TLS

Subscribe   🔍

Thomas Bayes and the crisis in science

David Papineau argues that it is crucial for scientists to start heeding the lessons of Thomas Bayes

By **David Papineau**

𝕏   f   ✉

Footnotes to Plato is a TLS *Online series appraising the works and legacies of the great thinkers and philosophers*

We are living in new Bayesian age. Applications of Bayesian probability are taking over our lives. Doctors, lawyers, engineers and financiers use computerized Bayesian networks to aid their decision-making. Psychologists and neuroscientists explore the Bayesian workings of our brains. Statisticians increasingly rely on Bayesian logic. Even our email spam filters work on Bayesian principles.

It was not always thus. For most of the two and a half centuries since the Reverend Thomas Bayes first made his pioneering contributions to probability theory, his ideas were side-lined. The high priests of statistical thinking condemned them as dangerously subjective and Bayesian theorists were regarded as little better than cranks. It is only over the past couple of decades that the tide has turned. What tradition long dismissed as unhealthy speculation is now generally regarded as sound judgement.

We know little about Thomas Bayes's personal life. He was born in 1701 into a well-to-do dissenting family. He entered the Presbyterian ministry after studying logic and theology at Edinburgh and lived in Tunbridge Wells for most of his adult life. Much of his energy seems to have been devoted to intellectual matters. He published two papers during his lifetime, one on theology, and the other a defence of Newton's calculus against Bishop Berkeley's criticisms. The latter impressed his contemporaries enough to win him election to the Royal Society.

The work for which he is best known, however, was published posthumously. Bayes died in 1761, but for some time up to his death he had been working on a paper entitled "An Essay towards Solving a Problem in the Doctrine of Chances". His work was passed on to his friend Richard Price, who arranged for it to be presented to the Royal Society in 1763. Bayes's essay marks a breakthrough in thinking about probability.

Probability theory was in its infancy in Bayes's day. Strange as it may seem, before the seventeenth century nobody could calculate even such simple chances as that of a normal coin landing five heads in a row. It wasn't that the information wouldn't have been useful. There was plenty of gambling before modernity. But somehow no one could get their head around probabilities. As Ian Hacking put in his groundbreaking *The Emergence of Probability* (1975), someone in ancient Rome "with only the most modest knowledge of probability mathematics could have won himself the whole of Gaul in a week".

By Bayes's time, the rudiments of probability had finally been forged. Books such as Abraham de Moivre's *The Doctrine of Chances* (1718) explained the basic principles. They showed how to calculate the probability of five heads on a normal coin (it is 1/32) and indeed more complex probabilities like five heads on a coin biased 75 per cent in favour of heads (that would be 243/1024 – about ¼). At last it was possible for gamblers to know which bets are good in which games of chance.

Not that the Reverend Bayes was any kind of gambler. What interested him was not the probability of results given different causes (like the probability of five heads given different kinds of coin). Rather he wanted to know about the "inverse probability" of the causes given the results. When we observe some evidence, what's the likelihood of its different possible causes? Some commentators have conjectured that Bayes interest in this issue was prompted by David Hume's sceptical argument in *An Enquiry Concerning Human Understanding* (1748) that reports of miracles are more likely to stem from inventive witnesses than the actions of a benign deity. Be that as it may, Bayes's article was the first serious attempt to apply mathematics to the problem of "inverse probabilities".

Bayes's paper analyses a messy problem involving billiard balls and their positions on a table. But his basic idea can be explained easily enough. Go back to the coins. If five tosses yield five heads in a row, then how likely is it that the

coin is fair rather than biased? Well, how long is a piece of string? In the abstract, there's no good answer to the question. Without some idea of the prevalence of biased coins, five heads doesn't really tell us anything. Maybe we're spinning a dodgy coin, or perhaps we just got lucky with a fair one. Who knows?

What Bayes saw, however, was that in certain cases the problem is tractable. Suppose you know that your coin comes from a minting machine that randomly produces one 75 per cent heads-biased coin for every nine fair coins. Now the inverse probabilities can be pinned down. Since five heads is about eight times more likely on a biased than a fair coin, we'll get five heads from a biased coin eight times for every nine times we get it from a fair one. So, if you do see five heads in a row, you can conclude that the probability of that coin being biased is nearly a half. By the same reasoning, if you see ten heads in a row, you can be about 87 per cent sure the coin is biased. And in general, given any observed sequence of results, you can work out the probability of the coin being fair or biased.

Most people who have heard of Thomas Bayes associate him primarily with "Bayes's theorem". This states that the probability of A given B equals the probability of B given A, times the probability of A, divided by the probability of B. So, in our case, Prob(biased coin/five heads) = Prob(five heads/biased coin) x Prob(biased coin) / Prob(five heads).

As it happens, this "theorem" is a trivial bit of probability arithmetic. (It falls straight out of the definition of Prob(A/B) as Prob(A&B) / P(B).) Because of this, many dismiss Bayes as a minor figure who has done well to have the contemporary revolution in statistical theory named after him. But this does a disservice to Bayes. The focus of his paper is not his theorem, which appears only in passing, but the logic of learning from evidence.

What Bayes saw clearly was that, in any case where you can compute Prob(A/B), this quantity provides a recipe for adjusting your confidence in A when you learn B. We start off thinking there's a one-in-ten chance of a biased coin but, once we observe five heads, we switch to thinking it's an even chance. Bayes's "theorem" is helpful because it shows that evidence supports a theory to the extent the theory makes that evidence likely – five heads support biasedness because biasedness makes five heads more likely. But Bayes's more fundamental insight was to see how scientific methodology can be placed on a principled footing. At bottom, science is nothing if not the progressive assessment of theories by evidence. Bayes's genius was to provide a mathematical framework for such evaluations.

Bayes's reasoning works best when we can assign clear initial probabilities to the hypotheses we are interested in, as when our knowledge of the minting machine gives us initial probabilities for fair and biased coins. But such well-defined "prior probabilities" are not always available. Suppose we want to know whether or not heart attacks are more common among wine than beer drinkers, or whether or not immigration is associated with a decline in wages, or indeed whether or not the universe is governed by a benign deity. If we had initial probabilities for these hypotheses, then we could apply Bayes's methodology as the evidence came in and update our confidence accordingly. Still, where are our initial numbers to come from? Some preliminary attitudes to these hypotheses are no doubt more sensible than others, but any assignment of definite prior probabilities would seem arbitrary.

It was this "problem of the priors" that historically turned orthodox statisticians against Bayes. They couldn't stomach the idea that scientific reasoning should hinge on personal hunches. So instead they cooked up the idea of "significance tests". Don't worry about prior probabilities, they said. Just reject your hypothesis if you observe results that would be very unlikely if it were true.

This methodology was codified at the beginning of the twentieth century by the rival schools of Fisherians (after Sir Ronald Fisher) and Neyman-Pearsonians (Jerzy Newman and Egon Pearson). Various bells and whistles divided the two groups, but on the basic issue they presented a united front. Forget about subjective prior probabilities. Focus instead on the objective probability of the observed data given your hypothesized cause. Pick some level of improbability you won't tolerate (the normally recommended level was 5 per cent). Reject your hypothesis if it implies the observed data are less likely than that.

In truth, this is nonsense on stilts. One of the great scandals of modern intellectual life is the way generations of statistics students have been indoctrinated into the farrago of significance testing. Take coins again. In reality you won't meet a heads-biased coin in a month of Sundays. But if you keep tossing coins five times, and apply the method of significance tests "at the 5 per cent level", you'll reject the hypothesis of fairness in favour of heads-biasedness whenever you see five heads, which will be about once every thirty-second coin, simply because fairness implies that five heads is less likely than 5 per cent.

This isn't just an abstract danger. An inevitable result of statistical orthodoxy has been to fill the science journals with bogus results. In reality genuine predictors of heart disease, or of wage levels, or anything else, are very thin on the ground, just like biased coins. But scientists are indefatigable assessors of unlikely possibilities. So they have been rewarded with a steady drip of "significant" findings, as every so often a lucky researcher gets five heads in a row, and ends up publishing an article reporting some non-existent discovery.

Science is currently said to be suffering a "replicability crisis". Over the last few years a worrying number of widely accepted findings in psychology, medicine and other disciplines have failed to be confirmed by repetitions of the original experiments. Well-known psychological results that have proved hard to

reproduce include the claim that new-born babies imitate their mothers' facial expressions and that will power is a limited resource that becomes depleted through use. In medicine, the drug companies Bayer and Amgen, frustrated by the slow progress of drug development, discovered that more than three-quarters of the basic science studies they were relying on didn't stand up when repeated. When the journal *Nature* polled 1,500 scientists in 2016, 70 per cent said they had failed to reproduce another scientist's results.

This crisis of reproducibility has occasioned much wringing of hands. The finger has been pointed at badly designed experiments, not to mention occasional mutterings about rigged data. But the only real surprise is that the problem has taken so long to emerge. The statistical establishment has been reluctant to concede the point, but failures of replication are nothing but the pigeons of significance testing coming home to roost.

Away from the world of academic science and its misguided anxieties about subjectivity, practical investigators have long benefited from Bayesian methods. When actuaries set premiums for new markets, they have no alternative but to start with some initial assessments of the risks, and then adjust them in the light of experience. Similarly, when Alan Turing and the other code-breakers at Bletchley Park wanted to identify that day's German settings on the Enigma machine, they began with their initial hunches, and proceeded systematically on that basis. No doubt the actuaries' and code-breakers' initial estimates involved some elements of guesswork. But an informed guess is better than sticking your head in the sand, and in any case initial misjudgements will tend to be rectified as the data come in.

The advent of modern computers has greatly expanded the application of these techniques. Bayesian calculations can quickly become complicated when a number of factors are involved. But in the 1980s Judea Pearl and other computer scientists developed "Bayesian networks" as a graph-based system for simplifying Bayesian inferences. These networks are now used to

streamline reasoning across a wide range of fields in science, medicine, finance and engineering.

The psychologists have also got in on the act. Statisticians might be ideologically resistant to Bayesian logic, but the unconscious brain processes of humans and other animals have no such scruples. If your visual system is trying to identify some object in the corner of the room, or which words you are reading right now, the obvious strategy is for it to begin with some general probabilities for the likely options, and then adjust them in the Bayesian way as it acquires more evidence. Much research within contemporary psychology and neuroscience is devoted to showing how "the Bayesian brain" manages to make the necessary inferences.

The vindication of Bayesian thinking is not yet complete. Perhaps unsurprisingly, many mainstream university statistics departments are still unready to concede that they have been preaching silliness for over a century. Even so, the replicability crisis is placing great pressure on their orthodoxy. Since the whole methodology of significance tests is based on the idea that we should tolerate a 5 per cent level of bogus findings, statistical traditionalists are not well placed to dodge responsibility when bogus results are exposed.

Some defenders of the old regime have suggested that the remedy is to "raise the significance level" from 5 per cent to, say, 0.1 per cent – to require, in effect, that research practice should only generate bogus findings one time in a thousand, rather that once in twenty. But this would only pile idiocy on stupidity. The problem doesn't lie with the significance level, but with the idea that we can bypass prior probabilities. If a researcher shows me data that would only occur one time in twenty if geography didn't matter to hospital waiting times, then I'll become a firm believer in the "postcode lottery", because the idea was reasonably plausible to start with. But if a researcher shows me data that would only occur one time in a 1,000 if the position of Jupiter were irrelevant to British election results, I'll respond that this leaves

the idea of a Jovian influence on the British voter only slightly less crazy than it always was.

No sane recipe can ignore prior probabilities when telling you how to respond to evidence. Yes, a theory is disconfirmed if it makes the evidence unlikely and is supported if it doesn't. But where that leaves us must also depend on how probable the theory was to start with. Thomas Bayes was the first to see this and to understand what it means for probability calculations. We should be grateful that the scientific world is finally taking his teaching to heart.

David Papineau's most recent book *is* Knowing the Score: How sport teaches us about philosophy (and philosophy about sport)

Keep reading

| *Essay*

Parmenides: Magnificent depths

The Presocratic philosopher who insisted that we must follow the rigours of an argument, no matter how surprising the conclusion - setting in motion...

By **Raymond Tallis**

| *Essay*

Roger Penrose: Beauty above all

On the life and work of the Nobel Prize-winner, who has revolutionized our understanding of black holes

By **Graham Farmelo**

| *Essay*

G. E. Moore: A great philosopher?

The realist who championed simplicity and common sense, and who discovered 'Moore's Paradox'

By **Thomas Baldwin**

Footnotes to Plato | *Essay*

Jonathan Edwards: Total depravity and empiricist philosophy

A thinker who was both 'a nemesis and an avatar' of the Enlightenment

By **Crispin Sartwell**

SUBSCRIBE TO THE PODCAST

Google Products   Spotify   Apple Podcasts

Terms & Conditions   Privacy

Cookie Settings