > c) = \frac{Pr(a \wedge (a > c))}{Pr(a)} \equiv$$
$$Pr(a > c) = Pr(a > c|a)$$

- As Stalnaker (1974) pointed out, once we see that, it's not really clear why we would expect The Thesis to hold in general, or why we would want it to.
- Not clear that this restricts the Thesis enough to avoid the triviality arguments.

That raises the question: can we construct cases where the conditional is intuitively not probabilistically dependent of its antecedent? Something we'll return to.

A different reason to restrict: *context-sensitivity*.

- We've seen abundant reason to think the sentence $\ulcorner$a > c $\urcorner$ is *context sensitive*.
  - In a context where the relevant information is i it might express the proposition $A >_i C$; if the relevant information is some other set of worlds i' it will express some potentially distinct proposition $A >_{i'} C$
- Once we make this distinction, we are forced to make some choices.
  - Given a particular conditional $>_i$, which rational probability measures yield the Thesis for $>_i$?
  - Likewise, given a particular probability measure Pr, which conditionals $>_i$ yield the Thesis for Pr?

A natural idea: the conditional should be coordinated with your evidence:

- Let E_{Pr} be the strongest proposition such that $Pr(E) = 1$.
- Assume that for each rational probability function, $>_{E_{Pr}}$ yields Stalnaker's Thesis wrt to Pr. Call this assumption *Locality*.

  - Roughly, coordinate your conditional probabilities with *your* conditional.

- Seems pretty intuitive:

  - Suppose E_A is Alice's total evidence and E_B is Billy's.
  - Why should Alice have to coordinate her probabilities in $A >_{E_B} C$ — Billy's conditional — with her conditional probabilities?

9 Back to the Triviality Arguments

Response to Triviality 1:

- Assumption 1 fails: there's no particular conditional which yields Stalnaker's Thesis on all reasonable probability functions.

- Particular step where the argument fails:

  - Suppose Pr satisfies the other assumptions.
    - Again, say that E_{Pr} is the strongest proposition such that $Pr(E) = 1$.
  - Locality says that $Pr(A >_E C) = Pr(C|A)$.
  - It does not follow from Locality that $Pr_C(A >_E C) = Pr(C|A)$
    - Instead Pr_C yields the Thesis for a *different* conditional, namely $A >_{E \cap C} C$.

Triviality 1 clearly fails then. What about the others?

References

Adams, E. (1975). *The Logic of Conditionals*. Dordrecht.

Ellis, B. (1978). A unified theory of conditionals. *Journal of Philosophical Logic*, pages 107–124.

Hajek, A. and Hall, N. (1994). The hypothesis of the conditional construal of conditional probability. In Eells, E., Skyms, B., and Adams, E. W., editors, *Probability and conditionals: Belief revision and rational decision*. Cambridge University Press.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85(3):297–315.

Rothschild, D. (2013). Do indicative conditionals express propositions? *Noûs*, 47(1):49–68.

Stalnaker, R. (1974). Letter to Bas Van Fraassen.

Stalnaker, R. C. (1970). Probability and conditionals. *Philosophy of Science*, 37(1):64–80.