Bayesian Approach

Aditya F. Ihsan

Imagine someone from 17th century accidentally arrive to present time. He will be amazed to almost everything that science and technology is capable to do, while we see all of it as normal.

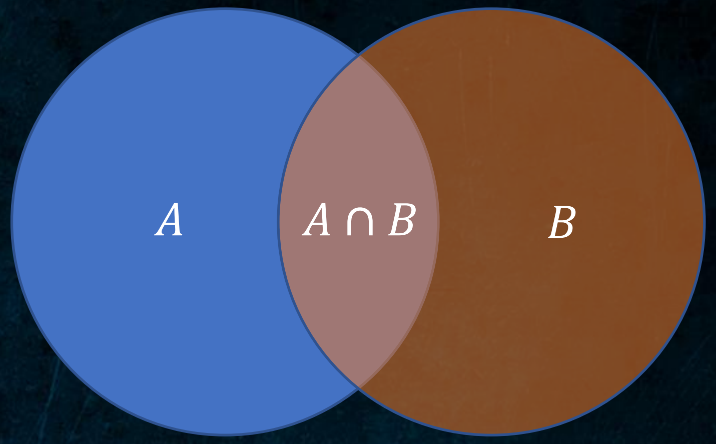
This is
Bayesian Framework

"Evidence update beliefs, not determine it"

1.
What is Bayesianism?

The core behind Bayesian framework is Bayes Theorem

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$



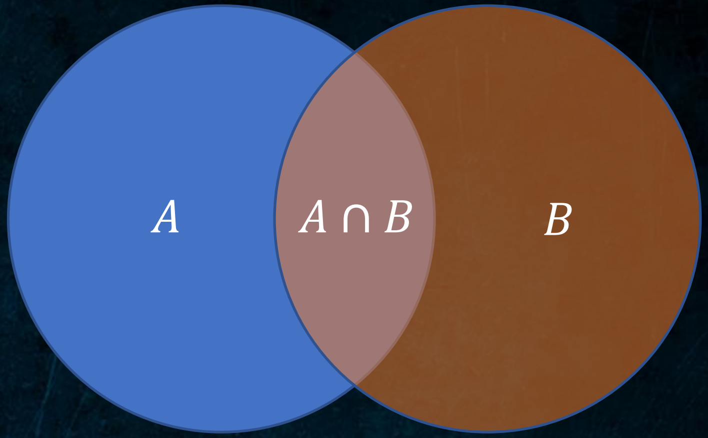
And the core concept behind Bayes Theorem is conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

What is Bayes Theorem

If A and B are independent events, then
$P(A \cap B) = P(A)P(B)$, which implies

$$P(A|B) = P(A)$$



$P(A|B) = 1$ means A directly follows from B, and

$P(A|B) = 0$ means B is against A

What is Bayes Theorem

Bayes Theorem is direct derivation of definition of conditional probability, and the fact that intersection is commutative:

$$P(A \cap B) = P(B \cap A)$$

Thus,

$$P(A|B)P(B) = P(B|A)P(A)$$

And voila, we have Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

What is Bayes Theorem

2.

If it is just a probability, then why is this important?

The formula of Bayes Theorem is just a mathematical fact.

What makes it special is its interpretation.

We can write it as

$$P(A|B) = P(A)\frac{P(B|A)}{P(B)}$$

which intuitively can be seen as

$$(A\ after\ B) = (A\ before\ B) * some\ factor\ of\ B$$

It tells us how probability of A changes after an event B

Behind Bayesian Statistics

In general sense, we write it as

$$Posterior = Prior * \frac{Likelihood}{Normalization}$$

In Science, it can be used to see how evidence e confirms a hypothesis h:

$$P(hypothesis|evidence) = P(hypothesis) * \frac{P(e|h)}{P(e)}$$

Behind Bayesian Statistics

If $P(e|h) > P(e)$ or $P(h|e) > P(h)$ then e confirms h

If $P(e|h) < P(e)$ or $P(h|e) < P(h)$ then e disconfirms h

If $P(e|h) = P(e)$ or $P(h|e) = P(h)$ then e is neutral with h

Behind Bayesian Statistics

3.

Is it objective to be applied in Science?

Bayesian statistics changes radically how probability is interpreted.

In classical sense, probability of an event is the frequency of the event occurring in some experiments or processes.

In Bayesian interpretation, probability of an event become the degree of belief of that event.

It is because in light of a new information, the probability updates itself (prior to posterior)

Objective vs subjective

Objective side of Bayesians:

Prior should be defined “objectively”, and thus posterior represent probabilities that rational agents ought to accept

However, it is unclear how we define prior in objective way. To define prior objectively, we may need to list all possible hypotheses, which is practically insuperable.

Objective vs subjective

The possible choice: Subjective Bayesian

Prior probabilities is dependent on subjective degree of belief of an individual.

Alternative (Dorling, 1979): probability is interpreted as measure in scientific practice that reflect scientist behaviour (similar to gambling). However, it is unclear what is within scientific practice that corresponds.

Objective vs subjective

The possible choice: Subjective Bayesian

Prior probabilities is dependent on subjective degree of belief of an individual.

The objective side of this is that how the prior, as subjective as it is, will objectively changes to posterior in light of new evidence.

Objective vs subjective

4.

What are some of its properties?

Another way to see

$$P(h|e) = \frac{P(h)P(e|h)}{P(h)P(e|h) + P(\neg h)P(e|\neg h)}$$

The yellow part is actually $P(e)$

Observe,

If $P(h) = 0$, then whatever the evidence is, $P(h|e) = 0$

If $P(h) = 1$, then whatever the evidence is, $P(h|e) = 1$

Some Observations

Convergence of different prior

Example: Let $P(e|h) = 0.8$

Prior 1: $P(h) = 0.3$. Then

$$P(h|e) = \frac{0.3 \times 0.8}{0.3 \times 0.8 + 0.7 \times 0.2} = 0.63$$

Prior 2: $P(h) = 0.6$,

$$P(h|e) = \frac{0.6 \times 0.8}{0.6 \times 0.8 + 0.4 \times 0.2} = 0.857$$

Different scientist start with different beliefs are eventually converges to higher belief with good evidence

Some Observations

Multiple evidences

$$P(h|e_1 \cap e_2 \cap \cdots \cap e_n) = \frac{P(h)}{P(e_1)P(e_2|e_1)\cdots P(e_n|e_1 \cap \cdots \cap e_{n-1})}$$

First evidence:

$$P(h|e_1) = P(h)\frac{P(e_1|h)}{P(e_1)}$$

Second evidence:

$$P(h|e_1 \cap e_2) = P(h|e_1)\frac{P(e_2|e_1 \cap h)}{P(e_2|e_1)}$$

If $e_2 = e_1$, then $P(e_2|e_1) = 1$, thus the factor is not that large.

Some Observations

Multiple evidences

The point is that there are diminishing returns from efforts to confirm a theory by a single kind of evidence.

By contrast, the prior probability of a theory being confirmed by some new kind of evidence may be quite low. However, once it occurs, it leads to a significant increase in the probability ascribed to the theory.

Some Observations

5.

What is the advantage of this approach?

1. Duhem-Quine Problem:

It is impossible to experimentally test a scientific hypothesis in isolation, because an empirical test of the hypothesis requires one or more auxiliary/background assumption

Bayesian approach can “solve” this problem

Bayesian Explanation

Prout's Case Example:

William Prout hypothesize that atomic weight of all elements are whole number multiples of atomic weight of hydrogen (h).

Let also a is the accuracy of atomic weight measurement.

We set

$P(a) = 0.6$ (chemist are quite confident with the measurement even thought some uncertainties and impurities are possible

$P(h) = 0.9$ (Prout is very confident with the hypothesis)

Objective vs subjective

Prout’s Case Example:

Prout then take an experiment e to compute the atomic weight of Chlorine. By his hypothesis it should be around 35.83 - 3.6

By uniform distribution, we set then

$$P(e|\neg h \cap a) = 0.01$$
$$P(e|\neg h \cap \neg a) = 0.01$$
$$P(e|h \cap \neg a) = 0.02$$

Objective vs subjective

Prout's Case Example:

Compute some probabilities

$$P(e|\neg h) = P(e|\neg h \cap a)P(a) + P(e|\neg t \cap \neg a)P(\neg a)$$
$$= 0.01 \times 0.6 + 0.01 \times 0.4 = 0.01$$
$$P(e|h) = P(e|h \cap a)P(a) + P(e|h \cap \neg a)P(\neg a)$$
$$= 0 + 0.02 \times 0.4 = 0.008$$
$$P(e|a) = P(e|\neg t \cap a)P(\neg t) = 0.01 \times 0.1 = 0.001$$

Objective vs subjective

Prout's Case Example:

Invoking the Bayes

$$P(h|e) = P(h)\frac{P(e|h)}{P(e)} = 0.9 \times \frac{0.008}{0.0082} = 0.878$$

$$P(a|e) = P(a)\frac{P(e|a)}{P(e)} = 0.6 \times \frac{0.001}{0.0082} = 0.073$$

Objective vs subjective

1. Duhem-Quine Problem:

Bayesian approach can see how different assumptions/hypotheses are updated by new evidence.

In the Prout case, Prout's hypothesis can be kept while the falsification is put to the measurement process.

Objective vs subjective

2. On Ad hoc hypothesis

Recall: *"a modification in a theory, such as the addition of an extra postulate or a change in some existing postulate, that has no testable consequences that were not already testable consequences of the unmodified theory"*

In the sense of Bayesian:

Let a theory t is refuted by fact e. Then additional explanation a is given so that $t \cap a$ implies e

Objective vs subjective

2. On Ad hoc hypothesis

Adhoc-ness has at least two types of criterion:

- It possess no independent test implications
- It does have such implications, but none has been verified.

Ad hoc hypothesis is sometimes difficult to rule out because it is too weak to have independent test. It also admits hypothesis in a way that at least clashes with our intuitions

Objective vs subjective

2. On Ad hoc hypothesis

From Bayesian PoV, a theory can be scientific even if it is ad hoc (which tends to be regarded otherwise). Accceptability of ad hoc hypothesis is evaluated by the value its probability.

Suppose a theory t is modified by adding ad hoc hypothesis a. It is straightforward that $P(t \cap a) \leq P(a)$.

The modified theory, $t \cap a$, will gives low probability simply on the grounds that $P(a)$ is unlikely. In fact, there is no restriction to value of $P(t \cap a|e)$.

Objective vs subjective

Recall, in falsificationism, examples of new advance:

Bold Conjecture
(Low prior):

General Theory of Relativity

Confirmed

Eddington's experiment during eclipse that shows bending of star light

Cautious Conjecture
(High prior):

Naïve set theory is consistent axioms

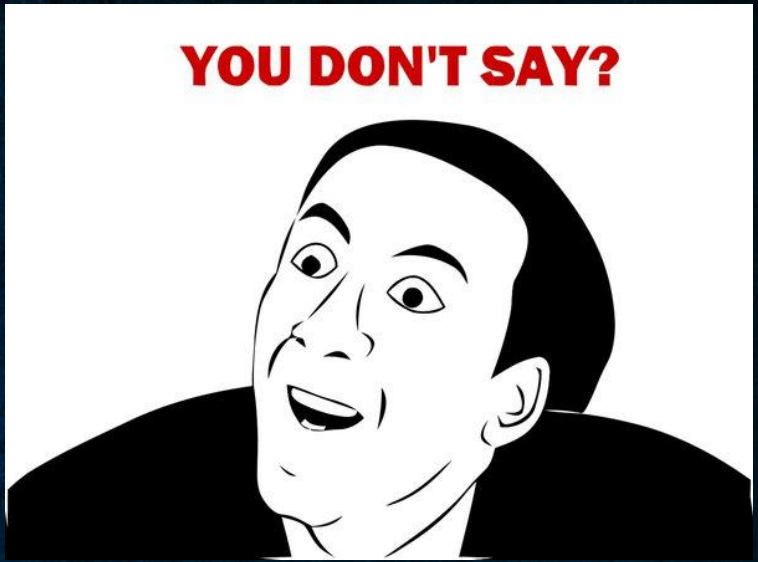
Falsified

Russel's paradox about possibility of self-referenced set

In context of falsificationism

Low prior will directly give low posterior if falsified. So, it is unsignificant

Falsified



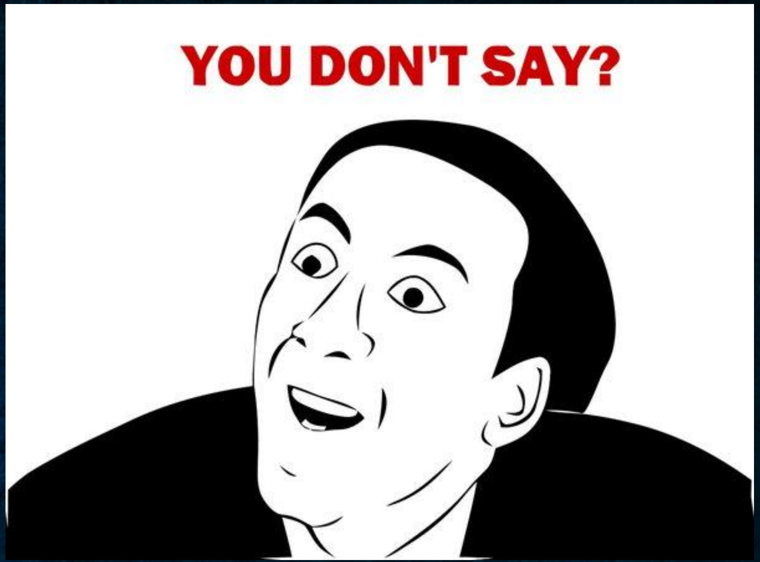
Bold conjecture:

Turtle Earth Theory

In context of falsificationism

High prior, but with repeated evidence, gives unsignificant update

Confirmed



Cautious Conjecture:

Paper burnt

In context of falsificationism

6.

Is there no issue with Bayesian?

1. Disagreements only reflect various belief of scientists

It follows that any disagreements in science, between proponents of rival research programs, paradigms or whatever, reflected in the (posterior) beliefs of scientists, must have their source in the prior probabilities held by the scientists. But the priors are themselves totally subjective and not subject to a critical analysis. They simply reflect the various degrees of belief each individual scientist happens to have.

Consequently, those of us who raise questions about the sense in which science can be said to progress will not have our questions answered by the subjective Bayesian, unless we are satisfied with an answer that refers to the beliefs that individual scientists just happen to have started out with.

Critiques on Subjective Bayesian

2. Difficulties to access degrees of belief

One of the most important sources of information that we need to have access to in order to acquire that understanding is the degrees of belief that scientists actually do or did hold.

Two problems:

1. Problem of gaining access to a knowledge of these private degrees of belief.
2. The implausibility of the idea that we need to gain access to these private beliefs in order to grasp the sense in which, say, the wave theory of light was an improvement on its predecessor. The problem is intensified when we focus on the degree of complexity of modern science, and the extent to which it involves collaborative work.

Objective vs subjective

3. Taqlid

It would seem that, provided a scientist believes strongly enough in his or her theory to begin with (and there is nothing in subjective Bayesianism to prevent degrees of belief as strong as one might wish), then this belief cannot be shaken by any evidence to the contrary, however strong or extensive it might be.

Objective vs subjective

The Bayesian theory we are proposing is a theory of inference from data; we say nothing about whether it is correct to accept the data or even whether your commitment to the data is absolute. It may not be, and you may be foolish to repose in it the confidence you actually do.

The Bayesian theory of support is a theory of how the acceptance as true of some evidential statement affects your belief in some hypothesis. How you come to accept the truth of the evidence and whether you are correct in accepting it as true are matters which, from the point of view of the theory, are simply irrelevant.

Remarks

- Alan Chalmers frequently cited Howson and Urbach in this chapter, thus primary source is also used in this presentation

BAYESIAN VERSUS NON-BAYESIAN APPROACHES TO CONFIRMATION*

Colin Howson and Peter Urbach

Remarks

- Bayesian statistics is practical only if used to many theories that use quantified proportions/probability, e.g. medical theory. In the case of general Science, it is difficult to put exact value of “the degree of belief”.
- It is an alternative way to explain inferential testing without bias (cancer test “paradox”)

To be continued next month on the next section:

Ch.14: The New Experimentalism

Thank you