Day 12 (B) Thurs Aug 8

Diagnostic Screening, Probabilistic instantiation, base rates and all that

The computations for the DS model stem from Berger and Sellke (1987).

They claim it's just a heuristic, not that you'd use prevalences to assign priors.

FALLACIOUS ARGUMENT:

Pr(the randomly selected null hypothesis is true) = .5

The randomly selected null hypothesis is H_{51}

Pr(H_{51} is true) = .5

Each null either is true or not! My selecting it from an urn by means of a chosen selection procedure does not give evidence for its truth or probable truth

Equivocal: I can model an experiment of selecting hypotheses from an urn: if it satisfies a Bernouilli model, I might say, the probability a (generic) outcome has the property (true) = the % true.

But the event (of being red, being true) isn't a statistical hypothesis; so isn't what you need for the likelihoods.

A statistical hypothesis H assigns probabilities to all possible outcomes

Consider this in relation to some criticisms of severity

A “hypothesis” that consists of asserting that a sample possesses a characteristic such as “having a disease” or “being college-ready.”

The point is to give it a frequentist prior.

Students From the Wrong Side of Town

Isaac, has passed comprehensive tests of mastery of high school subjects regarded as indicating college readiness S…

The battery of tests is assumed to be very capable of uncovering lack of readiness, so that such high scores *S* could very rarely result among high school students who are not sufficiently prepared to be deemed ‘college ready’.

Take S to be good evidence

H(I): Isaac is not deficient but is college ready.

And against

H'(I): Isaac's mastery of high school subjects is deficient, i.e., he is not college-ready.

Pr(*S*|*H(I)*: Isaac is college ready) ≈ 1, (practically 1)

Pr(*S*|*H*'(I): not college ready (i.e., deficient)) = .05 (very low)

- We should really consider degrees of readiness, but here I keep to the supposed counterexample.

Note: Thse numbers do not by themselves lead us to say H(I) has passed severely.

- Need to know of selection effects
- Not to check that they translate into a process that probes readiness, not an "isolated result"

Suppose a case where H(I) is warranted by dint of scores S.

"But wait a minute!" says the critic, Isaac was randomly selected from a population wherein college-readiness is exceedingly rare, Fewready Town where only 1 in 1000 are college ready. e.g.,

(*) P($H(I)$) =.001.

Thus the posterior probability for $H(I)$ is still low and H'(I)(deficient), the posterior is high.

e.g., Pr(H'(I)|S) = .95.

Fallacy of probabilistic instantiation

The critic – for example, Howson, Achinstein – sees the conclusion as problematic for the severity account as, it's assumed the frequentist would also accept (∗) P(*H(I)*) = .001.

Although the probability of college readiness in a randomly selected student from high schoolers from Fewready Town is .001, it doesn't follow that Isaac, the one we happened to select, has a probability of .001 of being college-ready

To suppose it does is to commit a kind of a fallacy of division:

The prevalence of readiness in Fewready Town is low
Isaac comes from Fewready Town

Thus, there’s a low probability that Isaac is ready

We need not preclude that H(I) has a legitimate frequentist prior; it might refer to generic and environmental factors that determine the chance of his deficiency

Achinstein’s “response to the probabilistic fallacy charge is to say that it would be true if the probabilities in question were construed as relative frequencies. [but] I am concerned with epistemic probability.”

Achinstein’s Rule for Objective Epistemic Probabilities: If (we know only that) Isaac is randomly selected from a population where p% have property C, then the objective epistemic probability that Isaac has C equals p.(2010, p. 187)

"If all we know is that Isaac was chosen at random from a very disadvantaged population, very few of whose members are college ready, say one out of one thousand,then we would be justified in believing that it is very unlikely that Isaac is college-ready"

(i.e., $\Pr(H(I)) = .001$ and, hence $\Pr(H(I)|S)$ is very low)

Even though $\Pr(H(I)|S)$ has increased from $P(H(I))$

For Achinstein, unless the posterior reaches a threshold of a fairly high number, he claims, the evidence is “lousy.”

The example considers only two outcomes: reaching the high scores or not, i.e., S or $\sim$S.

Clearly a lower grade gives even less evidence of readiness; that is, $\Pr(H(I)|{\sim}S) < \Pr(H(I)|S)$

Therefore, whether Isaac scored a high score or not, Achinstein’s epistemic probabilist reports justified high belief that Isaac is not ready.

The probability of Achinstein finding evidence of Isaac's readiness even if in fact he is ready (H is true) is zero.

Therefore, Achinstein's account violates what we have been calling the most minimal principle for evidence:

- The weak severity principle: Data x fail to provide good evidence for the truth of H' if the inferential procedure had very little chance of providing evidence against H', even if H' is false.

Reverse discrimination?

- If Isaac had been selected from a population where college-readiness is common, Manyready suburbs, the same set of passing scores *S* would be regarded as strong evidence for *H(I)*, Isaac being ready.
- Using this way of *evaluating evidence*, a high school student would have to have scored quite a bit higher on these tests than one selected from the affluent neighborhood in order for his scores to be considered evidence for his readiness!

I consulted with Lehmann after the first round of examples in 1996-7

- I was visiting him in Princeton where his wife J. Shaffer was at the Institute for Educational Testing Service, and this type of case could arise in policy-making
- He said: the test hasn't done its job if it can't make distinction in cases of rare diseases or rare assets

Actually if you actually had a large proportion of unready students amongst those who get passing scores S, there would be many reasons to deem the tests too lax for severity to be satisfied for Isaac.

So the prior enters, and is grounds to question those numbers

Severity for Problem-Solving (p. 300): Souvenir U

Note that there's no reason the problem at hand can't be providing an ordinary conditional probability

Severity then enters to assess if there is adequate warrant to take the problem as solved

The Case of General Hypotheses

When we move from hypotheses like “Isaac is college-ready” (which are really events) to generalizations – which Achinstein makes clear he regards as mandatory the difficulty for obtaining epistemic probabilities via his frequentist straight rule become more serious

The percentages “initially true” will vary considerably, and each would license a distinct “objective epistemic” prior.

The problems with the diagnostic screening.

Fisher: The Function of the p-value is Not Capable of Finding Expression…...

Discussing a test of the hypothesis that the stars are distributed at random, Fisher takes the low p-value (about 1 in 33,000) to "exclude at a high level of significance any theory involving random distribution" (Fisher (1956), p. 42).

Even if one were to imagine that H_0 had an extremely high prior probability, Fisher continues,--never minding "what such a statement of probability a priori could possibly mean "—the resulting high posteriori probability to H_0, would only show that "our reluctance to accept a hypothesis strongly contradicted by a test of significance"… (Fisher (1956), p. 45) "is not capable of finding expression in any calculation of probability a posteriori" (Fisher (1956), p. 43).

Indeed, if one were to consider the claim about the a priori probability to be itself a hypothesis, Fisher suggests, it would be rejected by the data.

So, if there were a case where H severely passes test T with x, and yet the posterior of H given x is low, we are free to take this as evidence that the posterior fails to do the job demanded by the severity requirement.

(1) The priors given in all the examples I've seen are not legitimate frequentist priors for the statistical hypothesis being tested;

(2) Using the procedures recommended by those who advocate using those priors, we would endorse inferences that fail on severity grounds.

(3) If there is a legitimate frequentist prior for the H under test, then, we would assign those posteriors in the same way we assign probability to events—values of random variables.

(4) If the posterior probability of "not-H" is high even though I regard H as having passed a severe test, it might well be seen as showing the inability of the posterior probability concept to capture the notion of evidence held by a severe tester! (R.A. Fisher's position).

I think many people who think they want a probability of a hypothesis really just want ordinary probabilities of events

The way to get them on the error statistical account is to obtain good evidence for the statistical hypothesis (or model) that assigns these probabilities!

Take a common linear regression model M:

$$Y = a + bX_1 + cX + u$$

This might let you predict the expected value of Y given a value for X, e.g., salary, given numbers of years of training, sex, etc. given the model has passed severely.