Probability in Philosophy

Brian Weatherson

Rutgers and Arché

June, 2008

Questionnaire

I'm not sure how much knowledge everyone already has, so I'd like to start with a little questionnaire. On a card, say for each of the following topics whether you're familiar with the topic, have heard of it but aren't familiar with it, or have never heard of it.

1. Countable vs Uncountable Sets
2. Axiom of Choice
3. Set Theories Intermediate between ZF and ZFC
4. Lesbegue Measure
5. Kripke Semantics for Intuitionistic Logic
6. Dutch Book Arguments
7. Representation Theorems concerning Utility Functions

Probability Functions

- Domain: Elements of some lattice
- Assume that we can define conjunction, disjunction, top and bottom

Probability Functions

- Domain: Elements of some lattice
- Assume that we can define conjunction, disjunction, top and bottom
- Range: [0, 1]
- Note that the range, at least classically, doesn't include non-reals, especially infinitesimals

Possible Domains

Lots of things have the right structure to be the domain.

- Sentences
- Propositions
- Sets

Possible Domains

Lots of things have the right structure to be the domain.

- Sentences
- Propositions
- Sets

We'll largely follow (mathematical) orthodoxy and take sets to be the domain.

Possible Domains

Lots of things have the right structure to be the domain.

- Sentences
- Propositions
- Sets

We'll largely follow (mathematical) orthodoxy and take sets to be the domain.

This has consequences, for, e.g. the status of the claim that logically equivalent propositions have the same probability.

Axioms

- A field of sets is a subset of the powerset of some set X that is closed under intersection, union and (relative to X) complementation.
- A probability function is any function from a field $F \subseteq P(X)$ to $[0, 1]$ satisfying

1. $Pr(X) = 1$
2. If $A, B \in F$ and $A \cap B = \emptyset$, then $Pr(A) + Pr(B) = Pr(A \cup B)$

Logical Equivalents

- We're taking the inputs to be sets
- So 'two' logically equivalent propositions are the same set
- So it's a structural fact that if A $\dashv\vdash$ B, then Pr(A) = Pr(B)

Logical Equivalents

- We're taking the inputs to be sets
- So 'two' logically equivalent propositions are the same set
- So it's a structural fact that if $A \dashv\vdash B$, then $Pr(A) = Pr(B)$
- Alternatively, we could take propositions to be sets of some kind, and we'll have the same result
- That's what I'll do for a while

Logical Consequences

- Assume $A \vdash B$
- Then $B \dashv\vdash A \vee (\neg A \wedge B)$
- Since $A \wedge (\neg A \wedge B) = \emptyset$, we get
- $Pr(B) = Pr(A) + Pr(\neg A \wedge B)$
- Since $Pr(\neg A \wedge B) \geq 0$, because the range of Pr is [0, 1], it follows that $Pr(B) \geq Pr(A)$

Independent Propositions

Similar proofs deliver each of the following results

1. $\Pr(A) = \Pr(A \wedge B) + \Pr(A \wedge \neg B)$
2. $\Pr(B) = \Pr(A \wedge B) + \Pr(\neg A \wedge B)$
3. $\Pr(A \vee B) = \Pr(A) + \Pr(\neg A \wedge B)$

Independent Propositions

Similar proofs deliver each of the following results

1. $\Pr(A) = \Pr(A \wedge B) + \Pr(A \wedge \neg B)$
2. $\Pr(B) = \Pr(A \wedge B) + \Pr(\neg A \wedge B)$
3. $\Pr(A \vee B) = \Pr(A) + \Pr(\neg A \wedge B)$

Putting those three things together we get an important result

- $\Pr(A) + \Pr(B) = \Pr(A \vee B) + \Pr(A \wedge B)$

Alternative Formulations

The above derivations, which are utterly foundational, are perhaps unfortunate in three respects

1. We used distinctively *classical* expansions of propositions
2. We took the equiprobability of equivalents as a structural fact, rather than an axiom
3. We took the non-negativity of probabiilties as a structural fact, rather than an axiom

There is a way to avoid all of these negative features

Alternative Formulations

A probability function with respect to an entailment relation $\vdash$ has as domain a set S of propositions closed under conjunction, disjunction and negation, and as range $\mathbb{R}$), and satisfies the following four axioms for all A, B $\in$ S

1. If A is a $\vdash$-theorem, then $\Pr(A) = 1$
2. If A is a $\vdash$-antitheorem, then $\Pr(A) = 0$
3. If $A \vdash B$, then $\Pr(A) \leq \Pr(B)$
4. $\Pr(A) + \Pr(B) = \Pr(A \vee B) + \Pr(A \wedge B)$

That's equivalent to our original formulation if $\vdash$ is classical implication. It's more general, because it allows for other interpretations of $\vdash$. We might come back to this next week.

Interpretations

There are many real-world functions that (arguably) satisfy these constraints

- Frequencies
- Propensities
- Chances
- (Rational) credences
- Degrees of evidential support

All of these are at least a little controversial, and especially for the epistemological cases, it's quite a substantial philosophical claim that the probability calculus, as so far defined, is of any significance at all.

Conditional Probability Introduced

- As well as being interested in the probability of events (e.g. whether it will rain tomorrow), we're sometimes interested in probabilities conditional on other events (e.g. whether it will rain tomorrow conditional on rain being forecast).
- There is a standard definition for the conditional probability of B given A.

$$Pr(A|B) =_{df} \frac{Pr(AB)}{Pr(B)}, if\, Pr(B) > 0$$

Bayes Theorem

For various purposes, it is useful to remember various conversions of this definition. The most commonly used is Bayes Theorem.

$$Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)}$$

Usually you can reconstruct this whenever you need, but it's helpful to remember.

- The Wikipedia page for Bayes Theorem has some related results that might be helpful to remember.

Independence

We say that A and B are probabilistically *independent* iff $\Pr(AB) = \Pr(A)\Pr(B)$.

This is equivalent to each of the following claims, which in turn justify the name.

- $\Pr(A|B) = \Pr(A)$
- $\Pr(B|A) = \Pr(B)$

The intuitive idea is that taking one as given doesn't change the other.

Division by Zero

- The *If Pr(B)* $>$ *0* constraint is fairly serious
- Let t be the time in seconds a particular particle takes to decay.
- We might wonder something about the nature of t given t's being rational
- Or imagine a coin will be tossed infinitely many times
- What's the probability that it will land heads exactly 20 times, conditional on landing heads at most 20 times?

Domain of Conditional Probability Function

- We might also worry that introducing conditional probability violated structural constraints
- Originally, Pr was a function from propositions/sets to numbers
- But $A|B$ is neither a proposition nor a set
- So it looks like we've somehow changed the subject matter

Lewis's Bombshell

- That A|B isn't a proposition was not always taken as given
- Some folks used to think it was equivalent to a kind of conditional
- Indeed, that it was equivalent to English *If B, then A*
- Lewis showed that could not be the case, given some very weak assumptions

Lewis's Bombshell

Assume there are propositions A, C such that AC, A¬C and ¬A all have positive probability. And assume that | is a conditional, writable as →.

1. $\forall Pr : Pr(A \to C) = Pr(C|A)$ (by hypothesis)
2. $Pr(A \to C|C) = Pr(C|AC)$ (from 1)
3. $Pr(C|AC) = 1$ (since AC entails C)
4. $Pr(A \to C|\neg C) = Pr(C|A\neg C)$ (from 1)
5. $Pr(C|A\neg C) = 0$ (since A¬C entails ¬ C)
6. $Pr(D) = Pr(D|C)Pr(C) + Pr(D|\neg C)Pr(\neg C)$ (Theorem)
7. $Pr(A \to C) = Pr(A \to C|C)Pr(C) + Pr(A \to C|\neg C)Pr(\neg C)$ (substituting A→C for D)
8. $Pr(C|A) = Pr(C)$ (from 1, 3, 5, 7)

But that is too strong in general; not all propositions are independent.

Multiplicative Axiom

We start to avoid the problems by a simple redefinition of conditional probability. The axiom is now.

$$Pr(AB) = Pr(A|B)Pr(B)$$

Conditional Forms of All Axioms

Better, we take all axioms to have conditional form. So our new axiomatisation looks like this, where Pr is a function from pairs of propositions (the second of which is not an antitheorem) to reals.

1. If $C \rightarrow A$ is a $\vdash$-theorem, then $Pr(A|C) = 1$
2. If $C \rightarrow A$ is a $\vdash$-antitheorem, then $Pr(A|C) = 0$
3. If $C \rightarrow A \vdash C \rightarrow B$, then $Pr(A|C) \leq Pr(B|C)$
4. $Pr(A|C) + Pr(B|C) = Pr(A \vee B|C) + Pr(A \wedge B|C)$
5. $Pr(AB|C) = Pr(A|BC)Pr(B|C)$

Recovering Unconditional Probability

We now take "Pr(A)" to be a shorthand for Pr(A|T), where T is some tautology

An Open Question

Say that for each n $\in$ 1, 2, ..., the probability that that there are exactly n jabberwocks is $\frac{1}{2^{n+1}}$.

- What is the probability that there is at least 1 jabberwock?

An Open Question

You might try to reason as follows

- The probability that there is at least 1 jabberwock is the probability that there's exactly 1, plus the probability that there's exactly 2 plus etc
- That is, it's $\frac{1}{4} + \frac{1}{8} + \ldots$
- That is, it's 0.5

That reasoning is not sound given the axioms to date.

Inequality or Equality

The 'addition' axiom we have only works for pairwise addition. You can't infer much from that about infinitary addition.

Inequality or Equality

The 'addition' axiom we have only works for pairwise addition. You can't infer much from that about infinitary addition.

- You can infer something
- Finite addition tells us the probability that there are between 1 and n jabberwocks, for any n
- And as n goes to infinity, that probability goes to 0.5
- Since that proposition entails there are some jabberwocks, the probability that there are some jabberwocks is at least 0.5

Inequality or Equality

In general, if p1, p2, ... are pairwise inconsistent, the most we can prove is

$$Pr(p1 \vee p2 \vee ...) \geq Pr(p1) + Pr(p2) + ...$$

If we want equality, we have to add it as an axiom.

The Flat Distribution Over $\mathbb{N}$

- There is a reason some people resist adding this as an axiom.
- It would rule out a flat distribution over $\mathbb{N}$

The Flat Distribution Over $\mathbb{N}$

- There is a reason some people resist adding this as an axiom.
- It would rule out a flat distribution over $\mathbb{N}$
- Let the domain be subsets of $\mathbb{N}$
- Let F_n be the number of numbers $\leq$n such that F(n), for any predicate F, and define Pr as follows

$$Pr(\mathbf{F}|\mathbf{G}) = \lim_{n \to \infty} \frac{F_n}{(F \wedge G)_n}$$

That would be ruled out by the equality axiom.