Uncertainty*

Will Lowe

Social scientists are doubly concerned with uncertainty: Formal research methods are designed to quantify the uncertainty that arises when building theories of an imperfectly predictable social world. But much of the variability in human behavior that demands these methods is generated by the judgments that individual society members make in the light of their own uncertainty.

Rational expectations theories in economics are a rare case where the two concerns are explicitly related: the price of a stock, a society-level phenomenon, is predicted to evolve randomly over time because all the available systematic information will have been used by individual investors to minimize their uncertainty about its value. Any remaining price movement reflects residual investor uncertainty.

Probability provides the basis for addressing both concerns with uncertainty: Most economic and political theory assumes that probabilistic inference is a reasonable behavioral model of individual decision making in an uncertain world. And uncertainty about particular social scientific theories is also expressed using probabilities, and the associated machinery of statistics. But probability is not the only possible choice for quantifying uncertainty, so it is worth considering why it is a good one.

There are several arguments motivating individuals to use probability theory for managing their uncertainty. Dutch Book arguments show that individuals who act on their uncertainty in a way that is inconsistent with probability theory can, in gambling situations, always be fleeced (Ramsey, 1960). A more general motivation for probability was provided by Richard Cox (1946). Cox showed that for any numerical measure of certainty, plausibility, or confidence in a proposition A that we might invent, if it has the following properties then it must be a simple rescaling of the probability that A is true. The properties, expressed in terms of plausibility, are

1. transitivity; if A is more plausible than B, and B is more plausible than C, then A is more plausible that C.

2. the plausibility of A is a function of the plausibility of not-A. (For example when the more plausible A is, the less plausible not-A is.)

3. the plausibility of A and B together is a function of the plausibility of A, and the plausibility of B when A is certain.

*This is a penultimate draft of a shorter piece written for the Sage Encyclopedia of Research Methods in the Social Sciences

If these properties hold of the measure, then the number we assign to A will be a constant multiple of the probability of A. If principles 1–3 are accepted as minimum requirements then probability theory is the correct way to quantify uncertainty. Since Cox's argument does not depend on whether the uncertainty being quantified is personal or social scientific, it provides a general motivation for quantifying uncertainty with probability. Despite the wide application of rational expectations theories in economics, most psychologists doubt that probability theory is a good descriptive model of individual reasoning under uncertainty (Tversky and Kahneman, 1974). For example, people are overconfident relative to available data, and partially neglect the prior probability of events when they make predictions (base-rate neglect). These psychological facts are not necessarily problematic for rational expectations theories since individual expectations need only be consistent with probability theory on average. Nor do they prevent probability-based theories of personal inference; people may be performing probabilistically correct inference on an incorrect internal model or be taking into account causal constraints (Cheng, 1997). People also seem to process information in a way that trades some of the advantages of probabilistically correct inference for speed, by using an incorrect but less computationally demanding inference process that is only approximately correct (Gigerenzer and Goldstein, 1996).

The question of whether scientific uncertainty should be quantified in the same way as personal uncertainty defines the debate in statistics between Frequentists and Bayesians. We focus this question with two interpretations of a familiar example.

A linear regression model of the relationship between some quantity Y and two potentially explanatory variables X_1 and X_2 can be written as

$$\begin{aligned} Y &= X_1\beta_1 + X_2\beta_2 + \epsilon \\ \epsilon &\sim \text{Normal}(0, \sigma^2) \end{aligned}$$

In a Bayesian treatment of this model, parameter values are uncertain and this uncertainty is described by a Prior probability distribution, $p(\beta_1, \beta_2, \sigma^2)$. But even if the parameters were known with certainty, the outcome Y would be uncertain. This uncertainty is factored into a deterministic part, the additive relation $X_1\beta_1 + X_2\beta_2$, and a stochastic part, by assuming a particular probability distribution for the random variable ϵ. Uncertainty about Y can be partly reduced by observing values of X_1 and X_2, down to the minimum level determined by ϵ. ϵ may represent the work of a truly random social process or our belief that there are variables relevant to Y not included in the model. The equation defines a distribution over outcomes $p(Y \mid \beta_1, \beta_2, \sigma^2)$ that is conditional on particular values of the parameters (and explanatory variables, not shown here). Bayes theorem combines these two distributions into a posterior probability distribution $p(\beta_1, \beta_2, \sigma^2 \mid Y)$. The posterior probability distribution combines uncertainty about parameter values and uncertainty about the value of Y that would persist even if everything else was known. The peak of the posterior is the most probable value for the parameters, and a 95% interval has probability 0.95 of containing the true value.

The posterior distribution is conditional on the actual value of Y because after the data are observed, their values are known with certainty. Conditioning on the observed data is only possible because there is a prior distribution. For Frequentists, it is inappropriate to use prior distributions in science because it reduces the objectivity of scientific inference. Then probability is restricted to describing objectively random mechanisms, and uncertainty is confined to states of complete ignorance. Under this interpretation, ϵ describes an objective physically or socially random process, and the parameters are fixed but completely unknown.

To infer values for the parameters Frequentists choose a function of the available Y-values to be an estimator and then compute a point estimate and confidence interval. Since Y has a probability distribution in virtue of containing ϵ, parameter estimates (but not the parameters themselves) also have a distribution because they are functions of the objectively random Y. Confidence intervals reflect uncertainty using probability only indirectly: it is the method of interval construction rather than the interval itself that is associated with a probability, say 0.95. A 95% confidence interval is an interval computed in such a way that it will contain the true parameter value in 95% of repeated samples. Since observed Y-values cannot be conditioned on, a confidence interval depends on an infinite number of hypothetical outcomes, values of Y that did not actually occur but could have done.

Although a definition of probability in terms of repeated samples allows Frequentists to avoid quantifying levels of uncertainty, it can be problematic. For example, much social science research is concerned with explaining essentially unique situations where the idea of a repeat sample may not make sense. In contrast, interpreting ϵ as an uncertainty measure is consistent with any position on the possibility of repeated sampling, since there will be uncertainty about outcomes in either case.

As for the uncertainty represented by priors over parameters, it is important to see that a posterior distribution states what it would be rational to believe on the basis of observed data under a set of prior beliefs. Given a different set, the rational beliefs might well be different. And although the prior is inevitably subjective, it is explicitly represented, introduces no ambiguity into inference, and the consequences of changing it may be informative about the information content of the data.

The debate between Frequentists and Bayesians about social scientific inference turns on whether probability is always the best representation of uncertainty. In other words, whether the benefits of being able to condition on data that is actually observed, freedom from sampling theoretic constraints, and consistency with foundational arguments for probability as a measure of uncertainty, should outweigh concerns about introducing subjective elements into the process of increasing scientific knowledge about the social world.

References

Berger, J. O. and Wolpert, R. L. (1984). *The Likelihood Principle: A Review, Generalizations, and Statistical Implications.* Institute of Mathematical Statistics, Hayward CA.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104(2):367–405.

Cox, R. T. (1946). Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14(1):1–13. Reprinted in *Readings in Uncertain Reasoning*, G. Shafer and J. Pearl (eds.), 1990, Morgan Kaufmann.

Edwards, A. W. F. (1972). *Likelihood: An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference*. Cambridge University Press, Cambridge.

Gigerenzer, G. and Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103:650–669.

Ramsey, F. P. (1960). Truth and probability. In *The Foundations of Mathematics and Other Essays*. Littlefield Adams and Co., Paterson NJ.

Tversky, A. and Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185:1124–1131.