A Comment on Jon Williamson's "Inductive Influence"

Stephan Hartmann
London School of Economics

Bayesianism

- Bayesianism comes in many forms, and the variant of *Objective Bayesianism* that Jon defends is one of them.
- Objective Bayesianism is not w/o problems though. Jon proposes a solution to one of these problems - the problem of learning. I'll give a critical assessment of his solution.
- I'll also make some more general remarks about the prospects of Objective Bayesianism.

Overview

I. Jon's Objective Bayesianism
II. The Problem of Learning and Jon's Solution
III. Problems with Jon's Solution
IV. Objective Bayesianism Reconsidered

I. Jon's Objective Bayesianism

Objective Bayesianism

- While subjective Bayesians leave the priors completely unconstraint, objective Bayesians think that they should be (uniquely) fixed.
- To do so, Jon invokes the Principle of Maximum Entropy (PME):

  (i) PME makes sure that our degrees of believes are as far away as possible from the extremes 0 and 1 (satisfying constraints by “empirical knowledge” or “background knowledge”).

  (ii) PME is a *logical* constraint.

  (iii) PME implies the *Principle of Indifference*.

Qualitative constraints

- In order to deal with qualitative constraints, Jon employs *objective Bayesian nets* that allow the modeling of inductive influences and go with an efficient algorithm for assigning the probabilities in the net. This algorithm solves a major problem of earlier accounts that employ PME and partly rebuts Pearl’s criticism thereof.
- I will not repeat how the technicalities work.

II. The Problem of Learning and Jon’s Solution

The Problem of Learning

- If the agent has no background knowledge, PME yields

  $p(b_n^1) = p(b_n^0) = p(b_{101}^1 | b_1^1 \ldots b_{100}^1) = 1/2$.
- This is counter-intuitive as we do seem to learn - at least sometimes.
- Solution: Introduce *dependencies* between the variables enforced by the background knowledge.

Dependencies (1)

- Where do the dependencies come from?
- Jon: “[T]he variables are all instantiations of the same predicate.”
- Is this plausible? Consider this sequence:

  B(this raven), B(Christian’s T-shirt), B(my pen),… - They all instantiate the same predicate, but there is hardly any dependence, and the next thing I observe (e.g. an elephant) might not be black.

Dependencies (2)

- I find it much more plausible to ground the dependencies in the raven example on the overlap in the gene pools of ravens which causes the color of ravens.

The main equation

- Assuming that

  (i) $p_\varepsilon \geq p_{\varepsilon'} + \tau_n$

  with $p_\varepsilon =_{df} p(b_{n+1}^1 | b_1^{\varepsilon 1}, \ldots, b_1^{\varepsilon n})$ and the influence threshold $\tau_n \neq 0$ and

  (ii) an *influence graph* in which B_{n+1} has as its parents all previous variables $B_1, \ldots, B_n$, PME gives us:

  (E) $p_\varepsilon = 1/2 + n\,\tau_n\,(r_n/n - 1/2)$

  with the number of positive instances r_n.
- This equation also shows that τ_n has to be $\neq 0$ to allow for learning. (E) has important consequences.

Claim 1

- If we know the frequency

  $freq_{F,\,a}(F(a)) = x$,

  then all priors $p(b_n^1) = x$, $\tau_n = 1/n$ (and hence all $\lambda_n = 0$) and $p_\varepsilon = r_n/n$.

  Hence, we should set our degrees of belief according to the observed frequencies r_n/n.

Claim 2

- If we do not know the frequency (and if we do not have any knowledge about how the B_n are produced), then we should set our degrees of belief according to the observed frequencies (in the short run as well as in the long run).

III. Problems with Jon's Solution

Claim 2 is false!

- To prove claim 2, Jon assumes that (i) the (troublesome) Principle of Indifference tells us that all priors $p(b_k^1)$ are equal and (ii) that they are equal to x, the frequency!
- But we do not know the frequency (assumption!) and so we have to set $p(b_k^1) = y$ (with arbitrary y, PME does not help us here!). From this, we can only conclude, using (E), that $k\ \tau_k = m\ \tau_m$ and hence $\tau_n = c/n$ with $0 \leq c \leq 1$. *PME does not fix c.* For λ_n we get $\lambda_n = (1-c)/c \bullet n$.

What about claim 1?

- Claim 1 only obtains if $\tau_n \neq 0$. An inductive skeptic will not be willing to make this assumption. For her, Jon's procedure is question begging as it is based on the assumption that there are dependencies.
- Is claim 1 plausible? No! Recall, for example, Popper's sun-rising example (as quoted in Gillies (2000, 72f)).

VI. Objective Bayesianism Reconsidered

Should we all assign the same priors? (1)

- According to Objective Bayesianism, we should assign the same priors once we agreed on the background knowledge. Why should we?
- One motivation for having a rule that helps us to fix priors comes from *machine learning*. To run the programs, priors have to be specified before the machinery can start running.
- But does it always make sense to fix the priors uniquely? Do we (human beings) do so and should we do so?

Should we all assign the same priors? (2)

- The answer to the *descriptive question* is probably no.
- So what about the *normative question*? I do not see why we should assign the same priors. And why should our priors be as non-committal as possible?
- It is correct that the evidence does not support more, but, as we are ultimately interested in good posteriors, assigning priors should be *relative to a certain goal* (e.g. high speed of convergence) and I do not see how such a claim can be established using PME.

How objective is Objective Bayesianism?

- Jon does not distinguish between background knowledge and a probability model. Indeed, we always have to make assumptions about the probabilistic model (e.g. which variables should be included) and these assumptions might not all be fixed by background knowledge. Two agents might agree on the background knowledge, but not on the probability model. And so they'll assign different priors using PME. Probabilities are relative to a probability model, which is bad news for the objectivist.

How objective… ? (cont'd)

- Note that it might also make sense to change our probability model once observations come in. (Cf. again Popper's example).
- Jon seems to agree with this, but I doubt that PME can be of any help here.

Three final questions

1. What do the probabilities mean? Are they subjective degrees of belief, or rather logical probabilities? (Remember that Jon takes PME to be a logical constraint.)
2. How can more realistic cases of evidence acquisition be modeled? What about testimony?
3. Shouldn't we also condition on background theories? How can this be modeled?