Dpto. de Logica y Filosofia de la Ciencia, Valencia 2007
LECTURE 2 "BAYESIAN STATISTICAL INFERENCE"

Jan-Willem Romeijn, University of Groningen
Email: j.w.romeijn@rug.nl
Web: http://www.philos.rug.nl/~romeyn

1. Bayes' theorem and inverse probability
Yesterday I have introduced Bayesian inference as a means to derive new probability valuations from a set of given valuations. Today we apply this idea to inductive inference.

The real genius of Bayes shows in the inductive application of his theorem. Up to the time of Bayes, probability theory was only used to derive the probability of events from a known cause, such as a game of chance. By contrast, Bayes used his theorem to derive probabilities for possible causes, e.g. a range of possible games of chance, from an observed series of events. He invented so-called inverse probability.

One might argue that the application of Bayes' theorem to scientific confirmation provides an answer to the Humean problem of induction. But to assess this view, we first look at a slightly different answer.

2. Carnapian inductive logic
For repeated observations Laplace, and later Reichebach, proposed the so-called rule of succession, which allows observers to compute the probability of a subsequent observation on the basis of a past series of observations. Such rules have been studied extensively by Carnap and his followers.

Carnap employs an observational language, associated with the observation algebra introduced earlier. The language is an expression of all salient distinctions. Over the language we can therefore distribute the probability symmetrically. The completely symmetric distribution leads to the exchangeable inductive rule c^*. This generalises to a continuum of inductive methods including the parameters λ and γ.

The inductive rules of Carnap may indeed be viewed as a solution to Hume's problem of induction. But as Goodman's new riddle reveals, the solution works on the assumption that the language employs so-called projectable predicates. Furthermore, general hypotheses cannot be accommodated in the inductive rules. Finally, the rules seem to connect rather badly with statistics as used in the sciences.

3. A Bayesian representation of Carnapian inductive logic
To get a Bayesian grip on Carnapian inductive rules, it is convenient to define an algebra for repeated observations. The relevant algebra here is a so-called cylindrical algebra.

Once repeated observations are captured in terms of an algebra of sets, we may apply the Kolmogorovian axioms to the probabilities over repeated observations.The Carnapian inductive rules can then be understood as specific probability assignments. The predictions can be derived from these assignments by applications of Bayes' rule, but the probability assignment itself must be assumed at the onset.

Thus, in a Bayesian logic, the derivation of the predictions is uncontroversial. But the probability assignment over the language shows up as an explicit premise.

4. A frequentist semantics for statistical hypotheses
Recall that Bayes' original idea on induction involved reasoning about probabilistic setups. It turns out that there is an intimate relation between that original idea, and the Carnapian rules introduced above. To present this relation, it is convenient to elaborate on the concept of a statistical hypothesis first, making use of the frequentist interpretation of probability by von Mises.

Von Mises' theory centers around the notion of a 'Kollektiv': an infinitely long sequence of observations with specific limiting relative frequencies of the possible outcomes, which is otherwise completely random and therefore does not show any other kind of pattern or periodicity. This latter requirement is conveniently expressed by an assumption known as the 'law of excluded gambling systems'.

In the theory of von Mises, probability is expressed as a property of 'Kollektiv's': the probability of a result is defined by the limiting relative frequency of that result in the Kollektiv. In my thesis I argue that we may identify a statistical hypothesis with the set of all the 'Kollektiv's' with the corresponding probabilities in the cylindrical algebra.

The driving force behind the frequentist interpretation of probability is logical empiricism. The obvious problem with this view is that it does not allow for single-case probabilities. Also, the theory employs the rather unempirical notions of randomness and infinite sequence. But such problems do not hamper the use of frequentism in the formal semantics of Bayesian logic.

5. Bayesianism and prior probabilities

Bayesian statistical inferences take as input a statistical model, comprising of a collection of statistical hypotheses, and a prior probability over these hypotheses. Observations determine the likelihoods of the hypotheses, and together with the prior we can then compute the posterior probability over the hypotheses.

The prior is a subjective input component to Bayesian statistical inference. This subjectivity is sometimes taken to be at variance with the objectivity strived for in science. However, the prior also offers an opportunity for encoding background knowledge that is not captured in the statistical model. Moreover, the influence of the prior diminishes with a sufficiently large number of observations.

In Jeffreys' view of the Bayesian method, and against subjectivist views, the prior probability over hypotheses is completely objective. It is determined not by personal preference, but by an objective ranking in terms of simplicity. The general gist of this has recently become popular again as objective Bayesianism. It is intimately connected to formal methods of model selection.

6. The representation theorem

The Bayesian statistical inferences sketched above also lead to predictions on next observations, like those generated by Carnapian prediction rules. A special class of prediction rules concerns those predictions that are invariant under the permutation of past observations, the so-called exchangeable prediction rules.

An important link between Bayesian inferences over statistical hypotheses and exchangeable prediction rules is provided by De Finetti's representation theorem: every exchangeable rule can be represented uniquely by a prior probability density over Bernoulli hypotheses in a Bayesian inference. As a special case, the family of Dirichlet distributions coincides exactly with the Carnapian continuum of rules.

De Finetti argued that we can therefore avoid using statistical hypotheses altogether, and make do with exchangeable prediction rules and generalisations of them without loss of generality. Strictly speaking this is correct, but I want to argue there are conceptual advantages to using the hypotheses after all.

7. A Bayesian solution to Hume's problem?

We saw that the Bayesian representation of Carnapian logic was only a partial solution to the problem of induction: the derivation of predictions is justified by the logic, but the probability assignment has to be assumed. The situation for the Bayesian inference is not much different: it provides a solution to the logical problem of induction but it suggests nothing towards solving the epistemic problem of induction.

Still the Bayesian statistical inference has some conceptual advantages over the Carnapian solution. Its methods may be viewed as logical inferences from a statistical model and a prior. The model, moreover, can be understood as a projectability assumption, on a par with Carnap's choice for an observation language. Finally, Bayesian statistics has a place both for physical and epistemic probability.

These conceptual advantages make Bayesian statistics an attractive alternative to classical statistics. Moreover, due to the invention of the computer Bayesian methods are also computationally feasible. But the problem that we must choose our priors remains.