The Problem of Induction: Handout

Mathias Winther Madsen

January 11, 2015

Bertrand Russell on induction:

> It is obvious that if we are asked why we believe that the sun will rise to-morrow, we shall naturally answer 'Because it always has risen every day.' We have a firm belief that it will rise in the future, because it has risen in the past.
>
> ...
>
> But the real question is: Do any number of cases of a law being fulfilled in the past afford evidence that it will be fulfilled in the future? If not, it becomes plain that we have no ground whatever for expecting the sun to rise to-morrow, or for expecting the bread we shall eat at our next meal not to poison us, or for any of the other scarcely conscious expectations that control our daily lives. It is to be observed that all such expectations are only probable; thus we have not to seek for a proof that they must be fulfilled, but only for some reason in favour of the view that they are likely to be fulfilled.
>
> ...
>
> We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken.
>
> (*Problems of Philosophy*, 1912, Chapter VI)

Nelson Goodman on "lawlike" predictions:

> Now let me introduce another predicate less familiar than "green." It is the predicate "grue" and it applies to all things examined before t just in case they are green but to other things just in case they are blue. Then at time t we have, for each evidence statement asserting that a given emerald is green, a parallel evidence statement asserting that that emerald is grue. And the statements that emerald a is grue, that emerald b is grue, and so on, will each confirm the general hypothesis that all emeralds are grue.
>
> ...
>
> Thus although we are well aware which of the two incompatible predictions is genuinely confirmed, they are equally well confirmed according to our present definition. Moreover, it is clear that if we simply choose an appropriate predicate, then on the basis of these same observations we shall have equal confirmation, by our definition, for any prediction whatever about other emeralds-or indeed about anything else.

("The New Riddle of Induction," 1955)

Karl Popper on falsification:

> Psychoanalysis is a very different case. It is an interesting psychological metaphysics ... but it never was a science. There may be lots of people who are Freudian or Adlerian cases ... But what prevents their theories from being scientific in the sense here described is, very simply, that they do not exclude any physically possible human behaviour. Whatever anybody may do is, in principle, explicable in Freudian or Adlerian terms.
>
> ...
>
> The point is very clear. Neither Freud nor Adler excludes any particular person's acting in any particular way, whatever the outward circumstances. Whether a man sacrificed his life to rescue a drowning, child (a case of sublimation) or whether he murdered the child by drowning him (a case of repression) could not possibly be predicted or excluded by Freud's theory; *the theory was compatible with everything that could happen* [...]

("The Problem of Demarcation," 1974; emphasis in original)

Jakob Bernoulli on probability estimation:

> Let the number of fertile cases and the number of sterile cases have exactly or approximately the ratio $\frac{r}{s}$, and let the number of fertile cases to all the cases be in the ratio $\frac{r}{r+s}$ or $\frac{r}{t}$, which ratio is bounded by the limits $\frac{r+1}{t}$ and $\frac{r-1}{t}$. It is to be shown that so many experiments can be taken that it becomes any given number of times (say c times) more likely that the number of ferile observations will fall between these bounds than outside them, that is, that the ratio of the number of fertile to the number of all the observations will have a ratio that is neither more than $\frac{r+1}{t}$ nor less than $\frac{r-1}{t}$.
>
> ...
>
> [For example, if $r/t = 3/5$, then] by what has been demonstrated, it is inferred that if $25,550$ experiments are taken, it will be more than 1000 times more likely that the ratio of the number fertile observations to the number of all the observations will fall between these bounds, $\frac{31}{50}$ and $\frac{29}{50}$, than outside them. On the same understanding, if c is set equal to $10,000$ or $100,000$, it may be seen that it will be more than ten thousand times more probable, if there are $31,258$ experiments, and more than a hundred thousand times more probable, if there are $36,966$, and so forth to infinity, continually adding to the $25,550$ another 5708 experiments. Whence at last this remarkable result is seen to follow, that if the observations of all events were continued for the whole of eternity (with the probability finally transformed into perfect certainty) then everything in the world would be observed to happen in fixed ratios and with a constant law of alternation. Thus even in the most accidental and furtuitous we would be bound to acknowledge a certain quasi-necessity and, so to speak, fatality.
>
> (*The Art of Conjecturing*, 1713, Part 4, Ch. 5)

Abraham Wald on the minimization of maximum risk:

> A decision function $\delta_0$is said to be a minimax solution of the decision problem if it minimizes the maximum of [the average loss] $r(F, \delta)$ with respect to [the probability measure] F, i.e., if
>
> $$\sup_F r(F, \delta_0) \le \sup_F r(F, \delta)$$
>
> for all δ, where the symbol $\sup_F$ stands for supremum with respect to F.

> In the general theory of decision functions, as developed in Chapter 3, much attention is given to the theory of minimax solutions for two reasons: (1) a minimax solution seems, in general, to be a reasonable solution of the decision problem when an a priori distribution in [the sample space] Ωdoes not exist or is unknown to the experimenter; (2) the theory of minimax solutions plays an important role in deriving the basic results concerning complete classes of decision functions [i.e., classes whose expansion lead to no improvements].
>
> There is an intimate connection between minimax solutions and Bayes solutions. It will be seen in Chapter 3 that under general conditions a minimax solution is also a Bayes solution. More precisely, a minimax solution is, under some weak restrictions, a Bayes solution relative to a least favorable a prior distribution.
>
> (*Statistical Decision Functions*, 1950, Chapter 1.4.2)

Vapnik and Chervonenkis on the uniform law of large numbers:

> According to the classical Bernoulli theorem, the relative frequency of an event A in a sequence of independent trials converges (in probability) to the probability of that event. In many applications, however, the need arises to judge simultaneously the probabilities of events of an entire class S from one and the same sample. Moreover, it is required that the relative frequency of the events converge to the probability uniformly over the entire class of events S. More precisely, it is required that the probability that the maximum difference (over the class) between the relative frequency and the probability exceed a given arbitrarily small positive constant should tend to zero as the number of trials is increased indefinitely. It turns out that even in the simplest of examples this sort of uniform convergence need not hold. Therefore, one would like to have criteria on the basis of which one could judge whether there is such convergence or not.
>
> This paper first indicates sufficient conditions for such uniform convergence which do not depend on the distribution properties and furnishes an estimate for the speed of convergence. Then necessary and sufficient conditions are deduced for the relative frequency to converge uniformly to the probability. These conditions do depend on the distribution properties.
>
> ("On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities," 1968, translated 1972)