Null, Alternative, and Informative Hypotheses
Utrecht 2007

A Philosophical Analysis of Inequality-constrained Models

Jan-Willem Romeijn,
Philosophy, Groningen University

Rens van de Schoot,
Psychology, Utrecht University

Contents

1 Bayes factors for models

To choose between a model $\mathcal{M}_1$ with inequality constraints and its unconstrained rival $\mathcal{M}_0$, we may compute the Bayes factor:

$$BF_{01} = \frac{p(\mathcal{M}_1|E)}{p(\mathcal{M}_0|E)}$$

Depending on whether $BF_{01} > 1$ or $BF_{01} < 1$, we choose the constrained or unconstrained model respectively.

Example
An example of this use of Bayes factors is the discussion by Boelen and Hoijtink of a study of dissociative identity disorder (DID):

Constrained model For DID-patients, the probability of recollection in a set memory task is strictly larger than that of DID-simulators.

Unconstrained model There is no constraint on the relation between probabilities of recollection in DID-patients and DID-simulators.

Central questions

In this talk I discuss the following questions about this use of Bayes factors:

- What kind of support or confirmation is provided by a Bayes factor?
- How does the use of Bayes factors fit with general scientific methodology?
- Can we treat Bayes-factors for models in the same way as for statistical hypotheses?

2 The problem of induction

The sun has risen on every morning of my life up until now. Will it rise tomorrow? Can we say it will always rise?

Inductive inference

Inductive inference runs from the given data to predictions of future data and empirical generalizations. How can we justify this mode of inference?

- A justification based on the past success of induction is itself based on inductive inference.
- The assumption that the world is uniform, as such, is too general.
- The problem emerges again as the problem that we cannot justify our choice of the properties in which the world is uniform.

Falsificationism

The philosopher Popper reacted to this problem by a total rejection of inductively derived claims.

Only negative claims about the empirical world can be justified. For example, on the basis of a single black swan we can conclude that “All swans are white” is false.

Inductive logic

The philosopher Carnap attempted to justify inductive inference by means of probability theory.

$$\frac{\text{All the swans until now were white}}{\text{All swans are white}}$$

$$\downarrow$$

$$\frac{\text{All the swans until now were white}}{\text{The probability of ``All swans are white'' is large.}}$$

By fixing the language of the data set, and by applying the appropriate symmetry requirements to the probability assignments over it, Carnap derived justified inductive predictions.

3 Bayesian inductive logic

Statistical inference can be viewed as an answer to the problem of induction. Under the assumption of a statistical model, it takes us from data to probabilistic predictions and generalisations.

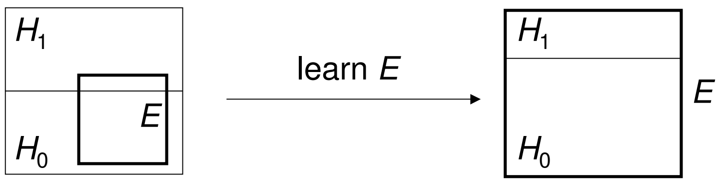
Bayesian inductive logic provides a view on statistical inference in between Popper and Carnap.

Bayesian logic

Bayes' theorem determines deductively how to combine evidence E with an assumed model:

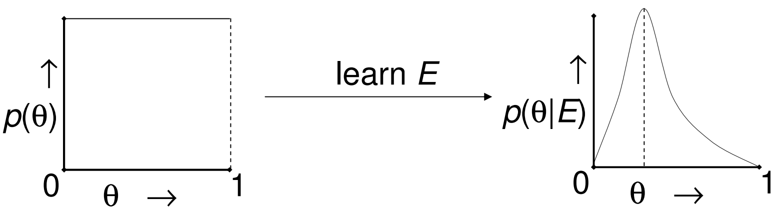
$$p(H_j|E) = p(H_j)\frac{p(E|H_j)}{p(E)}.$$

This can be represented conveniently in a diagram. The probabilities are expressed by the size of the sets.



Models as premisses

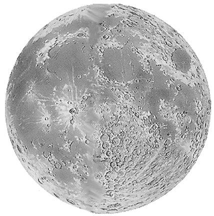
What we learn from the data, is determined not just be Bayes' theorem, but also by the specific models that we choose.



For example, if we choose a model Θ with multinomial distributions, we assume that the observations are independent and identically distributed.

4 Model selection

Assume that if the moon is made of blue cheese, then Brigitte Bardot is a man. Now assume the moon is made of blue cheese. We derive that Brigitte Bardot is a man.




But Brigitte Bardot is not a man. Hence something must be wrong with the premisses.

Goodness of fit and other criteria

The predictive performance of a model is a measure of how well it accords with the world. If the performance is low, we have reason to discard the model. But there are other further criteria as well:

Simplicity More complicated models will have a better fit but risk overfitting, and are therefore penalized.

Causality We may prefer a model because it accords better with the causal story we want to tell about the system.

Informativeness Sometimes a model fits less well but provides answers where other models remain silent.

The methodological status of Bayesian statistics

Bayesian statistics unites the inductivist and falsificationist elements of Carnap and Popper respectively.

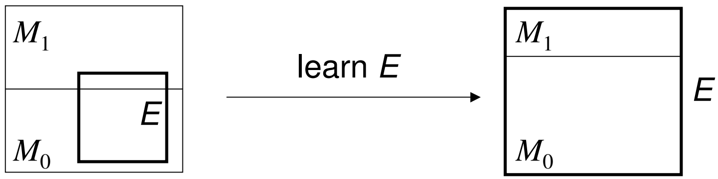
- It is inductivist in the sense that it allows for probabilistic support of hypotheses by the data.
- It is falsificationist in the sense that it is tested against the data on its performance.

In unifying these aspects, it may be viewed as an improved version of the hypothetico-deductivist method of Hempel.

5 Extending Bayesian logic

Hoijtink and co-workers use Bayesian logic on the level of checking the premisses.

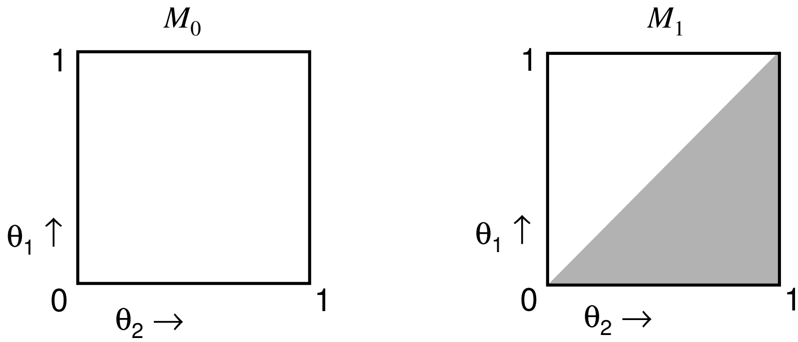
$$BF_{01} = \frac{p(\mathcal{M}_1|E)}{p(\mathcal{M}_0|E)} = \frac{p(E|\mathcal{M}_1)p(\mathcal{M}_1)}{p(E|\mathcal{M}_0)p(\mathcal{M}_0)} = \frac{p(E|\mathcal{M}_1)}{p(E|\mathcal{M}_0)}.$$



Models with differing inequality constraints are treated as if they were statistical hypotheses.

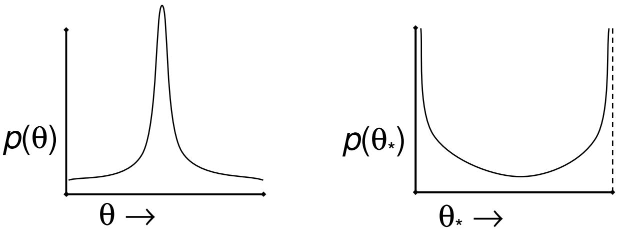
Testing the scope of a model

In the case of two parameters and a simple inequality constraint, we are testing whether the inclusion of a certain range of hypotheses in the model leads to a better predictive performance.



Testing the priors?

We may even allow the same range of hypotheses in both models, and simply test which of two priors fits the data better!



Solely in virtue of the differing priors, the two models Θ and Θ^* have different predictive properties, that is, different likelihood functions.

6 Models as hypotheses?

Intuitively, it makes sense to test hypotheses against the data because ultimately these statistical hypotheses have different empirical content.

$$\theta = 1/2 \;\rightarrow\; E = 101001\ldots \qquad \theta = 1/6 \;\rightarrow\; E = 000100\ldots$$

But does it make sense to extend Bayesian logic towards model selection?

Convergence as criterion

One possible criterion for deciding whether it makes sense to test two models against each other is that for at least some sequences of evidence E, the Bayes factor must tend to 0 or ∞.

- For models of which the included statistical hypotheses do not completely overlap there are such evidence statements.
- For models that only differ in terms of their non-zero priors, there are no such sequences of evidence sentences.

On the other hand, the latter fact may be a reason to drop the relation between probability and empirical content.