Bayesian Confirmation Theory

January 11, 2021

Probability Review

The Probability Axioms

[NORMALITY]

Every proposition (over which *Pr* is defined) is assigned a probability somewhere between 0 and 1.

$$0 \leq \Pr(X) \leq 1 \tag{1}$$

[CERTAINTY]

Any proposition Ω that is certain to be true (e.g., a logical truth) is assigned probability 1.

$$\Pr(\Omega) = 1 \tag{2}$$

[ADDITIVITY]

If propositions X and Y are *mutually exclusive*, then the probability of their disjunction is equal to the sum of their probabilities.

Two propositions are *mutually exclusive* just in case they cannot *both* be true.

$$\text{If } X\&Y \text{ are mutually exclusive, } \Pr(X \vee Y) = \Pr(X) + \Pr(Y) \tag{3}$$

The Overlap Rule

What is the probability of a disjunction when its disjuncts are *not* mutually exclusive?

[OVERLAP]

The probability of a disjunction is equal to the sum of the probabilities of its disjuncts minus the probability its disjuncts' overlap.

$$\Pr(X \vee Y) = \Pr(X) + \Pr(Y) - \Pr(X \wedge Y) \tag{4}$$

This is the intuitive idea behind The Overlap Rule: if the propositions X and Y are not mutually exclusive, then by adding $\Pr(X)$ to $\Pr(Y)$ in order to get $\Pr(X \vee Y)$, we are "double counting" the possibility in which they are *both* true, i.e., $(X \wedge Y)$; to correct for this, we need to subtract out $\Pr(X \wedge Y)$.

Conditional Probability

Let $\Pr(X \mid Y)$ be the probability of X *conditional* on Y.
It is defined as follows:

$\Pr(X|Y)$ is, roughly, the probability that X is the case on the assumption that Y is the case.

$$\Pr(X \mid Y) = \frac{\Pr(X \wedge Y)}{\Pr(Y)} \tag{5}$$

More Rules and Definitions

The Multiplication Rule: If $\Pr(E) > 0$, then

$$\Pr(X \wedge E) = \Pr(X \mid E) \cdot \Pr(E) \tag{6}$$

The Total Probability Rule: If $0 < \Pr(E) < 1$, then

$$\Pr(X) = \Pr(X \mid E) \cdot \Pr(E) + \Pr(X \mid \neg E) \cdot \Pr(\neg E) \tag{7}$$

The Logical Consequence Rule: Suppose that Y logically entails X. Then

$$\Pr(Y) \leq \Pr(X) \tag{8}$$

Statistical Independence. X and Y are said to be *statistically independent* just in case $\Pr(X \mid Y) = \Pr(X)$.

If X and Y are statistically independent, $\Pr(X \wedge Y) = \Pr(X) \cdot \Pr(Y)$.

Proof. From **The Multiplication Rule:**

$$\Pr(X \wedge Y) = \Pr(X \mid Y) \cdot \Pr(Y)$$

And, from the definition of **Statistical Independence:**

$$\Pr(X \mid Y) = \Pr(X)$$

So, $\Pr(X \wedge Y) = \Pr(X) \cdot \Pr(Y)$.

Inductive Logic and Confirmation

Project: develop a purely syntactic theory of *inductive reasoning*, which improves on the limitations of the “deductive" method.

see: Hempel and Carnap, in the first half of the twentieth century.

HYPOTHETICO-DEDUCTIVE METHOD

Work out which observation statements Os are entailed by a theory T.

1. If T entails O, then O *confirms* T.
2. If T entails O, then $\neg O$ *refutes* T.

It would be nice to say more—especially about the cases in-between these two extremes.

In addition, inductive reasoning was often assumed to validate the following principle.

INSTANCE PRINCIPLE:

Observations of instances of a generalization *confirm* that generalization.

Hempel called this "Nicod’s Criterion".

Generalizations have the form

All Fs are G.

Observing an instance of this generalization would be to learn of some particular F, a, that a is G.

Hempel identified various formal features that we might want the confirmation-relation to possess.

Carnap developed an account of confirmation in terms of probabilistic relevance. Carnap’s view is an early version of Bayesian Confirmation Theory.

Three Puzzles of Confirmation

The Paradox of the Ravens

Consider the following hypothesis:

$H =$ *All Fs are G.*

Hypothesis H is logically equivalent to hypothesis H^*

$H^* =$ *All non-G things are non-Fs.*

If the INSTANCE PRINCIPLE is correct, then observing that some non-G thing is also a non-F confirms the hypothesis H^*. But, because H and H^* are logically equivalent, such an observation also confirms hypothesis H.

Example. Observing that this white shoe is not a raven confirms the hypothesis that all ravens are black.

But that seems absurd! Indoor ornithology?

We can translate H into

For all things x, if x is an F, then x is a G.

And we can translate H^* into

For all things x, if x is not a G, then x is not an F.

And both statements can be understood as saying "for all things x, either x is not F, or x is G." They are logically equivalent.

The "Problem" of Irrelevant Conjunctions

Consider the following two plausible principles of inductive logic

TWO PRINCIPLES OF INDUCTIVE LOGIC

Special Consequence Condition (SC): If E confirms H, and H entails H^*, then E confirms H^*.

Converse Consequence Condition (CC): If E confirms H, and H^* entails H, then E confirms H^*.

Problem: These two principles entail that if E confirms something, then E confirms anything!

Proof of the Problem. Suppose that E confirms H. We will show that for any A, E then confirms A.

From **(CC)**, E confirms $H \wedge A$. And $H \wedge A$ entails A; so, by **(SC)**, E confirms A.

Nelson Goodman's "New Riddle of Induction"

Consider the following properties:

$$x \text{ is } \textit{grue} \text{ iff}_{df} \begin{cases} x \text{ is green} & \text{if } x \text{ is observed before 2022} \\ x \text{ is blue} & \text{if } x \text{ is not observed before 2022} \end{cases}$$

$$x \text{ is } \textit{bleen} \text{ iff}_{df} \begin{cases} x \text{ is blue} & \text{if } x \text{ is observed before 2022} \\ x \text{ is green} & \text{if } x \text{ is not observed before 2022} \end{cases}$$

Suppose that you observe a green emerald. According to the INSTANCE PRINCIPLE, this confirms the hypothesis *"All emeralds are grue"*. But should it?

Bayesian Confirmation Theory

What is it for some evidence E to provide some confirmation for a hypothesis H?

We are using 'confirms' in a technical sense to mean something like "E is evidence for H" or "E supports H," etc.

BAYESIAN CONFIRMATION THEORY.

Evidence E *confirms* hypothesis H just in case

$$\Pr(H \mid E) > \Pr(H) \tag{9}$$

Bayesian Confirmation Theory makes great use of *Bayes' Theorem*.

Bayes' Rule. Assume that $\Pr(E) > 0$. Then,

$$\Pr(H \mid E) = \frac{\Pr(E \mid H) \cdot \Pr(H)}{\Pr(E)} \tag{10}$$

The theorem follows directly from the definition of conditional probability. Let's try to prove it!

$$\Pr(X \mid Y) = \frac{\Pr(X \wedge Y)}{\Pr(Y)}, \text{ if } \Pr(Y) > 0$$

Because of the *Total Probability Rule*, the theorem can be re-written as follows:

$$\Pr(H \mid E) = \frac{\Pr(E \mid H) \cdot \Pr(H)}{\Pr(E \mid H) \cdot \Pr(H) + \Pr(E \mid \neg H) \cdot \Pr(\neg H)} \tag{11}$$

This allows us to calculate the confirmatory support that some evidence might supply for a hypothesis by making use of information that we might very well have available to us—e.g., the *likelihood* of that evidence according to each hypothesis plus your *priors* in those hypotheses.

Feature of Bayesian Confirmation Theory

Bayesian Multiplier. Here's a helpful way to think about what Bayesian Confirmation Theory says about confirmation.

Bayesian Multiplier + **Priors** Assume that $\Pr(E) > 0$. Then,

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)}{\Pr(E)} \cdot \Pr(H)) \tag{12}$$

If $\frac{\Pr(E|H)}{\Pr(E)} > 1$, then evidence E confirms hypothesis H. If the Bayesian Multiplier is less than 1, it disconfirms the hypothesis.

This allows us to notice some general, interesting properties that confirmation has, e.g.,:

1. The Hypothetico-Deductive Principle: If H entails E, then observing E confirms H.

2. The Surprising Evidence Principle: The more surprising the evidence, the more it confirms the hypothesis

3. The Likelihood Lover's Principle: The higher the physical probability that H assigns to E, the more strongly H is confirmed by the observation of E.

This can be especially useful if we notice that hypothesis (sometimes with the help of auxiliary assumptions) entail physical probabilities for bits of evidence: $\Pr^*_H(E)$.

Bayesian Confirmation Theory & the Three Puzzles

1. Ravens.

2. Conjunctions

3. Grue

Worries About Bayesian Confirmation Theory

What constrains our priors? Is the view too subjective?

Problem of Old Evidence.

Problem of New Theories.

Appendix

Proof of Bayes' Theorem

Definition of Conditional Probability:

$$\Pr(X \mid Y) = \frac{\Pr(X \wedge Y)}{\Pr(Y)}, \text{ if } \Pr(Y) > 0$$

$$\begin{aligned} \Pr(X \mid Y) \cdot \Pr(Y) &= \frac{\Pr(X \wedge Y)}{\Pr(Y)} \cdot \Pr(Y) \\ &= \Pr(X \wedge Y) \end{aligned}$$

This is one way to derive *Bayes' Rule* from the definition of Conditional Probability. There are other ways to do it, too.

$$\begin{aligned} \frac{\Pr(X \mid Y) \cdot \Pr(Y)}{\Pr(X)} &= \frac{\Pr(X \wedge Y)}{\Pr(X)} \\ &= \Pr(Y \mid X) \end{aligned}$$