Having conceded loss in the battle for justifying induction, philosophers appeal to logic to capture scientific method

Inductive Logics	Logic of falsification
“Confirmation Theory” Rules to assign degrees of probability or confirmation to hypotheses given evidence e	Methodological falsification Rules to decide when to “prefer” or accept hypotheses
Carnap C(H,e) Inductive Logicians	Popper Deductive Testers

Inductive Logicians	Deductive Testers
we can build and try to justify "inductive logics" straight rule: Assign degrees of confirmation/credibility	we can reject induction and uphold the "rationality" of preferring or accepting H if it is "well tested"
Statistical affinity	Statistical affinity
Bayesian (and likelihoodist) accounts	Fisherian, Neyman-Pearson methods: probability enters to ensure reliability and severity of tests with these tests.

Brief Popper Notes

Demarcation Criterion:

Popper's problem is how does one determine (demarcate) a scientific theory from a non- or pseudo-scientific theory? I

(Popper) reject the common answer, namely….

We need to distinguish between empirical methods that are scientific from those that are not

The theories that got me started involved: Marx's theory of history, Freud and Adler's theories of psychology, and Einstein's theory of General Relativity (GTR).

My friends were impressed with these theories' "explanatory power": once your eyes were opened, they seemed to explain everything

I worked for Adler but was disappointed in his handling a case that seemed to go against Adler's theory.

The upshot is that all Adler's *observations were interpreted in light of his theory.*

In fact, one can use the theory to explain human actions that are exactly opposite, e.g.,

(a) a man drowns a child,

(b) a man sacrifices his life to save a drowning child.

Both may be explained using the resources of either theory (e.g., inferiority complex or Oedipal complex.)

<u>Freud</u>:

In case (a), Freud might say the man suffered from repression, while in (b) man had achieved sublimation.

There was no behavior that couldn’t be interpreted in terms of either theory.

In contrast, Einstein’s theory was impressive because it had a risky prediction: if the predicted light deflection effect was observed to be absent, the theory would be refuted.

Popper concluded (36):

1. It is easy to obtain confirmations, if we look for them.

2. Confirmations should only count if they are the result of risky predictions, if without the theory, we should have expected an event incompatible with the theory.

3. Every good theory is a prohibition, the more it forbids, the better it is.

4. A theory that is not refutable by any conceivable event is not scientific. Irrefutability is not a virtue.

5. Every genuine test of a theory is an attempt to falsify it or refute it. There are degrees of testability.

6. Confirming evidence should not count except when it is the result of a genuine test of the theory—it must be able to be presented as a serious but unsuccessful attempt to falsify it (corroborating evidence).

7. Some testable theories when found false are upheld by their admirers, e.g., by introducing *ad hoc* some auxiliary assumptions or reinterpreting it *ad hoc* so that it escapes refutation. The price paid is to destroy or lower the scientific status of the theory.

The upshot is “The criterion of the scientific status of a theory is its falsifiability.”

Do #2, and maybe #6 include a positive side….how can Popper argue for them?

Before pondering this….let's look at the big picture of how Popper claims to deal with Hume…

Popper’s attitude toward Hume’s Problem of Induction:

Hume was right. We can’t infer from constant conjunction to the next case

Popper criticizes Hume on psychological (e.g., the puppies) and logical grounds, p 43- 44.

What came first, the theory or the observation/

Two Options or Stances Given Hume's Problem: p. 45

1. We obtain knowledge non-inductively.

2. We obtain knowledge by induction and thus by a logically invalid and rationally unjustifiable procedure.

By non-inductive he means not by enumerative induction; we use observations to try to refute theories, and accept those that pass severe tests.

We actively try to impose regularities upon the world: Theory of trial and error, of conjectures and refutations.

Because no more rational course of action is open to us.
p. 51 “Assume we have made it our task to live in this unknown world of ours; …”

nothing "safer" than accepting the "best tested" theory.

(Question: is this convincing? What’s so good about a hypothesis that has passed a severe test, as Popper defines it? How can Popper even know a test is severe given especially that he thinks there are always infinitely many hypotheses?)

Popperians or Critical Rationalists

Critical rationalists: We may justify the rationality of (tentatively) accepting, or preferring, or believing the “best tested” theory *T*, without justifying *T* itself (as true, probably true)

According to Popper: Although it is not reasonable to believe that the method of conjecture and refutation will succeed, or is likely to succeed (i.e., that it is reliable), *No better than the amoeba*

p. 52: the critical attitude may be described as the conscious attempt to make our theories, our conjectures, suffer in our stead in the struggle for the survival of the fittest…a more dogmatic attitude would eliminate theories by eliminating us.

Summary of Popper's conclusions: p. 53-4

We can preserve the principle of empiricism: the fate of a theory is decided by observation and experiment, by the result of tests.

So long as a theory stands up to the severest tests we can design, it is accepted if it does not, it is rejected; but it is never inferred from the evidence, only the falsity of the theory can be inferred, and this inference is purely deductive.

What about corroboration?

Popper sets out those “confirmation measures” alongside Carnap.
But he declares that none of them measures corroboration *unless* they may be seen as the result of severe testing

Doubts that severity can be formalized any more than the rule of “total relevant evidence” among inductivists (i.e., probabilists)

Popper in a letter to me: I regret not having learned statistics

In looking to statistics to solve induction, Bayes theory philosophers generally looked to a Bayesian confirmation theory (logic or subjective)

(a) The overview of Popper: (pp. 75-82),
(b) The Live Exhibit (vi) (p. 88): Revisiting Popper's Demarcation of Science;
(c) (p. 108): "What Warrants Inferring a Hypothesis that Passes Severe Tests?

Any questions in relation to that assignment.

Bernoulli trials: Plain Jane Version

Let's start with a specific example and generalize, then go back to specifics and generalize some more (SIST p. 33)

4 Bernoulli trials. These have 2 possible outcomes, "success" or "failure", S or F (heads or tails, correct guess if milk put in first, winning ticket, etc.)

observed sample $x_0 = <S,S,F,S>$ (or x_{obs})

We can use a random variable, which takes value 1 whenever the trial is S, 0 when it's F.

$$x_0 = <1,1,0,1>$$

equivalently,

x_0 = <X_1=1, X_2 =1, X_3=0, X_1=1>

Let Pr(X = 1) = θ for any trial, and that trials are independent

θ a parameter; in the Bernoulli case it's from 0 to 1

If we knew θ, if we could compute

Pr(x_0; θ) = Pr(observed x_0; assuming prob of success at each trial = θ)

f(x_0; θ)

The **joint** outcome involves series of “ands”

x_0 = the 1st trial is 1 and 2nd trial is 1 and 3rd trial is 0 and 4th trial is 1

So, $Pr(x_0; \theta)$
= $Pr (X_1 = 1 \text{ and } X_2 = 1 \text{ and } X_3 = 0 \text{ and } X_4 = 1; \theta)$

Because the trials are *independent*, the probability multiplies

$Pr(x_0; \theta) = Pr(X_1 = 1; \theta)Pr (X_2 = 1; \theta)Pr (X_3 = 0; \theta)Pr (X_4 = 1; \theta)$

Suppose $\theta = .2$ (as in Royall’s example)

(e.g., 100 balls, 20 are red and we randomly draw, and success is getting a red ball)

What's Pr(X = 1) assuming the probability of X = 1 is $.2$?

Who is buried in Grant's tomb?

Therefore, Lik($\theta = .2$; x_0) = Pr(1,1,0,1; .2) = (.2)(.2)(.8)(.2)

Where did .8 come from?
If Pr(S = .2) then Pr(not-S) = .8
(since by the axioms, Pr(S or ~S) = 1 = Pr(S) + Pr (~S))

Note SIST error last line p. 33, it should be Lik(.2) because Royall is about to use H_0: $\theta \leq .2$ vs H_1: $\theta > .2$ to compare his likelihoodist inference with the frequentist significance test

We want to compare Lik($\theta = .2$; x_0) with the likelihood given $\theta = .8$ (measure of comparative "support")

Lik($\theta = .8$; x_0) = Pr(1,1,0,1; .8) = (.)(.)(.)(.)

.0064 vs. .1024

In general, with this x_0,

Lik(θ; x_0) = Pr(1,1,0,1; θ) = (θ)(θ)(1 - θ)(θ) =

$\theta^3(1 - \theta)$ order doesn't matter

So Lik($\theta = .2$; x_0) = Pr(1,1,0,1; .2) = (.2)(.2)(.8)(.2)
and Lik($\theta = .8$; x_0) = Pr(1,1,0,1; .8) = (.8)(.8)(.2)(.8)

LR ($\theta = .2$ over $\theta = .8$) = .0064 / .1024

$(.2)^3(.8)/ (.8)^3(.2) = (.25)^3(4) \sim .06$

Can also write the LR reverse LR ($\theta = .8$ over $\theta = .2$) = 16.6

It’s useful to start with the Likelihoodist, because it’s a key example of a logic of (comparative) evidence, and hits one of the big “wars”

Still we don’t usually crank out numbers;
My book does because it’s taking the criticisms in their actual location and the people arguing use numbers

.1054

Generalize for 4 Bernoulli trials

More generally, still for 4 trials, say we don't know the result,

Write the result of the kth trial is x_k as $X_k = x_k$
Random variable, capital X_k and lower case x_k is its value

$x_{obs} = (X_1 = x_1$ and $X_2 = x_2$ and $X_3 = x_3$ and $X_4 = x_4)$

$\Pr(x; \theta) = \Pr(x_1; \theta)\Pr(x_2; \theta)\Pr(x_3; \theta)\Pr(x_4; \theta)$

These should really be frequency distributions:
$f(x_1; \theta)\, f(x_2; \theta)\, f(x_3; \theta)\, f(x_4; \theta)$

Shortcut abbreviation for multiplying:
$f(x_1; \theta)\ f(x_2; \theta)\ f(x_3; \theta)\ f(x_4; \theta)$

$$\prod_{k=1}^{4} f(x_k; \theta)$$

Now take the Royall example on p. 34, n = 17, there are 9 successes and 8 failures (ugly numbers, they're his)

$\text{Lik}(x; \theta) = \theta^9 (1 - \theta)^8$

Observed proportion of successes = .53

Even without calculating,

$\theta = .53$ makes the observed outcome most probable, it's the maximally likely θ value

He fixes $\theta = .2$ and considers the Likelihood ratio of .2 and various alternatives

Since the sample proportion is .53, any value of θ further from .53 than .2 is will be less well supported than .2

Start with .2, .33 more takes us to .53, another .33 goes to .86

So any $\theta > .86$ is less likely than is .2

Likelihood ratio of .2 and .9

LR ($\theta = .2$ over $\theta = .9$) $= [.2^9 (.8)^8]/[[.9^9 (.1)^8] = 22.2$

(top p. 36)

both are too hideously small, we would never be computing them. But we can group

$(2/9)^9 \, (8)^8 \sim 22$ top of p. 36

Royall:
"Because H_0: θ ≤ .2 contains some simple hypotheses that are better supported than some hypotheses in H_1 (e.g., θ = .2 is better supported than θ = .9)…the law of likelihood does not allow the characterization of these observations as strong evidence for H_1 over H_0."

The significance tester tests H_0: θ ≤ .2 *vs.* H_1: θ > .2

So he rules out composite hypotheses.

The significance tester tests H_0: $\theta \leq .2$ *vs.* H_1: $\theta > .2$

She would reject H_0 and infer some (pos) discrepancy from .2

(observed mean M – expected mean under H_0) in standard deviation or standard error units

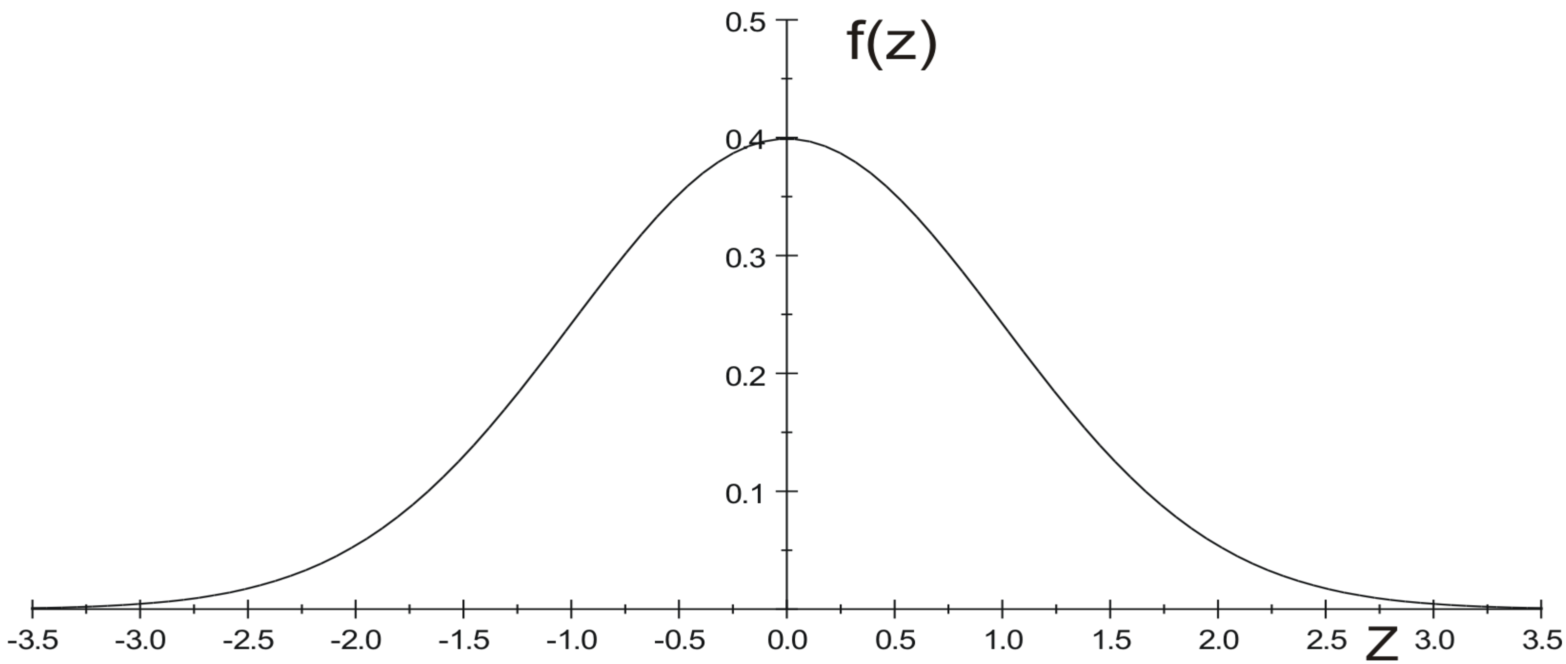
(.53 - .2)/.1 <u>~ 3.3</u>

Here 1 SE is .1

<u>Test Statistic</u> $d(x_0)$ is (.53 - .2)/.1

Lets us use the Standard Normal curve (we're using a Normal approximation)

(area to the right of 3) ~0, very significant.



$\Pr(d(X) \geq d(x_0); H_0) \sim .003$

$Pr(d(X) < d(x_0); H_0) \sim .997$

(see p. 35)

Admittedly, just reporting there's evidence H_1: $\theta > .2$, as our significance tester, doesn't seem so informative either.

In inferring $H_{1,}$ she is only inferring *some* positive discrepancy from .3

A 95 % confidence interval estimate, which we have not discussed, would be $.53 \pm 2SE$
$[.33 < \theta < .73]$

We'll see how severity also gives a report of discrepancy and has some advantages.

The Likelihoodist gives a series of comparisons: this is better supported than that, less strongly than some other value.

If you give enough comparisons, maybe our inferences aren't so different.

Is this really a statistical inference? Or just a report of the data? For the Likelihoodist it is, and the fact that a significance test is not comparative even precludes it from being a proper measure of evidence.

One Stat War Explained

Likelihoodists maintain that any genuine test or “rule of rejection” should be restricted to comparing the likelihood of H versus some point alternative H′ relative to fixed data x

No wonder the Likelihoodist disagrees with the significance tester.

Elliott Sober: “The fact that significance tests don’t contrast the null with alternatives suffices to show that they do not provide a good rule for rejection” (Sober 2008, p. 56).

The significance test has an alternative H_1: $\theta > 0.2$! (not a point)
(STINT p. 35)

While we're at notation: let's generalize for *n* Bernoulli trials

x_{obs} a member of the sample space: x_{obs} ε R real numbers)

$x_{obs} = X_1 = x_1$ and $X_2 = x_2$ and $X_3 = x_3$ ….and $X_n = x_n$

$Pr(x_{obs}; \theta) = f(x_1; \theta)\, f(x_2; \theta)\, f(x_3; \theta) \ldots f(x_n; \theta)$

Shortcut abbreviation:

$$\prod_{k=1}^{n} f(x_k; \theta)$$

$$\prod_{k=1}^{n} f(x_k; \theta)$$

I’ve run out of letters, let z = number of success out of n, n – k failures Lik(x; θ)

$$\theta^{z} (1 - \theta)^{n-z}$$

More notation z = $\sum_{k=1}^{n} x_k$