Notes on Scientific Method

Ray Solomonoff
transcribed by Jannik Michelfeit

The method we are using is essentially that of Rudolf Carnap ("Logical Foundations of Probability" University of Chicago Press, 1950), one of the founders of "Logical Positivism".

He represented each possible universe by a long string of symbols. For d symbol types and strings of length n there are just d^n possible universes to consider. To do prediction he postulated a function that assigned a probability to each sequence, and he computed probabilities of continuations of sequences just as we do. He felt that scientific laws were, strictly speaking, unnecessary - That this probability function was all that was needed for science.

He began his investigation with a very simple probability function, that could only deal with regularities found in a Bernoulli sequence but he regarded this as only the beginning of his investigation.

We now use a more complex probability function that is able to deal with any describable regularity in the symbol string.

Unfortunately, our function is usually impossible to calculate exactly, and even approximations can be very difficult. We make these approximations by finding short codes for the data. Any regularity in our data is expressible as a short code. All of our scientific laws are means for writing short codes. A fair amount of the intellectual work in science consists of finding short codes.

Scientific work is of 3 kinds:

1. Finding regularities in data (as we've just discussed).

2. Deciding what new data to get.

3. Getting the new data. (This can be passive observation or active experimentation.)

Step 3 is most often a kind of engineering problem.

Step 2 is more difficult to characterize. It is guided by regularities found in old data... Scientific laws that seem to work.

Often there will be several theories that fit past data equally well, but extrapolate differently in new areas. We can then gather new data to distinguish between the theories. If there are many among which to choose and the needed experiments are expensive in time, money and/or manpower, we can use algorithmic probability to decide which theories are most likely and help decide which experiments to do.

However, in general, deciding what new data to get is a difficult problem - it guides the direction in which our science moves. It is critical in the scientific process.

One possible basis for decision is utilitarian. The idea that our ultimate purpose is that of simply surviving in a hostile, often unpredictable environment - that many species of living creatures have been evolutionarily selected for properly directed sense of curiosity - a characteristic that turns out to have survival value.

But ultimately the question that controls the direction of our experiments is "What do we want from science?".

Open problems, conjectures

1. If we have enough time to get all the short codes for some data, we have a good idea about rate of convergence of errors in probability. When we do not have all of the short codes (which is the usual situation) what is the expected rate of convergence?

   A key might be this conjecture: If we search for codes for our data in a "honest, unbiased" manner, then the mean bit cost per symbol given by these codes on the known data will be an unbiased estimate of the mean bit cost per symbol for the future data. "Honest, unbiased", will of course have to be defined. "Unbiased estimate" means the expected value of its error is zero.

2. In science, up to now, all laws that have been proposed are of essentially fixed length - They do not grow as the data that they are to explain grows. Algorithmic probability makes it possible to have data descriptions that surpass this limitation. - e.g. Grammars for ethnic languages could increase in size as the size of the body of data grew. Conceivably this sort of thing could be used in physics as well ...

3. Another limitation in current scientific theories is that they all use total recursive functions - functions that must have a value for all possible arguments. Algorithmic probability makes it possible to consider partial recursive functions - These do not have this limitation. Can they be usefully applied in science?

4. I have mentioned that algorithmic probability can be used to give good approximate probabilities to scientific theories and conjectures. The exact details of this have yet to be worked out.

5. Levin's search method enables us to deal with the halting problem in searching for predictive codes. Since this gives it the capability to go beyond any prediction methods that have ever been tried, it looks like a promising direction of investigation. I do not know of anyone having tried it.

   To implement this, one might design a simple machine or instruction set that could assign probability values to simple sequences - Then increase the complexity of the machine to become universal.

6. It is not difficult to show that algorithmic probability satisfies the first 5 axioms of Kolmogorov ["Foundations of Probability" 1950 translation of 1933 book]. It has not been shown for the 6th axiom, however. Although I suspect it is true, I have not proved it.

7. In linear regression, algorithmic probability makes a prediction based on the weighted sum of the predictions made using all possible number of coefficients. Up to now, statisticians have always simply selected the "best" number of coefficients and made predictions based upon it.

   Are there any cases in which using more than one set of coefficients is superior?