Statistics and the Bayesian mind

Thomas L. Griffiths
Department of Psychology
University of California, Berkeley

Joshua B. Tenenbaum
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

When people mention statistics and human cognition in the same sentence, it is usually to complain about the limitations of the latter when applied to the former. For example, during World War II the city of London was struck by German V-1 and V-2 rockets. The locations at which the rockets landed seemed to cluster in poorer parts of the city, and it was a widespread belief that these areas were being specifically targeted. After the war, the locations of rocket landings in a portion of the city were subjected to statistical analysis, which showed that there was not sufficient evidence to reject the null hypothesis that the rockets had fallen uniformly at random. The apparent clustering that people had perceived was just a coincidence.[i]

This reaction to the bombing of London is just one instance of a more general phenomenon of human thinking: a sense of coincidence that, all too often it seems, leads us to false conclusions. This tendency might suggest that statistics and human cognition are fundamentally not in accord, a view that gained prominent scientific support through the psychological research of Daniel Kahneman and Amos Tversky in the 1970s.[ii] Correct statistical reasoning may appear computationally demanding for the untrained mind – after all, even smart undergraduates often make mistakes on the simplest textbook problems – and Kahneman and Tversky set out to discover the heuristics that the mind uses to efficiently approximate these calculations. The evidence for a heuristic is the biases it induces – deviations from the normative predictions provided by probability theory. In the course of their research, Kahneman and Tversky uncovered a number of biases that seem to characterize human reasoning and decision-making, providing a significant challenge to economic models that assume people simply apply statistical decision theory. The great success of the "heuristics and biases" research program has meant that subsequent psychological research has tended to place greater emphasis on the ways in which people deviate from sound statistical inference than on the ways in which they are consistent.

Recently, cognitive scientists have begun to return to statistics as a guide to understanding human cognition, and to look at much more sophisticated kinds of statistical computations than were the focus of previous work. This turn has been motivated by the goal of trying to explain how human minds, as a routine but essential part of getting by in the real world, are able to solve extremely challenging inference problems – problems that have withstood decades of attempted solutions in machine learning, artificial intelligence, and other areas of computer science. Perhaps most striking is the human ability to draw rich inductive inferences from very sparse data, where the conclusions seem so underdetermined from the evidence available that even our imperfect level of success is remarkable. Many of the deepest questions in cognitive science take this form. For example, learning language, learning to categorize the objects in our environment, and learning about the causal relationships that involve those objects all require drawing underdetermined conclusions from data provided to us by the world and by other people. Statistics provides a way to explore these questions rigorously, specifying how a rational agent would approach a particular inductive problem and what form its solution could take. When reduced to its essence, statistics is the science of induction: it tells us under what circumstances a particular conclusion is warranted.

Exploring connections between statistics and the mind has been the theme of an increasing number of papers that aim to provide a "rational analysis" of different components of human cognition.[iii] The assumption of rationality as a guide to studying human behavior is not novel – it is commonplace in economics, statistical decision theory, and even in some of the earliest work done by statisticians. The novel contribution of rational analysis is to expand the scope of behaviors (and thoughts) analyzed in this way, going beyond decision-making and economic behavior, to explore deep questions about how we acquire and use knowledge about our world. One of the major tools in this enterprise is Bayesian statistics, which prescribes how a rational person should update their beliefs in the light of new evidence. Letting h denote a hypothesis about the structure of the world, a person's beliefs can be encoded in a "prior" probability distribution $P(h)$, where the probability assigned to each hypothesis reflects the degree to which the person believes that hypothesis to be true. Bayes' rule indicates how this distribution should be updated upon observing some data d, to produce the posterior distribution $P(h|d)$. The posterior probability of each hypothesis is given by

$$P(h \mid d) = \frac{P(d \mid h)P(h)}{\sum_{h' \in H} P(d \mid h')P(h')}$$

where $P(d|h)$, known as the likelihood, is the probability of observing d if h is true, and H is the set of all hypotheses entertained by the person. The degree to which one should believe in a particular hypothesis h after seeing data d is thus determined by two factors: the degree to which one believed in it before seeing d, as reflected by the prior probability $P(h)$, and how well it predicts the data d, as reflected in the likelihood, $P(d|h)$.

In statistics, the hypotheses being evaluated are typically different models, or different values of parameters, and Bayes' rule provides one way of selecting a hypothesis. This approach is increasing in popularity, but the use of Bayesian inference in statistics still remains a controversial topic. Part of the issue is the need to choose a prior distribution over hypotheses – something that has historically been seen as an obstacle to defining appropriately objective statistical methods. However, when used as an account of human cognition, this apparent weakness becomes an important strength. The prior distribution plays an important role in encoding the knowledge that people bring to bear on a task. By looking at the inferences that people make, we can work backwards to try to work out what kind of knowledge informed those inferences.

We have recently begun to use this approach to explore whether people's judgments are sensitive to the statistics of their environment in the way predicted by Bayes' rule.[iv] We did this by presenting people with a problem in which the data, *d*, were the current extent of a quantity, and the hypotheses, *h*, concerned its total extent. For example, one problem asked people to predict the total lifespans of clients at a life insurance company, based on their current age, while another asked them to predict the total box office take of movies based on how much money they had made so far. For such problems, the prior distribution over hypotheses, $P(h)$, should reflect the general distribution that a quantity follows. For example, the distribution of human lifespans is approximately Gaussian, while box office totals follow a distribution with power-law tails, with most movies making only a small amount of money but a few movies making a great deal of money. Different priors result in different posterior distributions, and as a consequence the predictions that one should make for human lifespans take quite a different form from the predictions one should make for movie grosses (see Figure 1). We asked people to make predictions of this kind for a variety of different kinds of quantities, following quite different distributions. The results of this experiment showed that people's predictions are in remarkably close correspondence with the predictions that result from applying Bayes' rule with an appropriate prior, suggesting that people are both sensitive to the statistics of their environment and able to use those statistics in a rational fashion when forming predictions.

These prediction studies were originally inspired by a classic textbook example from the Bayesian statistics literature: estimating the total number of trolleys or taxicabs in a given city, after seeing just the serial number on a single randomly chosen trolley or cab (and assuming that they are numbered consecutively from 1 on up).[v] This example of interval estimation has traditionally been offered as an intuitive demonstration of the power of Bayesian estimation over more conventional maximum-likelihood techniques. Essentially what we have shown is that people's intuitive estimates are indeed closely in tune with Bayesian prescriptions on this problem. Other classic problems from statistics also have cognitive correlates. For instance, consider the problem of evaluating whether a causal relationship exists between a certain drug and a certain side effect, based on observing contingency data: how many times the effect has been observed in patients taking the drug versus patients not taking the drug. Judging whether a causal relationship exists between two variables is a central capacity of everyday human cognition, but of course it also the bread and butter of statisticians. Standard statistical approaches are based on testing the null hypothesis of non-interaction, using a χ^2 statistic or some related quantity. A Bayesian approach is also possible. We can identify two possible hypotheses about the structure of the world: the hypothesis that the relationship exists, which we will call h_1, and the hypothesis that the relationship does not exist, which we will call h_0. We could assign prior probabilities to these hypotheses, $P(h_1)$ and $P(h_0)$, and then upon observing some data, *d*, compute posterior probabilities, $P(h_1|d)$ and $P(h_0|d)$ by applying Bayes' rule. With just two hypotheses, we can write Bayes' rule in "odds" form, with

$$\frac{P(h_1 \mid d)}{P(h_0 \mid d)} = \frac{P(d \mid h_1)}{P(d \mid h_0)} \frac{P(h_1)}{P(h_0)}$$

where the term on the left hand side is the posterior odds in favor of h_1, and the two terms on the right hand side are the likelihood ratio and prior odds in favor of h_1 respectively.

The likelihood ratio has a natural interpretation as a measure of the evidence that *d* provides in favor of h_1, being the only term in which *d* appears on the right hand side of the equation above. We could thus measure the evidence for a causal relationship by computing the likelihood ratio for two models, one in which the drug increases the probability of the side effect (by some unknown amount) and another in which the probability of the effect is independent of the drug. We have shown that this Bayesian computation provides an accurate account of people's intuitive judgments about the strength of evidence for causal relations, and also predicts several surprising qualitative patterns of judgment that are not predicted by a conventional χ^2 statistic, or previous heuristic accounts of causal judgment.[vi]

Modern Bayesian statistics also provides a way to re-examine some of the classic examples of human irrationality, such as our sensitivity to coincidences. We have tried to understand why it is that people are so often led astray by coincidences by providing a rational account of the role of coincidences in causal learning.[vii] Imagine that we had observed a set of events (our data, *d*), and were trying to evaluate whether there was a common cause responsible for some of those events (the hypothesis h_1) or they were simply independent (the hypothesis h_0). With the odds form of Bayes' rule in hand, we can consider the different circumstances that might arise. With a high likelihood ratio and high prior odds, we will have high posterior odds, and unambiguously conclude that a common cause exists, consistent with h_1. With a low likelihood ratio and low prior odds, we will have low posterior odds, and accept h_0 instead. But what happens when the likelihood ratio and prior odds are in conflict? We have argued that an event that has this result produces the phenomenology of a coincidence: it provides us with strong evidence (high likelihood ratio) for the existence of a causal force that we had previously believed was extremely unlikely to exist (low prior odds). The likelihood ratio provides a measure of the "strength" of a coincidence, being the extent to which it suggests that h_1 is correct. To test this idea, we conducted a series of experiments in which we examined the correspondence between such likelihood ratios and people's judgments of the strength of coincidences for different stimuli. One of these experiments was based on the bombing of London, asking people to judge how big a coincidence it would be for a particular configuration of rocket locations to arise by chance. We compared these judgments with the log-likelihood ratio produced by comparing a hypothesis in which some of the bombs fell around a common target (a mixture of a uniform and a Gaussian distribution, integrating over all model parameters) with the hypothesis of a uniform distribution (see Figure 2). Our results indicate that, in fact, people are extremely good at evaluating the evidence that an event provides for the existence of an unexpected causal force – judgments of the strength of coincidences correlate extremely well with the appropriate likelihood ratios. This suggests that the errors we make in reasoning about coincidences are not a consequence of seeing evidence where none exists (i.e. miscalculating the likelihood ratio), but rather are due to a failure to recognize just how much evidence is needed to accept an outlandish conclusion (i.e. miscalculating the prior odds). The people of London were not at fault in recognizing that the rockets exhibited some evidence of clustering, but this evidence was certainly not sufficient to justify the conclusion that poorer parts of the city were being targeted (or even to reject the null hypothesis in the statistical tests performed after the war).

Predicting the future, inferring causal relationships, and identifying coincidences are just three examples of how Bayesian statistics can shed light on human cognition. The same methods are now beginning to be used throughout cognitive science, providing a formal foundation for asking questions about how people acquire and use language, form categories, and learn causal structure. This work draws on many of the technical ideas that are at the forefront of contemporary work in Bayesian statistics, including hierarchical Bayesian models, nonparametric Bayesian methods, Markov chain Monte Carlo, and sequential Monte Carlo. Each of these ideas makes a contribution to explaining some aspect of cognition, such as how we "learn to learn", how our knowledge can increase in complexity as we acquire more data, and how the finite computational resources of our brains might approximate complex posterior distributions.

The rational analysis of human cognition provides a route by which research in statistics and cognitive science can directly inform one another. Statistics can inform cognitive science via the traditional route of providing normative theories against which human behavior can be compared. But cognitive science can also inform statistics: thinking about the computational problems people face in everyday life is a rich source of challenging new statistical problems, and sometimes the investigation of human behavior suggests that we need to revise our normative theory, or at least the way we construe the problem that people are solving. The errors that we see in simple probabilistic reasoning tasks are sometimes the consequence of people doing something far more sophisticated than we expect. While people might have trouble performing some of the calculations that appear on page 6 of an introductory statistics book, their intuitive ability to solve inductive problems means that they can effortlessly produce judgments consistent with methods that might appear on page 300 of an advanced volume on Bayesian statistics.

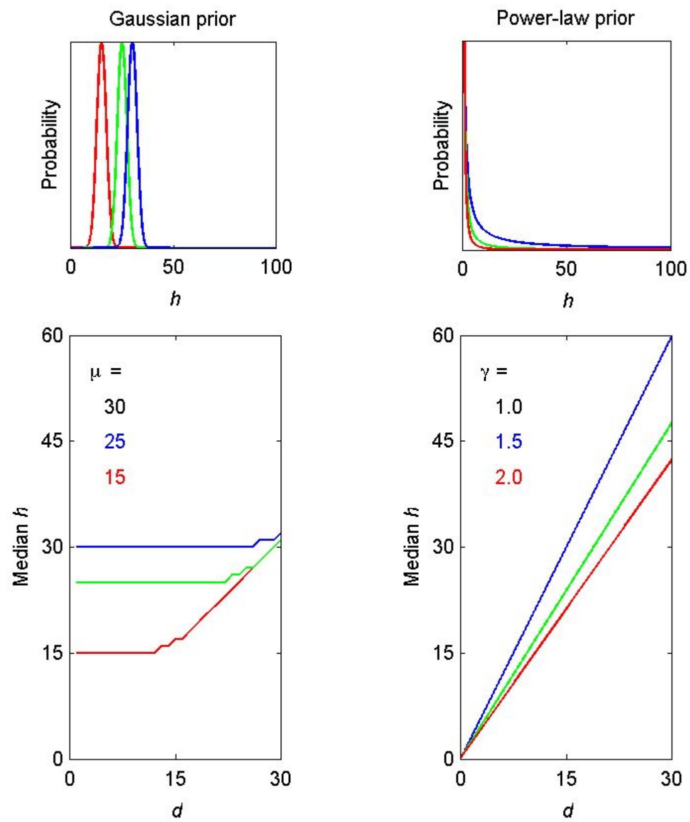

Figure 1. Gaussian and power-law priors result in different predictions. The upper panels show different priors $P(h)$ on hypotheses h as to the total value of a quantity (such as a human lifespan, or the box office take of a movie). The Gaussian priors have a standard deviation of five units, and vary in their mean, μ, falling off rapidly on either side of the mean. The power-law priors take $P(h) \propto 1/h^{\gamma}$, varying in γ, and fall off relatively slowly as h becomes large. Human lifespans follow a distribution that is approximately Gaussian, while the distribution of the box office take of movies is power-law. The lower panels show the median of the posterior distribution, $P(h|d)$, for different observed values of a quantity, d (such as the current age of a person, or the amount of money made by a movie so far). The function relating posterior medians to data is quite different for different priors, and people's predictions are sensitive to this fact.

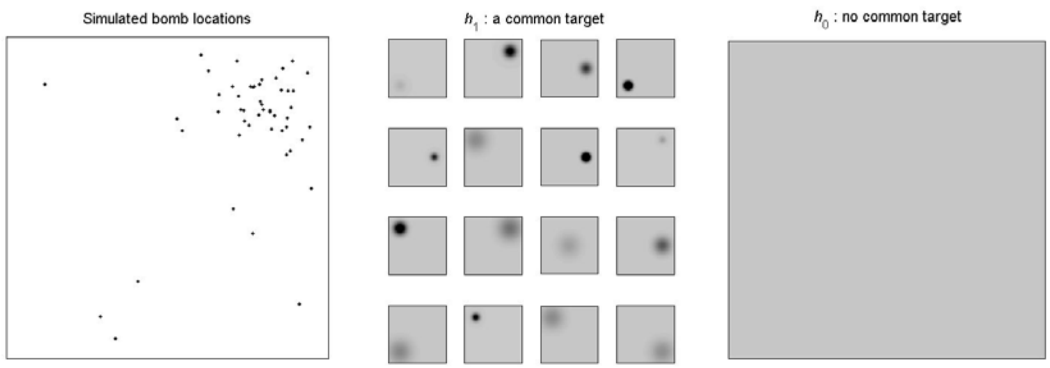

Figure 2. Evaluating coincidences using the bombing of London. In our experiments, people saw displays similar to that shown in the rightmost panel, depicting simulated bomb locations with dots, and were asked how big a coincidence it would be to observe such a pattern if the bombs actually fell at random. These judgments were compared with the likelihood ratio in favor of the hypothesis that some of the bombs shared a common target (as opposed to having no common target, and falling independently at random). These two hypotheses are depicted schematically in the other two panels. The hypothesis of a common target, h_1, defined a distribution over bomb locations obtained by integrating over the different parameter values for a mixture of a uniform and a Gaussian distribution. Some sample distributions from this set are shown, with darker regions receiving higher probability. The hypothesis of no common target, h_0, simply defined a uniform probability over all locations. People's judgments of the strength of coincidences corresponded closely with the resulting likelihood ratios.

---

[i] A discussion of this example, and of the general propensity for people to believe things that they should not, appears in Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life.* New York: Free Press.
[ii] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.
[iii] The term "rational analysis" (and a set of early examples) is due to Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

A review of some of this work appears in Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Science*, 3, 57-65.
[iv] Griffiths, T. L., & Tenenbaum, J. B. (in press) Optimal predictions in everyday cognition. *Psychological Science.*
[v] Jeffreys, H. (1939). *Theory of probability*. Oxford: Oxford University Press. Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge: Cambridge University Press.
[vi] Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 354-384.
[vii] Griffiths, T. L., & Tenenbaum, J. B. (in press) From mere coincidences to meaningful discoveries. *Cognition.*