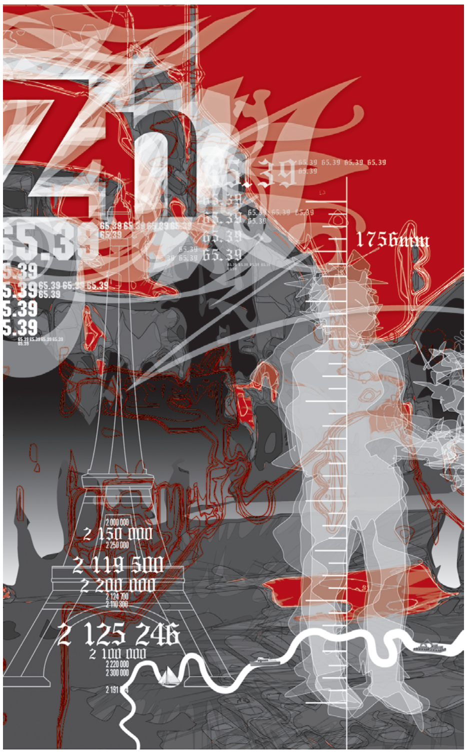
Dicing with the unknown



Charles Trevelyan

There are many things that I am uncertain about, says **Tony O'Hagan**. Some are merely unknown to me, while others are unknowable. This article is about different kinds of uncertainty, and how the distinction between them impinges on the foundations of Probability and Statistics.

Two kinds of uncertainty

There are things that I am uncertain about simply because I lack knowledge, and in principle my uncertainty might be reduced by gathering more information. Others are subject to random variability, which is unpredictable no matter how much information I might get; these are the unknowables. The two kinds of uncertainty have been debated by philosophers, who have given them the names *epistemic uncertainty* (due to lack of knowledge) and *aleatory uncertainty* (due to randomness).

Examples of aleatory uncertainty are familiar to students of probability theory, and include the outcomes of tossing dice and drawing cards from a shuffled pack. In statistics, aleatory uncertainty is present in almost all data that we obtain, due to random variability between the members of a population that we sample from, or to random measurement errors.

Examples of epistemic uncertainty are all around us. I am uncertain about the atomic weight of zinc, about the population of the city of Paris, and about whether the river Thames froze over in London during the winter of 1600–1601. At least for the first two of these, my uncertainty could be resolved by looking in a suitable reference book. It may be that there is no such source of information about the river Thames in 1600–1601, but in principle this is a question that might be resolved by historical investigation. Epistemic uncertainty about any given question varies from one person to another. For instance, I have negligible uncertainty about my height. Someone who has seen me might be able to guess reasonably accurately but could not be certain. Someone who knows only that I am a British male would have more uncertainty.

The distinction between aleatory and epistemic uncertainties is valuable in many areas where it is important to appreciate which uncertainties are potentially reducible by further investigation. But it is easy to see how much more fundamental it should be for statisticians, for whom randomness and uncertainty are their very *raison d'être*.

The theory of Statistics rests on describing uncertainties by using probability. A probability near 1 represents an event that is almost certain to occur, while a probability near 0 represents one that is almost certain *not* to occur. As we move away from these extremes towards the probability of ½, there is increasing uncertainty. Here, however, is where we meet the fundamental dichotomy between the two principal theories of Statistics: the frequentist and Bayesian theories. One characterisation of the difference between these two schools of statistical theory is that frequentists do not accept that aleatory uncertainty can be described or measured by probabilities, while Bayesians are happy to use probabilities to quantify any kind of uncertainty.

Two kinds of probability

Delving more deeply, the root of this disagreement lies in what probability means. Almost everyone who encounters probability for the first time in their education will be taught it using aleatory uncertainties, like the familiar games of tossing dice or coins. And the way probability will be taught is that it represents the long-run frequency with which the event in question occurs if it is repeated an indefinite number of times, and this is accordingly known as the *frequency* definition of probability. The nature of random events is that they are, at least conceptually, repeatable in this way. Epistemic uncertainties are generally not.

If probability is to encompass epistemic uncertainty it needs another definition.

In Bayesian statistics the probability of a proposition simply represents a *degree of belief* in the truth of that proposition. Notice in passing that we tend to use the word "proposition" (a statement that is either true or false) rather than "event" (something which may or may not occur) when discussing this kind of probability, since "event" has connotations of randomness and repeatability. A proposition might simply assert that an event will occur, but it may also refer to a statement with epistemic uncertainty.

The degree-of-belief interpretation of probability is sometimes referred to as *personal probability* or *subjective probability* because, as noted already, different people may have different degrees of uncertainty about a proposition. It is the "subjectivity" of this approach to probability that is most objected to by followers of the frequentist theory. Bayesians steadfastly defend this definition, and question the extent to which their methods are any more subjective than frequentist practice, or indeed scientific practice generally. However, that debate is beyond the scope of this article!

Two kinds of statistics

To see the implications of the frequentist position on probability, it is enough to note that uncertainty about parameters in statistical models is almost invariably epistemic. If, for instance, I was conducting experiments to measure empirically the atomic weight of zinc, the unknown parameter is that atomic weight. I cannot consider zinc as a randomly chosen element. Indeed, it is particularly zinc that I wish to know about. In a similar way, nearly all statistical analysis is to learn about parameters, and so to reduce our epistemic uncertainty about them. Since frequentist statistics does not and cannot quantify that uncertainty with probabilities, conventional statistical inferences (such as significance tests and confidence intervals) never make probability statements about parameters.

Yet the recipients of those conventional inferences almost universally interpret them as making probability statements about the parameters. When the null hypothesis is rejected with a p-value of 0.05, this is widely misunderstood as saying that there is only a 0.05 chance that the null hypothesis is true. If told that (3.2, 5.7) is a 95% confidence interval for a certain parameter, the interpretation that there is a 95% probability that the parameter lies between 3.2 and 5.7 is extremely common. It is enough to realise that our uncertainty about parameters is epistemic to appreciate that these have to be false interpretations.

The p-value of 0.05 and the confidence coefficient of 0.95 are aleatory probability statements about the *data*. The p-value says that in repeated sampling (creating an indefinite sequence of sets of data of the type being analysed) then if the null hypothesis were really true we would reject it in only 5% of those experiments. The confidence interval says that if this confidence interval were computed from each of the same infinite sequence of data sets, then 95% of those intervals will contain the true value of the parameter.

Neither of them says anything about the chance that the null hypothesis is true, or that the parameter lies in the interval, *for these data*. If we condition on the single set of data in front of us, there is no randomness in the problem, and so no frequentist probabilities can be stated.

In contrast, Bayesian inference *does* make probability statements about parameters. It can do so because the epistemic uncertainty in parameters can be quantified using the Bayesian's personal probability. Indeed, Bayesian inference describes how the acquisition of data modifies (and usually reduces) the uncertainty about a parameter, from "prior" uncertainty to "posterior" uncertainty. The Bayesian equivalent of a significance test asserts the probability that the null hypothesis is true. In the same way, the Bayesian analogue of the confidence interval (usually called a *credible interval*) has exactly the interpretation that is so often erroneously attributed to the frequentist confidence interval.

And two kinds of statistician

On a personal note, I have been both kinds of statistician in my career. It was my experience, as a young statistician, of analysing data and producing frequentist tests and confidence intervals for other scientists that convinced me that the Bayesian approach is the right one for statistical analysis. I had great difficulty persuading the scientists not to misinterpret the frequentist inferences I was giving them. And it was clear to me that this was because the correct interpretation was of no use to them. Frequentist inferences make only indirect statements about parameters, and can only be interpreted in terms of repeated sampling. Bayesian inferences directly answered the scientists' questions, making statements that were unambiguously about the parameters they wanted to learn about. Since that time (more than 30 years ago now), I have been an enthusiastic advocate and practitioner of the Bayesian approach.

Every statistician needs to understand the difference between the frequentist and Bayesian theories of statistics, and every practising statistician must (at least implicitly) choose between them. And whether something is unknown or unknowable, whether its uncertainty is due to fundamentally unpredictable randomness or to potentially resolvable lack of knowledge, turns out to lie at the heart of the debate.

Tony O'Hagan is a Professor of Statistics at the University of Sheffield. His research is in the theory and applications of Bayesian statistics. He has been involved in numerous applications, particularly in environmental statistics, asset management and health economics.

Six of one and half a dozen of the other

If I toss an ordinary coin, my probability that it will land Heads is 0.5. Suppose now that I have a bag of poker chips, and I know only that some are red and some are green. I have no reason to suppose that there are equal numbers of red and green chips. Indeed, almost certainly one colour will be more abundant than the other, but I have no idea which colour that will be, or how much more abundant it will be than the other colour. If one chip is to be pulled out of the bag my probability that it will be red is 0.5.

Now surely my uncertainties in the coin toss and the poker chip draw are different—the coin toss being very familiar and the bag of poker chips full of uncertainty—so why do I give both events the same probability? It is true that the uncertainties are different. The uncertainty about the coin toss is purely aleatory, whereas there is clearly epistemic uncertainty about the make-up of the bag of chips. Nevertheless, for a single coin toss and a single poker chip all the uncertainty is quantified in a single probability, that of Heads or red.

The difference emerges when I consider a sequence of tosses of that coin, and a sequence of chips drawn from the bag. My uncertainty about the coin tosses is still purely aleatory. No matter how many times I toss the coin, my uncertainty about getting Heads on the next toss is the same, and is expressed by a probability of 0.5. On the other hand, as I draw chips from the bag my epistemic uncertainty about its composition reduces, and my probability for the next chip being red changes according to the chips I have now seen.

The epistemic uncertainty lies in the proportion of red chips that I will see if I continue to pull chips from the bag until they are all removed. That proportion could be anything between 0 and 1. This is my unknown parameter, and it is this that I learn about as chips are drawn from the bag. For the coin, though, as I keep tossing it I know that the proportion of Heads will converge to 0.5. There is no epistemic uncertainty in the coin tossing, and no unknown parameter to learn about.

The whole purpose of Statistics is to learn from data, so there is epistemic uncertainty in all statistical problems. The uncertainty in the data themselves is both aleatory, because they are subject to random sampling or observation errors, and epistemic, because there are always unknown parameters to learn about.