Brendan Fleig-Goldstein
Confirmation Seminar
Presentation on:

Ockham's Razors: A User's Manual by Elliot Sober: Jeffreys's Simplicity Postulate and subsequent discussion pp. 87-102

Preliminaries

Razor of silence vs razor of denial: roughly agnosticism vs atheism. Razor of silence says e.g., "we don't need to postulate X" but doesn't deny X.

First priors vs non-first priors: first priors don't include empirical assumptions and can't be justified by observation

How should we assign first priors?

Jeffreys's simplicity postulate:

Jeffreys is concerned with scientific inference and how to learn from experience. He is concerned with all of the possible scientific hypotheses that could be true of the world and how to assign first priors over these.

In this context, there are an infinite amount of pair-wise incompatible hypotheses. The principle of indifference runs into trouble in this context (assuming axioms of probability, which he does).

Instead, assign higher prior probabilities to simpler hypotheses.

What counts as simpler? The number of adjustable parameters is inversely proportional to its simplicity (but Jeffreys had other requirements re what counts as simple, which he refined over time).

For example, in the case of the infinite ascending hierarchy of polynomial models, you can start from the bottom and assign ½ + ¼ + ⅛… all the way up (or a different infinite series summing to 1).

Jeffreys's simplicity postulate is global as opposed to local: simplicity leads to a single prior probability for a hypothesis; it is not relative to a particular inferential context.

Certain philosophers such as Leibniz, Newton*, and Kant seem to think there is an upper bound on how complex the world can be. Jeffreys just thinks that simpler hypotheses are *more probable*.

*I don't agree

Some minor problems:

Jeffreys's framework more or less requires that all scientific hypotheses are enumerable, and this is not the case…

Jeffreys's simplicity metric does not seem to comport with our intuitions about simpler hypotheses (Ackermann, 1963).

Big problem:

Jeffreys does not give an argument for assigning higher probabilities to simpler hypotheses. It seems he thinks that the simplicity postulate is a necessary condition for the possibility of learning from experience, and that we all assume that we can learn from experience.

Popper and simplicity

Popper primer: scientific theories aren't "confirmed" when their predictions agree with observations, they are "corroborated" when we try hard to falsify them and fail. Scientists should boldly conjecture strong and interesting theses and subject them to strong tests in an attempt to prove them wrong deductively.

For Popper, simpler theories are:

1. Stronger and more interesting!
2. Less probable
3. More easily falsifiable

Less probable: simpler hypothesis is often a special case of a more complicated set of postulates, and therefore *can't* be more probable

Line: $a_0 + a_1x$
Parabola: $a_0 + a_1x + a_2x^2$
Line (as a special case of a parabola): $a_0 + a_1x + a_2x^2$ & $a_2 = 0$

Popper correctly points out that a conjunct can't be more likely than the conjunction! Assigning a higher probability to a line violates axioms of probability.

More easily falsifiable: logically stronger → more entailments → more chances for a theory to be proven wrong

Stronger and more interesting?

Newton's theory of universal gravitation: applies to all matter everywhere

1. Simple
2. Strong and interesting
3. Easily falsifiable
4. Less probable?

In ascending hierarchy of polynomial models, lower polynomial degrees are:

1. Simpler: paradigm case?

2. More falsifiable: requires less data points to falsify (at least 3 points to falsify a line, 4 points to falsify a quadratic…)
3. Less probable: every lower polynomial degree is in a sense a special case of higher polynomial degree. A line is a quadratic where one of the adjustable parameters is 0…
4. Strong and interesting?

Ok, but what about:

H1: Polynomial degree 2
H2: Polynomial degree 2 & Jeffreys's middle name was Geoffrey

H1: simpler
H2: less probable, logically stronger, more falsifiable

"Popper thought all probabilistic theories are equally and infinitely complex" (p. 97)

Add an error term to a postulate and now you're in trouble…?

Popper had a measure of degree of corroboration that didn't behave in a way that matched his philosophy...and it was Bayesian.

Jeffreys's simplicity postulate reformulated to handle Popper's objection

Line: $a_0 + a_1x$
Parabola: $a_0 + a_1x + a_2x^2$
Line (as a special case of a parabola): $a_0 + a_1x + a_2x^2$ & $a_2 = 0$

Popper correctly points out that a conjunct can't be more likely than the conjunction! Assigning a higher probability to a line violates axioms of probability.

Try the following, prohibit nested models:

Line (as a special case of a parabola): $a_0 + a_1x + a_2x^2$ & $a_2 = 0$

Parabola (true parabola): $a_0 + a_1x + a_2x^2$ & $a_2 \neq 0$

Now, to say that the line is more probable is to say that $a_2 = 0$ is more probable than $a_2 \neq 0$. Sober says this is implausible, but that seems to beg the question. No violation of axioms of probability.

Parsimony and non-first priors

Consider a standard case of Bayesian reasoning:

E: cough and fever
H1: common cold (very common, usually leads to symptoms)
H2: post-alien abduction galactic space flu (very rare, always leads to symptoms)

Sometimes we use the word “simpler” to refer to H1, when all we mean is that it has a higher probability, because the prior is higher, justified by empirical data. I.e., parsimony can be surrogate for claims about non-first priors.

Sometimes this leads people to reason strangely…see the 20th century anaphylactic reaction to group selection. Confusing the razor of silence for the razor of denial?