Meeting #3

- Review of logic
- Problem of induction
- Confirmation Theory/ Formal Epistemology

This will have some simple formal computations which can look dizzying,but I will make them very simply (I hope)

Little Bit of Logic

(Double purpose: both for arguing philosophically and for understanding inductive/deductive methods)

Argument:

A group of statements, one of which (the conclusion) is claimed to follow from one or more others (the premises), which are regarded as supplying evidence for the truth of that one.

This is written:

$$P_1, P_2, \ldots P_n / \therefore C.$$

In a 2-value logic, any statement *A* is regarded as true or false.

A <u>deductively valid argument</u>: *if* the premises are all true *then*, necessarily, the conclusion is true: the premises entail the conclusion.

To use the "⊨" (double turnstile) symbol:

$$P_1, P_2, \ldots P_n \models C.$$

Equivalently, the premises *entail* the conclusion.
$(P_1 \;\&\; P_2, \ldots \;\&\; P_n) \to C$ is a tautology

(Simple truth tables provide an algorithm or computable method to determine validity; no knowledge of context needed)

Adding a premise to a valid argument results in the argument still being valid

If you can't have $(P_1 \& P_2, \ldots \& P_n)$ true and C false, then you can't have $(J \& (P_1 \& P_2, \ldots \& P_n))$ true and C false

*Relevant to a key point today about "tacking paradox"

Many examples are on the handout, here's a valid argument:

If (*H*) all swans are white, then the next swan I see will be white.
The next swan I see x is black.
Therefore (not-*H*) Not all swans are white

(falsification)

(prefer the above to ex (1) on p. 60, an error to fix)

Logic of Simple Significance Tests: Statistical Modus Tollens

(Statistical analogy to the deductively valid pattern *modus tollens*)

If the hypothesis H_0 is correct then, with high probability, 1-p, the data would *not* be statistically significant at level p.

x_0 is statistically significant at level p.

____

Thus, x_0 is evidence against H_0, or x_0 *indicates* the falsity of H_0.

<u>Invalid Argument</u>: Consider this argument:

If H then E

E

———

$\therefore H$

If (H) everyone is a Bayesian, then (E) Lindley is a Bayesian.

<u>(E) Lindley is a Bayesian.</u>

So, (H) everyone is a Bayesian.

Affirming the consequent.

(T premises, false conclusion)

Invalid argument: An argument whose form makes it possible to have all true premises and a false conclusion without contradiction.

Begin with the Traditional Problem of Induction

2.1 Traditional Problem of Induction

Inductive argument. With an inductive argument, the conclusion goes beyond the premises. So it's logically possible for all the premises to be true and the conclusion false: invalid.

The traditional problem of induction seeks to justify *enumerative induction (EI)* (*straight rule* of induction).

Infer from past cases of A's that were B's to all or most A's will be B's:

(EI) Prem: All observed A_1, A_2, ….A_n have been B's
Therefore, *H*: all (or most) A's are B's.

The premises might be experimental outcomes or data points:

$X_1, X_2, \ldots X_n$ (e.g., light deflection observations, drug reactions, radiation levels in fish)

Even if all observed cases had a feature F, or followed a law, there's no logical contradiction in the falsity of the generalization:

H: All *E*'s are *F*

This is also true for a statistical generalization:

90% in this class have property Q

Thus, *H*: 90% of all people have property Q.

***H* agrees with the data, but it's possible to have such good agreement even if *H* is false.**

(Asymmetry with falsification)

Exhibit (i). *Justifying Induction is Circular.* The traditional problem of induction is to justify the conclusion:

Conclusion: (EI) is rationally justified, it’s a reliable rule.

We need an argument for concluding (EI) is reliable.

That’s what is meant by “rationally” justifying it.

…….

Conclusion: The ‘inductive method’ (EI) is reliable (it will work in the future): inferring from past to future success is a reliable method.

What will the premises be?

a. Use an inductive argument to justify induction:

Premise: The inductive method has worked, has been reliable, in the past.

Conclusion: The inductive method is reliable (it will work in the future), i.e., inferring from past cases to future cases is a reliable method.

Problem: circular. (It uses the method in need of justification to justify that method.)

b. Use a deductive argument to justify induction:

Premise: If a method has worked (been reliable) in the past, then it will work in the future.

Premise: The inductive method has worked, has been reliable, in the past.

Conclusion: the inductive method is reliable (it will work in the future), i.e., inferring from past cases to future cases is a reliable method.

***Problem**: we cannot say it's a sound argument.*

In order to infer the truth of the conclusion of a deductively valid argument, the premises must be true, i.e., the argument must be sound.

You'd have to know the very thing which the argument was supposed to justify!

We need to deductively infer EI will be reliable *in general*: the known cases only refer to the past and present, not the future.

Alternatively put in terms of assuming the **uniformity of nature**

Logical problem of induction: can't use logic to solve it.

Attempts to dissolve the problem (not in SIST)

- It's asking for justification beyond where it's appropriate,
- It's converting induction to deduction (i.e., it's asking for a certainly true conclusion from true premises)
- That's just what we <u>mean</u> by rational.

To see what's wrong with the last, consider my friend the crystal gazer…

Counterinductive method:

Infer from All A's have been B's in the past to the next A will not be a B.

In terms of a method:

Infer from the fact that a method M has worked poorly (been unreliable) in the past that M will work well in the future. (The crystal gazing)

If a method M has worked poorly (been unreliable) in the past then M will work well in the future.

M has worked poorly (been unreliable) in the past

Therefore, M will work well in the future.

So, unless we allow this justification of counterinduction, we should not allow the appeal to 'that's what we mean by rational'

You might say, you just can't argue with someone who accepts counterinduction as rational.

You can't convince them that induction is superior to counterinduction, but all we care about is showing it's rational *for us to accept scientific induction*

This takes us to formal confirmation theory

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

we can build and try to justify "inductive logics"

straight rule: Assign degrees of confirmation/credibility

Statistical affinity

Bayesian (and likelihoodist) accounts

Deductive Testers

we can reject induction and uphold the "rationality" of preferring or accepting H if it is "well tested"

Statistical affinity

Fisherian, Neyman-Pearson methods: probability enters to ensure reliability and severity of tests with these tests.

Focus on inductive logics or formal confirmation theory

Exhibit (ii). *Probabilistic Affirming the Consequent.*

They didn't renounce enumerative induction (EI), they sought logics that embodied it (EI):

If ***H*** (all A's are B's), then **data**: all n observed A's are B's.

All n observed A's (A_1, A_2, ….A_n) are B's
Therefore, *H*: all A's are B's.

This is *affirming the consequent*.

Probabilistic affirming the consequent says only the the probability of the conclusion goes up.
It a boost in confirmation or probability–a *B-boost*.

This is what Bayes' theorem gives us.

How do we obtain the probabilities?

Rudolf Carnap tried to deduce them from the logical structure of a particular (first order) language.

The degree of probability, a rational degree of belief, would hold between two statements, a hypothesis and the data.

C(H,$\boldsymbol{x}$) symbolizes “the confirmation of H, given $\boldsymbol{x}$”.

Once you chose the initial assignments to core states of the world, calculating degrees of confirmation is a formal or syntactical matter, much like deductive logic.

They used logical probabilities:
Goal: measure the *degree of implication* or confirmation that ***x*** affords *H*.

> Carnap has stated that the ultimate justification of the axioms is inductive intuition. I do not consider this answer an adequate basis for a concept of rationality. Indeed, I think that *every* attempt, including those by Jaakko Hintikka and his students, to ground the concept of rational degree of belief in logical probability suffers from the same unacceptable ***apriorism***. (Salmon 1988, p. 13).

What does a highly probable claim, according to a particular inductive logic, have to do with the real world? How can it provide “a guide to life?” (E.g., Kyburg 2003, Salmon 1966.)

The hankering for an inductive logic remains.

It’s behind the appeal of the default Bayesianism of Harold Jeffreys, and other attempts to view probability theory as extending deductive logic.

Exhibit (iii). Hacking announced (1.4): “there is no such thing as a logic of statistical inference” (1980, p. 145),

Not only did all attempts fail; he recognized the project is “founded on a false analogy with deductive logic” (ibid.). He follows Peirce:

> In the case of analytic [deductive] inference we know the probability of our conclusion (if the premises are true), but in the case of synthetic [inductive] inferences we only know the degree of trustworthiness of our proceeding (Peirce 2.693).

In ampliative, or inductive reasoning, the conclusion should go beyond the premises; probability enters to qualify the overall “trustworthiness” of the method.

My argument from coincidence to weight gain in (1.3) inferred

H: I've gained at least 4 pounds

- The inference is qualified by the detailed data and information on how capable the method is at blocking erroneous pronouncements of my weight.
- *What is being qualified probabilistically is the inferring or testing process.*
- By contrast, in a probability or confirmation logic what is generally detached is the probability of *H*, given data. It is a *probabilism*.

Take note of a great quote by Fisher (p. 66)

In deductive reasoning all knowledge obtainable is already latent in the postulates. Rigour is needed to prevent the successive inferences growing less and less accurate as we proceed. The conclusions are never more accurate than the data. In inductive reasoning we are performing part of the process by which new knowledge is created. The conclusions normally *grow more and more accurate* as more data are included. **It should never be true, though it is still often said, that the conclusions are no more accurate than the data on which they are based.** (Fisher 1935, p. 54; my emphasis)

(linked vs convergent argument)

Basics of Probability (from Hacking handout)

I will try to make this painless

Basics of Probability (from Hacking handout)

The rules that follow are informal versions of standard axioms for elementary probability theory.

ASSUMPTIONS

The rules stated here take some things for granted:

- The rules are for finite groups of propositions (or events).
- If A and B are propositions (or events), then so are AvB, A&B, and ~A.
- Elementary deductive logic (or elementary set theory) is taken for granted.
- **If A and B are logically equivalent, then Pr(A) = Pr(B). [Or, in set theory, if A and B are events which are provably the same sets of events, Pr(A) = Pr(B).]**

NORMALITY

The probability of any proposition or event A lies between 0 and 1.

$$0 \leq Pr(A) \leq 1$$

CERTAINTY

An *event* that is sure to happen has probability 1.

A *proposition* that is certainly true (a tautology) has probability 1.

$$Pr(\text{certain proposition}) = 1$$

$$Pr(\text{sure event}) = 1$$

$Pr(\Omega) = 1$, Ω is all possible outcomes (outcome space)

ADDITIVITY

If two events or propositions A and Bare mutually exclusive (disjoint, incompatible), the probability that one or the other happens (or is true) is the sum of their probabilities.

If A and B are mutually exclusive, then

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

(e.g., Pr(6 or 4; single fair toss of a die) = 1/6 + 1/6 = 2/6

$Pr(H \vee \sim H) = 1$)

CONDITONAL PROBABILITY ("/" vs ";" :conditional on vs "under)

Now comes a definition.

If $Pr(B) > 0$, then $Pr(A/B) = \dfrac{Pr(A\&B)}{Pr(B)}$

Pr(5 /outcome is an odd #) = 1/3

It's gone up from Pr(5) (1/6)

Still need to know how to compute it, e.g., that it's a fair toss, still need ";"

Pr(5; toss of fair die) = 1/6

CONDITONAL PROBABILITY

The only basic rules are (1)-(3). Now comes a definition.

$$\text{If } Pr(B) > 0, \text{ then } Pr(A/B) = \frac{Pr(A\&B)}{Pr(B)}$$

3 axioms and a definition of conditional probability

MULTIPLICATION

The definition of conditional probability implies that:

If $Pr(B) > 0$, $Pr(A\&B) = Pr(A)Pr(B/A)$

$= Pr(B)Pr(A/B)$.

Using the propositions common in confirmation theory:

$$\Pr(H/x) = \frac{\Pr(H \,\&\, x)}{\Pr(x)}$$

Use the multiplication or “and” rule for the numerator

$$\Pr(H/x) = \frac{\Pr(H)\,\Pr(x/H)}{\Pr(x)}$$

Use Total Probability to spell out the denominator (p.24):

x ≡ ((x & H) or (x &~H)) __________

Pr(x) = P(x&H) + Pr(x&~H) __________

=Pr(H)Pr(x/H) + Pr(~H)Pr(x/~H) ___________

From SIST p. 24

A central way to formally capture probabilism is by means of the formula for conditional probability, where $\Pr(\boldsymbol{x}) > 0$:

$$\Pr(H \mid \boldsymbol{x}) = \frac{\Pr(H \text{ and } \boldsymbol{x})}{\Pr(\boldsymbol{x})}.$$

$\Pr(\boldsymbol{x}) = \Pr(\boldsymbol{x} \mid H)\Pr(H) + \Pr(\boldsymbol{x} \mid \sim H)\Pr(\sim H)$, we get:

$$\Pr(H \mid \boldsymbol{x}) = \frac{\Pr(\boldsymbol{x} \mid H)\Pr(H)}{\Pr(\boldsymbol{x} \mid H)\Pr(H) + \Pr(\boldsymbol{x} \mid \sim H)\Pr(\sim H)}$$

where $\sim H$ is the denial of H.

From SIST p. 24

All alternatives to H should be included in the "catchall" ~H

Each H_i is called the *prior probability*

Pr(H/***x***) = the posterior probability

Using Bayes' Theorem doesn't make you a Bayesian.

2.2 *Is Probability a Good Measure of Confirmation?* *(p. 66)*

(1) *Incremental* (B-boost)
H is confirmed by $\boldsymbol{x}$ iff $\Pr(H|\boldsymbol{x}) > \Pr(H)$,
H is disconfirmed iff $\Pr(H|\boldsymbol{x}) < \Pr(H)$.

(2) *Absolute*: H is confirmed by $\boldsymbol{x}$ iff $\Pr(H|\boldsymbol{x})$ is high, at least greater than .5

Note $\Pr({\sim}H|\boldsymbol{x}) = 1 - \Pr(H|\boldsymbol{x})$,

The most familiar (Bayesian) interpretation of confirmation is:
H is confirmed by ***x*** if ***x*** gives a boost to the probability of *H*, *incremental* confirmation.

The components of C(*H*,***x***) can be any statements, no reference to a probability model is required

If H entails $\boldsymbol{x}$, then assuming $\Pr(\boldsymbol{x}) \neq 1$, $\Pr(H) \neq 0$, we have $\Pr(H|\boldsymbol{x}) > \Pr(H)$.

This is an instance of probabilistic affirming the consequent.

(Note: if $\Pr(H|\boldsymbol{x}) > \Pr(H)$ then $\Pr(\boldsymbol{x}|H) > \Pr(\boldsymbol{x})$. Note 4, p. 69)

$$\frac{\Pr(H|x)}{\Pr(H)} = \frac{\Pr(x|H)}{\Pr(x)}$$

Simple way to see this (viewing *H* and ***x*** as statements*)

Thus

$$\Pr(\boldsymbol{x}) \Pr(H|\boldsymbol{x}) = \Pr(H) \Pr(\boldsymbol{x}|H)$$

Divide both sides by $\Pr(\boldsymbol{x})\Pr(H)$

$$\frac{\Pr(\mathrm{H}|\mathrm{x})}{\Pr(\mathrm{H})} = \frac{\Pr(x|H)}{\Pr(\boldsymbol{x})}$$

Airport alarm (p. 66)

From (1), ***x*** (the alarm) disconfirms the hypothesis *H*: the bag is clean, because its probability has gone down however slightly.

(it gives some confirmation to its violating TSA rules)

Yet from

(2) ***x*** confirms *H*: bag is clean, as Pr(*H*) is high to begin with.

Incremental confirmation is generally used in current Bayesian epistemology: Confirmation is a B-boost.

Note how easy it is to get a B-boost:

You've observed x some patients who improved
Let H be: all patients improve

H entails x

So Pr(x/H) = 1

So unless Pr(x) was 1, you get 'confirmation'

$$\frac{\Pr(x|H)}{\Pr(\boldsymbol{x})} = \frac{\Pr(H|\mathrm{x})}{\Pr(\mathrm{H})}$$

Exhibit (iv). *Paradox of Irrelevant Conjunctions* (Glymour 1980) "tacking paradox" (69-70)

If ***x*** confirms *H*, then ***x*** also confirms (*H* & *J*), even if hypothesis *J* is just "tacked on" to *H*

J is an *irrelevant conjunct* to *H*, with respect to evidence ***x***: Pr(***x***|*J*) = Pr(***x***|*J* & *H*).

For instance, ***x*** might be radioastronomic data in support of:

H: the GTR deflection of light effect is 1.75" and

J: the radioactivity of the Fukushima water being dumped in the Pacific Ocean is within acceptable levels.

(A) If $\boldsymbol{x}$ confirms H, then $\boldsymbol{x}$ confirms (H & J),where $\Pr(\boldsymbol{x}|H \,\&\, J) = \Pr(\boldsymbol{x}|H)$ for any J consistent with H.

The reasoning is as follows:

(i) $\Pr(\boldsymbol{x}|H)/\Pr(\boldsymbol{x}) > 1$ ($\boldsymbol{x}$ Bayesian confirms H)

(ii) $\Pr(\boldsymbol{x}|H \,\&\, J) = \Pr(\boldsymbol{x}|H)$ (J's irrelevance is given)

Substituting (ii) into (i) we get $[\Pr(\boldsymbol{x}|H \,\&\, J)/\Pr(\boldsymbol{x})] > 1$

Therefore $\boldsymbol{x}$ Bayesian confirms (H & J)—meaning, the prob of the conjunction goes up with **x**

In fact, the conjunction gets just as much of a boost as does H

p. 70 typo in the denominator of R: $\frac{\Pr(H|x)}{\Pr(H)}$

The conjunction (*H* & *J*) can get less of a confirmation boost than does *H* if we use a different measure of the boost.

But aren't they uncomfortable with (A), allowing (*H* & *J*) to be confirmed by ***x***?

Clark Glymour: we're not happy with an account that tells us deflection of light data confirms that GTR is true and GTR and the radioactivity of the Fukushima water is within acceptable levels, while assuring us that ***x*** does not confirm the Fukushima water having acceptable levels of radiation? (70)

Glymour goes further (interested participants can read):

It is plausible to hold what philosophers call the “special consequence” condition (p. 70):

If ***x*** confirms (H & J), then ***x*** confirms J.

It follows that if ***x*** confirms *H*, then ***x*** confirms *J* for any irrelevant *J* consistent with *H* (neither *H* nor *J* have probabilities 0 or 1).

Most Bayesian epistemologists reject special consequence, so they can avoid this.

What Does the Severity Account Say? (71)

Our account of inference disembarked way back at (1): that ***x*** confirms *H* so long as Pr(*H*|***x***) > Pr(*H*).

We reject probabilistic affirming the consequent.

The simplest case has *H* entail ***x***, and ***x*** is observed.
(We assume the probabilities are well defined, and *H* doesn't already have probability 1.)

H gets a B-boost, but there are many other "explanations" of ***x***.

It's the same reason we reject the Law of Likelihood (LL).

Unless they have been probed, finding an *H* that fits ***x*** is not difficult to achieve even if *H* is false: *H* hasn't passed severely.

Now the confirmation theorist is only saying a B-boost suffices for *some* evidence.

To us, to have *any* evidence, or even the weaker notion of an “indication”, requires a minimal threshold of severity be met.

How about tacking?

The process of tacking–one form–is once you have an incrementally confirmed *H* with data ***x***, tack on any consistent *J* and announce “***x*** confirms (*H* & *J*)”.

Let's allow that (*H* & *J*) fits or accords with ***x*** (since GTR entails or renders probable the deflection data ***x***).

But nothing has been done to measure the radioactivity of the Fukushima water being dumped into the ocean (to check J).

What They Call Confirmation We Call Mere "Fit" or "Accordance"

There are many other famous paradoxes of confirmation theory (e.g., the white shoe confirming all ravens are black and the grue paradox)

We make fun of them, but they often contain a puzzle of relevance for statistical practice.

There are two reasons the tacking paradox above is of relevance to us: There is a large-scale theory *T* that predicts ***x***, and we want to discern which portion of *T* to credit.

Severity says: do not credit those portions that could not have been found false, even if they're false. They are poorly tested.

Second, the question of measuring support with a Bayes boost or with posterior probability arises in Bayesian statistical inference

When you hear that what you want is some version of probabilism, be sure to ask if it’s a boost (and if so which kind) or a posterior probability, a likelihood ratio or something else.

True, the Bayesian epistemologist invites trouble by not clearly spelling out corresponding statistical models

They seek a formal logic, holding for statements about radiation, deflection, fish or whatnot: a mistake, I say.

We can have a general account for statistical inference, it just won't be purely formal.

A more adequate semi-formal epistemology can be gotten by considering the error probing properties of the overall method

Statistical Foundations Need Philosophers of Statistics

The idea of putting probabilities over hypotheses delivered to philosophy a godsend, an entire package of superficiality." (Glymour 2010, p. 334).

Given a formal epistemology, the next step is to use it to represent or justify intuitive principles of evidence (meta-methodology, meeting #1)

The problem to which Glymour is alluding is: you can start with the principle you want your confirmation logic to reflect, and then *reconstruct* it using probability.

We won’t be working in formal epistemology.
This is to give you a sense of that field.

But we do care about the very basic Bayesian computations we’ve seen today.