

Jan-Willem Romeijn
University of Groningen
j.w.romeijn@rug.nl
<http://www.philos.rug.nl/~romeijn>

1. Induction and probability

The first mention of probability in its modern meaning is found in a correspondence between Pascal and Fermat concerning a game of chance. This discussion concerns the estimation of the probability of events given a fixed chance setup. The probabilities are thus used to describe events in the world.

After being converted to the Catholic sect of Jansenists, Pascal devised an influential argument for believing in God, known as Pascal’s wager. Among other things the wager involves the probability that God exists. The thing to note is that in this case, probability is used to express rational beliefs.

The dual nature of probability facilitates its application to the problem of induction. Probability might be used to give a formal, and hence independently grounded account of the alignment of probability as rational belief with the probabilities ‘out there’.

2. Axiomatisation of probability

It is useful to specify the axiomatisation of probability due to Kolmogorov. In his treatment, probability is a measure of sets A , B , etc.

- $p(A) \geq 0$
- $p(\Omega) = 1$
- $p(A \cup B) = p(A) + p(B)$ if $A \cap B = \emptyset$.

We can conveniently represent the sets by means of Venn diagrams. Their areas are a natural measure, and thus represent the probabilities assigned to the sets. The areas nicely illustrate Kolmogorov’s axioms.

A collection of sets forms a so-called algebra if it is closed under a number of set theoretical operations. For Kolmogorov a probability measure is defined on such an algebra. For the applications we will consider, it is useful to think of sets as collections of possible worlds. Each set is characterised by a proposition that is true in exactly those possible worlds belonging to the set.

In this way we can associate sets with propositions, and thus assign probabilities to propositions. The beauty of Kolmogorov’s axiomatisation is that it is just a formal system. It does not suggest anything towards an interpretation of the probability measure.

3. Probabilistic logic

Following the work of De Finetti, Howson, and many others, we can view the theory of probability itself as a logic. There is a certain similarity between probability as a function over an algebra, and truth values as a function over a language. In this view Kolmogorov's axioms determine what probability assignments are consistent.

In the set-theoretical formulation of Kolmogorov it is rather easy to derive the theorem that Thomas Bayes painstakingly derived 250 years ago: $P(A | B) = P(A) P(B | A) / P(B)$. Inferences that employ this consistency criterium on subjective probabilities are often called Bayesian.

Probabilistic, or Bayesian logic dictates the probability values that must be assigned to specific propositions on the basis of certain values for other propositions. There are numerous philosophical applications of this idea, and often Bayes' theorem plays a central role. The Monty Hall dilemma is a case in point.

4. Induction and Bayesian logic

Our present interest is in the application of probabilistic inference to induction. Induction is a mode of inference that brings us from data to general conclusions on the system or mechanism from which the data is obtained.

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.

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 from the prior we can then compute the posterior probability over the hypotheses.

The prior is an epistemic, though not necessarily a subjective input component to Bayesian statistical inference. It is part and parcel of Bayesian or inverse probability that the two interpretations of probability, physical and epistemic, coexist. Bayes' theorem thus tells us how opinion can be aligned to physical probability.

5. A frequentist semantics for statistical hypotheses

There is still something peculiar about the use of statistical hypotheses as arguments of a probability function. Let me clarify the concept of a statistical hypothesis, making use of some ideas first developed by Gaifman and Snir and going back to the frequentist theory of 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. Though somewhat covertly, Gaifman and Snir argue that we may identify a statistical hypothesis with the set of all the 'Kollektiv's' with the corresponding probabilities in the cylindrical algebra.

6. Carnapian inductive logic

One might argue that Bayes thereby provides an answer to the Humean problem of induction. But to assess this view, we first look at a slightly different answer: the straight rule of Laplace and Reichebach. 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 evenly, respecting a distinct set of symmetries. This symmetric distribution leads to a continuum of inductive methods.

The inductive rules of Carnap may indeed be viewed as a solution to Hume's problem of induction, based entirely on the choice of a particular observation language. But general hypotheses cannot be accommodated in the inductive rules. And more importantly, as Goodman's new riddle reveals, the solution works on the assumption that the language employs so-called projectable predicates.

For present purposes, there is another way of saying why Carnapian inductive logic falls short of providing a solution. From the point of view of probabilistic logic, the Carnapian rules can be understood as specific probability assignments. The predictions can be derived from a given assignment, but the assignment itself must be assumed at the outset.

7. 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.

8. A Bayesian solution to Hume's problem?

We saw that Carnapian logic was only a partial solution to the problem of induction: the derivation of predictions is justified by probabilistic logic, but the probability assignment itself has to be assumed. Carnap found a motivation for constraints on the probability assignment in the observation language.

The situation for the Bayesian inductive logic 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. We must assume a probability assignment at the outset. The model and prior that fix this probability assignment and can both be understood as part of a projectability assumption, on a par with Carnap's choice for an observation language and specific symmetries.

Still the Bayesian representation of inductive inference has some conceptual advantages over the Carnapian inference rules. The model offers a particular handle on the projectability assumption. Bayesian inferences thereby separate the physical and epistemic components of the inductive assumptions, thus providing a better grip on the epistemic problem of induction.