Probability in Philosophy

Brian Weatherson

Rutgers and Arché

June, 2008

Probability Functions

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

Classical Probability

- If $\vdash$ is classical implication, those axioms can be simplified a lot.
- We can replace the third axiom with an axiom stating tautologies have probability 1
- And we can replace the fourth axiom with an axiom saying $\Pr(A) + \Pr(B) = \Pr(A \vee B)$ when A and B are disjoint

Classical Probability

- I showed last time that we could prove (classically) that if all equivalent propositions had the same probability, then if A entailed B, B's probability was not less than A
- I should have noted we could start with something (apparently) weaker
- We can prove that if all tautologies have probability 1, then equivalent propositions have the same probability
- Assume A and B are equivalent
- Then $A \vee \neg B$ is a tautology, and A and $\neg B$ are disjoint
- So $Pr(A) + Pr(\neg B) = 1$, so $Pr(A) = 1 - Pr(\neg B)$. But $Pr(B) = 1 - Pr(\neg B)$, so $Pr(A) = Pr(B)$

Truth Tables

- In circumstances where classical truth tables are useful modelling devices, there's a simple way to think about probability.
- So imagine you're in a setting where (a) you don't care about the internal structure of atomic propostions, and (b) you've only got finitely many atomic propositions to consider.
- In such circumstances, probability theory is basically measure theory over the rows of the truth table.
- That is, you generate probability functions by assigning a number (a measure) to each row of the truth tables such that each measure is non-negative, and the measures sum to 1.
- Then Pr(A) is the sum of the measures of the rows on which A is true.

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.

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

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.

Conditional Forms of All Axioms

So as to be clear on what the domain of Pr is, it is perhaps useful to take Pr to be always conditional. This gives the following axiomatisation, 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

- Again, if the logic is classical, the previous axioms can be simplified somewhat
- The idea still is that conditional on any C, probabilities are in [0, 1], logical truths/falsehoods take the extreme values, logically weaker propositions have a higher probability, and the probability of a disjunction of exclusive disjuncts is the sum of the probability of the disjuncts.
- When we don't care about the background C, we can use the simple axiom $Pr(AB) = Pr(A|B)Pr(B)$

An Important Theorem

$$Pr(A) = Pr(AB) + Pr(A\neg B)$$

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

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

$$Pr(A) = Pr(A|B)Pr(B) + Pr(A|\neg B)Pr(\neg B)$$

Note it follows from this that Pr(A|B) and Pr(A|¬ B) can't be on the 'same side' of Pr(A), either both greater than Pr(A) or both less than Pr(A). We'll prove this in two sections time.

Cleaning Up

- I mentioned last time that some of the foundational proofs relied on classical logic, but I didn't prove this.
- I don't have a particularly good answer as to how much work distributivity principles on the underlying lattice do.
- But I can say a bit about how much the fact that the underlying logic is classical does.
- In intuitionistic probability, it's impossible to get from the equality of the probability of equivalent propositions, to the claim that weaker propositions have greater probability.
- And it is impossible to get from the 'simple' addition axiom $\Pr(A) + \Pr(B) = \Pr(A \vee B)$ for disjoint A, B to the 'general' axiom $\Pr(A) + \Pr(B) = \Pr(A \vee B) + \Pr(A \wedge B)$.

Weaker Propositions are More Probable?

- This might be too much 'inside baseball' for those of you who don't care about intuitionistic logic, but it won't take too long.
- Consider a Kripke model with W $= \{1, 2, 3, 4\}$, R $= \{<1,1>, <1,2>, <1,3>, <1,4>, <2,2>, <2,3>, <3,3>, <4,4>\}$ and V(p) $= \{3\}$.
- Define a 'measure' m such that $m(1) = 0.3, m(2) = -0.1, m(3) = 0.5, m(4) = 0.3$.
- Then define Pr(A) = the sum of m(x) over the x that force A
- That satisfies the addition axiom, and equivalent propositions get equal probability, but Pr(p) = 0.5, while Pr($\neg\neg$p) = 0.4

Simple and General Addition

- This one is a bit easier
- Consider a language with just two atomic variables, p and q
- Define a Pr such that $\Pr(A) = 1$ if A is an intuitionistic consequence of $p \vee q$, and 0 otherwise
- That satisfies all the axioms we've considered except the general addition axiom, which fails because $\Pr(p) = \Pr(q) = \Pr(p \wedge q) = 0$, while $\Pr(p\ \mathit{vee}\ q) = 1$.

Infinitary Addition

- Sometimes infinite sums don't seem to be well defined
- For instance, no real number is the intuitive sum of 1+2+3+...
- For a different reason, no real number is the intuitive sum of 1-1+1-1+1...
- We could say that's (1-1)+(1-1)+... = 0+0+... = 0
- Or we could say that it is 1+(-1+1)+(-1+1)+... = 1+0+0+... = 1
- Better to say it is not defined at all

Limits of Sums

- But some sums *converge* on a stable value.
- That is, as we keep on adding terms we get closer and closer to a particular value.
- In that case, we say the sum is the value we converge on.
- More formally, if x_1, x_2, ... are such that

$$\exists s.\forall e > 0.\exists k \in \mathbb{N}.\forall n > k.\sum_{j=1}^{n} x_j \in (s - e, s + e)$$

We'll say that s $= x_1 + x_2 + \ldots$.

Sums of Probabilities

- One can prove (assuming classical mathematics), that if $x_1, x_2, \ldots$ are the probabilities of disjoint propositions $p_1, p_2, \ldots$, then $x_1 + x_2 + \ldots$ exists.
- This is a consequence of a more general result that if $\forall i . x_i \geq 0$ and $\exists s \forall n \sum_{k=1}^{n} x_k < s$, then $x_1 + x_2 + \ldots$ exists.
- So it might make sense to ask whether $Pr(p_1 \vee p_2 \vee \ldots) = Pr(p_1) + Pr(p_2) + \ldots$

Some Interesting Limiting Sums

- It's useful to remember a few examples of limits in order to work with various examples.
- The most crucial result for our purposes is

$$\sum_{k=1}^{\infty} r^{-k} = \frac{1}{r-1}$$

So, for instance, $\frac{1}{3} + \frac{1}{9} + \ldots = \frac{1}{2}$

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 $\frac{1}{2}$

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

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

That would be ruled out by the equality axiom.

The Flat Distribution Over $\mathbb{N}$

- The flat distribution says that Pr(a = n) = 0 for all n
- But it says that Pr(a = 1 ∨ a = 2 ∨ ...) = 1
- Indeed, any distribution that says that there are a countable infinity of choices, and each is equally probable, will have this feature
- Some people (including me!) take that as a proof that there can't be a probability distribution making each member of a countable infinity equally probable
- Other people take it as a proof that $Pr(p1 \vee p2 \vee ...) = Pr(p1) + Pr(p2) + ...$ if (p1, p2, ... are pairwise disjoint) is not an axiom

New Axiom

- But we will take that as an axiom
- So we now take the domain of a probability function to be closed under countable union
- And we replace the (special) addition axiom with

$$Pr(p1 \vee p2 \vee ... | C) = Pr(p1|C) + Pr(p2|C) + ...$$

(for any p1, p2, ... that are pairwise disjoint)

Terminology

- This is called the axiom of countable additivity
- A countably additive probability function is one that satisfies it
- Ideally, we'd use the phrase 'finitely additive' probability function for any function that satisfies the finite addition axiom
- In practice, many people use that phrase only for functions that are *not* countably additive, but which do satisfy the finite additivity axiom

Same Size Sets

- I should say a little more about what I mean by a 'countable' set, and here is as good a time as any to do it
- There is an interesting question about what it is to say sets are the same size
- The working definition that most mathematicians use is Cantor's
- Two sets are the same size iff there is a one-one mapping from one to the other

Cantor's Definition

We'll write $|S_1|$ for the size of S_1. Then Cantor's definition is

$$|S_1| = |S_2| \leftrightarrow$$

$$\exists f \subseteq S_1 \times S_2 :$$

$$(\forall x \in S_1 \exists ! y \in S_2 :< x, y > \in f) \wedge$$

$$(\forall y \in S_2 \exists ! x \in S_1 :< x, y > \in f)$$

- That is, there's a set of ordered pairs, the first of which is in S_1, the second of which is in S_2, and each member of S_1 is the first member in exactly one of the ordered pairs, and each member of S_2 is the second member of exactly one.

Nice Consequences

- This gives us just the results we want for finite cases
- Any two sets with n members are the same size, for any finite n
- And it does so without assuming there are such things as natural numbers
- So in principle it could be used to measure the size of infinite sets

Odd Consequences

- When we do so, we get some odd results
- Let S_1 be the set of natural numbers, and S_2 the set of even numbers
- Then there's an easy mapping from one to the other, the mapping $< 1, 2 >, < 2, 4 >, \ldots$.
- So it turns out these sets are the same size
- So a set can be the same size as one of its proper subsets

Cantor's Definition

- For reasons too numerous to go into here, mathematical orthodoxy is that we should accept this odd result
- And philosophers have (I think correctly) followed them
- There is a useful generalisation of Cantor's generalisation that we won't prove, but is useful to have
- If there is a mapping from S_1 to a subset of S_2, and from S_2 to a subset of S_1, then there is a mapping from S_1 itself to S_2.
- We'll often use that to prove sets are the same size

Powersets

- When you first see the odd result, you might be tempted to conclude that all infinite sets are the same size
- That's not a consequence of Cantor's definition
- In fact we can prove a rather important result inconsistent with it
- The powerset of S is the set of all subsets of S (including the null set and S itself)
- We'll write this $\mathcal{P}(\mathrm{S})$
- We can prove $|\mathcal{P}(\mathrm{S})| \neq |\mathrm{S}|$

Powersets

- Assume $|\mathcal{P}(\mathrm{S})| = |\mathrm{S}|$
- Then there is a mapping f from S to $\mathcal{P}(\mathrm{S})$
- Let $\mathrm{R} = x : x \in S \wedge x \notin f(x)$
- Since by definition $\mathrm{R} \subseteq \mathrm{S}$, $\mathrm{R} \in \mathcal{P}(\mathrm{S})$
- So there must be some $\mathrm{y} \in \mathrm{S}$ that is mapped onto R, i.e. $\mathrm{f(y)} = \mathrm{R}$
- If $\mathrm{y} \in \mathrm{R}$, then $\mathrm{y} \notin \mathrm{f(y)}$, so $\mathrm{y} \notin \mathrm{R}$
- But if $\mathrm{y} \notin \mathrm{R}$, then $\mathrm{y} \in \mathrm{S}$ and $\mathrm{y} \notin \mathrm{f(y)}$, so $\mathrm{y} \in \mathrm{R}$
- Contradiction

Countable and Uncountable Sets

- As always, let $\mathbb{N}$ be the set of natural numbers.
- One consequence of the above result is that some sets, for example $\mathcal{P}(\mathbb{N})$ are not the same size as $\mathbb{N}$.
- We'll say that sets that are the same size as $\mathbb{N}$ are **countable**.
- We won't prove it here, but it is provable that $\mathbb{R}$ is the same size as $\mathcal{P}(\mathbb{N})$, a fact we'll use a bit in what follows

Set of all Sets

- One consequence of $|\mathcal{P}(S)| > |S|$ is that there is no set of all sets
- To see this, assume S is the set of all sets
- Then $\mathcal{P}(S) \subseteq S$, since every member of $\mathcal{P}(S)$ is a set
- So $|\mathcal{P}(S)| \not> |S|$, contrary to what we proved above

Consequences for Probability

- Recall that when we were setting things up formally, probabilities were defined over sets of sets, the latter of which we sometimes took to be propositions
- The set that the probability is defined over can't be the set of all sets, because there is no such set
- As we'll go along, we'll see more and more reasons for thinking that there are serious limits to which propositions can have probabilities
- This matters philosophically I think; there are a lot of applications where people assume that every proposition has a probability, and this assumption is typically false.

Finite Conglomerability

- We say that a probability function is *finitely conglomerable* iff there is no proposition H and partition $E_1, E_2, ..., E_n$ such that either for all i, Pr(H|E$_i$) $>$ Pr(H), or for all i, Pr(H|E$_i$) $<$ Pr(H).
- By 'partition' here we mean a set of propositions that are mutually exclusive and jointly exhaustive
- Given the axioms (and classical logic, which we're assuming basically always from now on) we can prove that all probability functions are finitely conglomerable.

Proof of Finite Conglomerability

Assume that $\forall i.Pr(H|E_i) > Pr(H)$. (The other case works the same way.)

$$\begin{aligned}
Pr(H) &= Pr(HE_1 \vee HE_2 \vee ... \vee HE_n) \\
&= Pr(HE_1) + Pr(HE_2) + ... + Pr(HE_n) \\
&= Pr(H|E_1)Pr(E_1) + Pr(H|E_2)Pr(E_2) + ... + Pr(H|E_n)Pr(E_n) \\
&> Pr(H)Pr(E_1) + Pr(H)Pr(E_2) + ... + Pr(H)Pr(E_n) \\
&= Pr(H)(Pr(E_1) + Pr(E_2) + ... + Pr(E_n)) \\
&= Pr(H)Pr(E_1 \vee E_2 \vee ... \vee E_n) \\
&= Pr(H)
\end{aligned}$$

Finite Congolomerability

- Conglomerability has a nice translation into regular talk
- It says that the probability of H can't be outside the realm of the probability of H given possible evidence E_i
- So it is a nice result to have, because that seems like it should be the case

Exclusive and Inclusive Disjunction

- You might be reminded by conglomerability of principles like or-elimination in logic
- If we have $E_i \rightarrow H$ for each i in a partition $E_1, E_2, ..., E_n$, then we have H
- Similarly, if $Pr(H|E_i) > x$ for each i, then $Pr(H) > x$
- But these principles are not the same
- The logical rule holds for either inclusive or exclusive disjunction
- The probabilistic rule only holds for exclusive disjunction

Not a Counterexample

- Say that you're going to draw two cards at random from a deck containing the ace of spades, the ace of clubs and the king of hearts
- Let p = You draw the ace of spaces
- Let q = You draw the ace of clubs
- Let r = You draw both aces
- Then $\Pr(r) = \frac{1}{3}$
- But $p \vee q$ is guaranteed to be true
- And $\Pr(r|p) = \Pr(r|q) = \frac{1}{2}$

Countable Conglomerability

- We might wonder about generalising the conglomerability principle to larger partitions
- For instance, consider the case where the partition is a countable set $E_1, E_2, \ldots$
- Countable conglomerability is the principle that Pr(H) is not always above or always below $\Pr(H|E_i)$
- Given countable additivity, we can prove countable conglomerability using essentially the earlier proof

Proof of Countable Congolomerability

Assume that $\forall i. Pr(H|E_i) > Pr(H)$. (The other case works the same way.)

$$\begin{aligned}
Pr(H) &= Pr(HE_1 \vee HE_2 \vee \ldots) \\
&= Pr(HE_1) + Pr(HE_2) + \ldots \\
&= Pr(H|E_1)Pr(E_1) + Pr(H|E_2)Pr(E_2) + \ldots \\
&> Pr(H)Pr(E_1) + Pr(H)Pr(E_2) + \ldots \\
&= Pr(H)(Pr(E_1) + Pr(E_2) + \ldots) \\
&= Pr(H)Pr(E_1 \vee E_2 \vee \ldots) \\
&= Pr(H)
\end{aligned}$$

The Flat Distribution Again

- That proof used countable additivity twice over
- We might wonder whether it is necessary
- The answer is that it is
- Indeed the flat distribution we've seen provides a nice counterexample to countable conglomerability

Violating Conglomerability

- Define F to mean *is a multiple of 10*
- Consider each of the sets $\{n, 10n, 100n, \ldots\}$ where n is not a multiple of 10
- Those sets form a partition of $\mathbb{N}$
- For any such set S, $\Pr(Fa|a \in S) = 1$
- But $\Pr(Fa) = \frac{1}{10}$

A More Dramatic Violation

- Assume i and j are numbers randomly chosen from $\mathbb{N}$
- Consider $\Pr(i > j)$
- It's easy enough to show that $\forall n.\ \Pr(i > j \mid i = n) = 0$
- And $\forall n.\ \Pr(i > j \mid j = n) = 1$
- Whatever value $\Pr(i > j)$ takes, we'll have a massive conglomerability failure

Generalising This Result

- So we've shown that if we assume countable additivity, we can prove countable conglomerability
- And without countable additivity, this proof doesn't go through
- I was hoping to find a proof of something stronger
- Namely that if Pr is not countably additive and its domain is closed under countable union, then Pr is not countably conglomerable
- But I couldn't find a proof of that, or a citation of it, or a counterexample
- I'll keep looking and hopefully find something for next week

Conglomerability in Decision Theory

- A decision-based version of finite conglomerability is central to orthodox decision theory
- First some terminology
- Let $A \succ B$ mean that A is preferred to B
- Let $A \succeq B$ mean that A is at least as preferred as B
- Intuitively, conditional preferences are just preferences over conjunctions. So if A is preferred to B given C, that just means $AC \succ BC$

Conglomerability in Decision Theory

Assume throughout that $E_1, ..., E_n$ is a partition of possibility space.

Conglomerability for Preference $(\forall i . E_i A \succ E_i B) \rightarrow A \succ B$

That is, if A is better than B conditional on any member of the partition, A is better than B

Conglomerability in Decision Theory

Assume throughout that $E_1, \ldots, E_n$ is a partition of possibility space.

Conglomerability for Preference $(\forall i . E_i A \succ E_i B) \rightarrow A \succ B$

That is, if A is better than B conditional on any member of the partition, A is better than B

- Again, note the importance of this being a partition
- This really isn't just or-elimination
- On the other hand, it's a pretty plausible principle

Utility Functions

- To see how this generalises to the infinite case, we need a little decision theory
- Decision theorists attribute to each agent a *utility function* u. (We're skipping for now the reasons they do this.)
- If A $\succ$ B, then u(A) $>$ u(B)
- More importantly, u measures strength of preference
- Roughly, if A is preferred over B by as much as B is preferred over C, then u(A) - u(B) $=$ u(B) - u(C)

Expected Value

Here is a general definition of the expected value of a random variable X. We'll restrict our attention to the case where X takes on at most finitely many values, because that's all that's needed for now.

$$Exp(X) = \sum_{k} k \cdot Pr(X = k)$$

A random variable is just any expression that takes on different possible values. So we can talk about the expected height of the next person that walks into the room, or the expected age of the next U.S. President, or the expected utility of performing a particular action.

- Note that this will frequently *not* be a possible value of X

Expected Value and Action

The core of modern decision theory is the idea that action A is preferred to action B iff the expected utility of doing A is higher than the expected utility of doing B.

- Here's a quick example of this
- If p is true, then $u(A) = 10$, and $u(B) = 20$
- If p is false, then $u(A) = 5$ and $u(B) = 1$
- And $Pr(p) = 0.3$
- So $Exp(u(A)) = 10 \cdot 0.3 + 5 \cdot 0.7 = 6.5$
- And $Exp(u(B)) = 20 \cdot 0.3 + 1 \cdot 0.7 = 6.7$
- So B is better to do, although probably A will have the better outcome

The Two-Envelope Paradox

- If we assume that u is unbounded above, i.e. that utilities can be arbitrarily high, we get a rather odd paradox
- Formally, we can state the puzzle using two random variables, X and Y
- For each $n \in 0, 1, 2, \ldots$, $\Pr(X = 2^n) = \frac{2}{5} \times \frac{3^n}{5^n}$
- I'll leave it as an exercise to prove that those probabilities sum to 1
- $\Pr(Y=1) = \Pr(Y=2) = \frac{1}{2}$
- $u(A) = XY$, and $u(B) = X(3-Y)$

The Two-Envelope Paradox

- We can give a more intuitive characterisation of what's going on
- Take a biased coin, with a $\frac{3}{5}$ chance of landing heads, two unmarked envelopes, and put something worth 1 'util' in the first envelope
- Flip the coin, and if it lands tails, skip to the next step, while if it lands heads, double the value (in utils) in the first envelope and repeat
- Once the coin lands tails once, put double the amount in the second envelope
- Now shuffle the envelopes, and offer one of them to a friend

The Two-Envelope Paradox

- Intuitively, it should not matter which envelope the friend receives
- Let A be the action of taking the left-most envelope, and B taking the right-most envelope
- A little observation shows that the informal story we've told is well modelled by the formal story two slides back
- Now we'll try to argue that B is preferable to A.

The Two-Envelope Paradox

- We'll write things like A=4 meaning that envelope A contains something worth 4 utils
- Conditional on A=1, we know B=2, so B is preferable to A
- If A=2^n, for n $>$ 0, then there are two ways this could come about
- Either $X = 2^n$ and $Y = 1$, so $B = X = 2^{n+1}$, or $X = 2^{n-1}$ and $Y = 2$, so $B = 2^{n-1}$
- The prior probability of the first of these is $\frac{3^n}{5^{n+1}}$
- And the prior probability of the second is $\frac{3^{n-1}}{5^n}$

The Two-Envelope Paradox

- The prior probability of A $= 2^n$ and B $= 2^{n+1}$ is $\frac{3^n}{5^{n+1}}$
- And the prior probability of A $= 2^n$ and B $= 2^{n-1}$ is $\frac{3^{n-1}}{5^n}$
- Those are the only ways that A $= 2^n$ could come about
- So $\Pr(\mathrm{A} = 2^n) = \frac{3^n}{5^{n+1}} + \frac{3^{n-1}}{5^n} = \frac{8 \times 3^{n-1}}{5^{n+1}}$
- So $\Pr(B = 2^{n+1} | A = 2^n) = \frac{\frac{3^n}{5^{n+1}}}{\frac{8 \times 3^{n-1}}{5^{n+1}}} = \frac{3}{8}$
- And from that it quickly follows that $\Pr(B = 2^{n-1} | A = 2^n) = \frac{5}{8}$

The Two-Envelope Paradox

- So conditional on envelope A containing 2^n, there is a $\frac{3}{8}$ chance that B contains twice as much, and $\frac{5}{8}$ chance that it contains $1/2$ as much.
- So conditional on envelope A containing 2^n, the expected value of the contents of B is
- $\frac{3}{8} \times 2 \times 2^n + \frac{5}{8} \times \frac{1}{2} \times 2^n$
- And a little calculation shows that is equal to $\frac{17}{16} \times 2^n$

The Two-Envelope Paradox

- In preference terms that we put before, we have the following true for all values of n
- (Take envelope B)$\wedge A = 2^n \succ$(Take envelope A)$\wedge A = 2^n$
- The conglomerability principle we were considering would say that entails that (Take envelope B) $\succ$ (Take envelope A)

The Two-Envelope Paradox

- In preference terms that we put before, we have the following true for all values of n
- (Take envelope B)$\wedge A = 2^n \succ$(Take envelope A)$\wedge A = 2^n$
- The conglomerability principle we were considering would say that entails that (Take envelope B) $\succ$ (Take envelope A)
- Not so fast!
- The exact same calculations show that for all n
- (Take envelope A) $\wedge B = 2^n \succ$(Take envelope B)$\wedge B = 2^n$
- And conglomerability now says that (Take envelope A) $\succ$ (Take envelope B)

The Two-Envelope Paradox

- We can put this all more dramatically
- You know before looking that if you look in one of the envelopes, either of them, you'll prefer to have the other
- Indeed you'll *pay* to have the other
- And that's true whichever envelope you look in
- Some people take this to show that the setup is a money pump
- We'll say more about money pump arguments in subsequent lectures

A Puzzle About Conglomerability for Decisions

It seems we have to give up one of the following principles

1. $\succ$ is anti-symmetric
2. u is unbounded above
3. Countable conglomerability for decisions

And if we give up 3, we have to give up one of

1. Countable conglomerablity for credences
2. That credences are governed by the same principles as decisions

Bad Company Objection?

If we take option 3, we have two somewhat distinct reasons for giving up countable conglomerability for credences (i.e. subjective probabilities)

1. As we'll see in upcoming weeks, some theorists think that all constraints on credences follow from constraints on rational decisions. If that's right, and we have to give up countable conglomerability for decisions, there is a direct argument against countable conglomerability for credences
2. We might think that the intuitive support for countable conglomerability for credences is undermined by the inconsistency (relative to some assumptions) of countable conglomerability for decisions. Think of this as a 'bad company' argument.

We have to do a lot more work before we can evaluate these arguments. Some of that work has to do with 'flat' distributions.

Logical Probability

Traditionally people thought of probability as a kind of 'logical' relation

- Consider again the model I mentioned of probabilities as measures on truth tables
- Given the way I set things up, there is one very natural measure you might be pushed to
- That's the measure that assigns equal weight to each row
- Something like this idea is behind the traditional 'logical' interpretation of probability

Logical Probability

This might seem like a bad idea, because probabilities will now be massively model dependent

- It would be nice to have something more precise to say at this point about what was driving the old idea
- And by 'old' here, I mean a view that was looking creaky when Keynes wrote in the 1900s
- To the extent that the view has modern adherents, they tend to have complicated views about what is the right model

Flat Distributions

There is a modern descendent of this old view though

- That's the view that when you are completely without evidence, you should distribute credences/probabilities equally over open possibilities
- If there are countably many open possibilities, that leads to violations of countable additivity
- In other cases it leads to worse results

United Kingdom Puzzle

- Jack is told that Smith is from either the United Kingdom or the Republic of Ireland
- So he assigns probability $\frac{1}{2}$ to Smith's being from the United Kingdom
- Jill is told that Smith is from either England, Scotland, Wales or Ireland
- So she assigns probability $\frac{3}{4}$ to Smith's being from England, Scotland or Wales
- Oddly, despite being given in some sense the same information, Jill ended up assigning a higher probability to an (in some sense) weaker proposition
- Perhaps that isn't too odd if Jack and Jill don't know the structure of the United Kingdom
- But odder results are to follow

Breakfast Puzzle

What's the probability that I had Vegemite on toast for breakfast?

- You might think that, if you know that's a possibility but don't know anything else, that this probability should be $\frac{1}{2}$

Breakfast Puzzle

What's the probability that I had Vegemite on toast for breakfast?

- You might think that, if you know that's a possibility but don't know anything else, that this probability should be $\frac{1}{2}$

What's the probability that I had toast for breakfast?

- You might think that, if you know that's a possibility but don't know anything else, that this probability should be $\frac{1}{2}$

Breakfast Puzzle

What's the probability that I had Vegemite on toast for breakfast?

- You might think that, if you know that's a possibility but don't know anything else, that this probability should be $\frac{1}{2}$

What's the probability that I had toast for breakfast?

- You might think that, if you know that's a possibility but don't know anything else, that this probability should be $\frac{1}{2}$

But given that I definitely had toast if I had Vegemite on toast, it follows that the probability of my having Vegemite on toast, conditional on my having toast, is 1. And that shouldn't follow from minimal information.

Cube Factory

- Jack is told that a factory makes cubes, and that every cube has a side length between 0 and 2cm
- So he assigns probability $\frac{1}{2}$ to the proposition that the next cube has a side length between 0 and 1cm.
- Jill is told that a factory makes cubes, and that every cube has a volume between 0 and 8cm^3
- So she assigns probability $\frac{1}{8}$ to the proposition that the next cube has a volume between 0 and 1cm^3.

Cube Factory

- Jack is told that a factory makes cubes, and that every cube has a side length between 0 and 2cm
- So he assigns probability $\frac{1}{2}$ to the proposition that the next cube has a side length between 0 and 1cm.
- Jill is told that a factory makes cubes, and that every cube has a volume between 0 and 8cm^3
- So she assigns probability $\frac{1}{8}$ to the proposition that the next cube has a volume between 0 and 1cm^3.
- This is odd
- The information they got was provably equivalent
- And they end up assigning different probabilities to provably equivalent propositions

Chord Puzzle

Puzzle: Given a circle of unit radius, what is the probability that a chord randomly chosen on it has length >1?

Chord Puzzle

Puzzle: Given a circle of unit radius, what is the probability that a chord randomly chosen on it has length >1?

Arguably there are four different answers to this puzzle

1. $\frac{1}{2}$
2. $\frac{2}{3}$
3. $\frac{\sqrt{3}}{2}$
4. $\frac{3}{4}$

Chord Puzzle: Answer $\frac{1}{2}$

- Chord lengths are between 0 and 2
- So the probability that a given length is >1 is $\frac{1}{2}$

Chord Puzzle: Answer $\frac{2}{3}$

- Picking a chord is equivalent to picking two points
- When all we care about is the chord length, the first point is arbitrary
- So it's equivalent to given a point, picking another point
- If the second point is within $60^\circ$of the first, in either direction, the chord length will be < 1
- So the probability that it is > 1 is $\frac{1}{3}$

Chord Puzzle: Answer $\frac{\sqrt{3}}{2}$

- For each point in the circle, there is exactly one chord it is the midpoint of
- So if we select a radius (which will bisect the chord), then select a point on it, we'll have picked a unique chord
- Again, if all we care about is chord length, the choice of radius is irrelevant
- On any given radius, if we pick a point less than $\frac{\sqrt{3}}{2}$ of the way out from the centre, we'll have picked out a chord with length > 1. (Exercise: prove this)
- So the probability that the chord length is > 1 is $\frac{\sqrt{3}}{2}$

Chord Puzzle: Answer $\frac{3}{4}$

- Perhaps going via a two-step process is unnatural
- Better to just measure the area of points that are midpoints of chords with length > 1
- The result from the previous slide shows that is $\frac{3}{4}$ of the area of the original circle
- So the probability that our chord has length > 1 is $\frac{3}{4}$

Chord Puzzle

The point of these puzzles is not to suggest that one of these is the correct answer.

Rather, the point is that none of them are.

Philosophical Consequences

Probability is not a measure of ignorance.

- We can't just say that the correct probability measure assigns equal probability to all options we haven't ruled out
- We can't do that because there are too many incompatible ways to do it

Philosophical Consequences

Probability is not a measure of ignorance.

- We can't just say that the correct probability measure assigns equal probability to all options we haven't ruled out
- We can't do that because there are too many incompatible ways to do it

There's a philosophical argument for the same conclusion. (The following is not entirely flippant!)

1. Probability is a guide to life

Philosophical Consequences

Probability is not a measure of ignorance.

- We can't just say that the correct probability measure assigns equal probability to all options we haven't ruled out
- We can't do that because there are too many incompatible ways to do it

There's a philosophical argument for the same conclusion. (The following is not entirely flippant!)

1. Probability is a guide to life
2. Ignorance is not a guide to life

Philosophical Consequences

Probability is not a measure of ignorance.

- We can't just say that the correct probability measure assigns equal probability to all options we haven't ruled out
- We can't do that because there are too many incompatible ways to do it

There's a philosophical argument for the same conclusion. (The following is not entirely flippant!)

1. Probability is a guide to life
2. Ignorance is not a guide to life
3. So probability is not a measure of ignorance

Rotational Invariance

Let's try to define a measure on subsets of the unit 'circle' [0, 1) that is rotationally invariant.

- The unit circle is a picturesque way of thinking about the interval [0, 1)
- So think about the points arranged on a clockface, with 0 at the top, $\frac{1}{4}$ at 3 o'clock, etc
- The idea then is to find a measure such that if you can get from S to S' by rotating S around the circle, then S and S' should have the same measure
- The measure we'll end up with is called the Lesbegue measure

Formalisms

First, define what a rotation is. The binary function $\oplus$ is defined in [0, 1) $\times$ [0,1) as follows

$$x \oplus y = \begin{cases} x + y & \text{if } x + y < 1 \\ x + y - 1 & \text{if } x + y \geq 1 \end{cases}$$

Then the definition of rotational invariance is

$$\text{If } m(S) = x \text{ and } y \in [0, 1) \text{ then } m(\{x \oplus y : x \in S\}) = x$$

Formalisms

What we're interested in then is constructing as large a (normalised) measure as possible that is rotationally invariant

- Remember that a measure is a countable additive function over subsets of some universe
- So if A and B are disjoint sets, $m(A \cup B) = m(A) + m(B)$
- And more generally, if $A_1, \ldots$ are disjoint sets, then $m(A_1 \cup \ldots) = m(A_1) + \ldots$
- A normalised measure is one for which m(U) = 1
- Since our universe is [0, 1), we should restrict our attention to normalised measures

Goal

What we're going to show is that there is no rotationally invariant normalised measure definable over all members of $\mathcal{P}[0, 1)$.

In probabilistic terms, you can't even define a probability function over every subset of [0, 1) if you want to insist on rotational invariance.

We already saw some set-theoretic reasons why probability functions can't be complete. There are also reasons 'internal' to probability to think this.

Measures and Lengths

This measure has some nice properties

- If $S = [x, y)$, then $m(S) = y - x$
- The formal proof of this is a little long, and we'll basically skip it
- But note that there's a quick proof that if $y - x = \frac{1}{n}$ for $n \in \mathbb{N}$, $S = [x, y)$, then $m(S) = y - x$ follows from rotational invariance
- So if $y - x = \frac{m}{n}$ for $m, n \in \mathbb{N}$, the result follows from additivity
- And the complete result follows by countable additivity

ZF and ZFC

- When modern set theory was being developed in light of Russell's paradox, a number of axioms were fairly widely accepted
- These included powerset and infinity
- But one new axiom was controversial, the axiom of choice
- So much so that in the contemporary naming we distinguish ZF (the Zermelo-Frankel axioms) from ZFC (Zermelo-Frankel-Choice)

Axiom of Choice

There are a few different ways to formulate the axiom of choice. Here's the one we will use.

- Let S be any set of disjoint sets
- Then there exists a choice set C such that $\forall s \in S . \exists! x \in C . x \in s$
- As Russell put it, "The Axiom of Choice is necessary to select a set from an infinite number of socks, but not an infinite number of shoes"
- The axiom of choice has a *lot* of power when it comes to proving theorems.

Partition of [0, 1) by Rational Numbers

- Consider each set of numbers of the form $\{x : |x - s| \in \mathbb{Q}\}$ for $s \in [0, 1)$
- $\mathbb{Q}$ is the set of rational numbers
- So the sets above are sets of numbers that are separated from each other by a rational number
- Actually, they are numbers that are separated from some 'seed' s by a rational number, but since being separated by a rational number is an equivalence relation, what I wrote will do just as well

Partition of [0, 1) by Rational Numbers

- Let S be the set of every such set $\{x : |x - s| \in \mathbb{Q}\}$ for $s \in [0, 1)$
- Since $s \in \{x : |x - s| \in \mathbb{Q}\}$ for $s \in [0, 1)$, every $s \in [0, 1)$ is in one of these sets
- Moreover, since being separated by a rational number is an equivalence relation, every number is in exactly one of them
- So S is a partition of [0, 1), a set of sets such that every number in [0, 1) is in exactly one of its members

The Choice Set and Its Rotations

- Now choice tells us we can generate a 'choice set' from S
- Call this set C_0
- For every $x \in \mathbb{Q} \cap [0, 1)$ let $C_x = \{y \oplus x : y \in C_0\}$
- Let C be the set of all of these C_x
- We now want to prove that C is a partition of [0, 1)

The Choice Set and Its Rotations

- Assume $a \in C_x \wedge a \in C_y$ where $x \neq y$
- Then either $a - x$ or $a - x + 1$ is in C_0
- And either $a - y$ or $a - y + 1$ is in C_0
- So two of $a - x$, $a - x + 1$, $a - y$, $a - y + 1$ is in C_0
- But those numbers differ from each other by a rational number
- And that contradicts the assumption that C_0 contains exactly one member of each set in S

The Choice Set and Its Rotations

- For any $a \in [0, 1)$, there is some set $s \in S$ it is in
- And there is some number $x \in s$ in C_0
- So if $a \geq x$, then $a \in C_{a-x}$, and if $a < x$ then $a \in C_{a-x+1}$
- Either way, it is in some set in C

The Choice Set and Its Rotations

- So C is a partition
- Since there are countably many rational numbers in $[0, 1)$, it is a countable partition
- So if the sets in it have a measure, the measure of those sets must sum to 1
- But since the sets in it are constructed by rotation, they must have the same measure
- This is impossible
- If this measure is 0, then the sum of the measures of the set of C is 0
- If this measure is > 0, then the sum of the measures of the set of C is ∞

Conclusions about Unmeasurability

So C_0 simply does not have a Lesbegue measure

- Indeed, quite a lot of sets do not
- This follows from a simple assumption that the measure is countably additive, and rotationally invariant

Required Set Theory

- We used the Axiom of Choice here
- It was necessary to go beyond ZF; ZF is consistent with all sets being measurable
- But the result, that some sets are not measurable is not *equivalent* to Choice
- Indeed, it is a lot weaker
- But going into how much weaker would (a) take us a long way afield, and (b) go well beyond what I'm competent in

Philosophical Consequences

Not all propositions have probabilities

- As we noted above, this follows from the fact that there is no set of all sets
- But even if you ignored that complication, there is a purely probabilistic reason to think that not all propositions should have probabilities
- And again, this matters to philosophical applications that presuppose all propositions have probabilities