

COMMENT

How to be wrong but useful

Adam Kucharski*

Not long after roulette first appeared in the casinos of 18th century Paris, a new betting system emerged. Known as the 'martingale', it was adapted from a strategy used in bar games, and was supposedly foolproof. As a result, it soon became extremely popular among French players.

The martingale involved betting on black or red. The choice of color didn't matter, only the stake: rather than putting down a fixed amount each time, a player would instead double their bet after each loss. When the player eventually won, they would therefore recoup all previous losses as well as gaining their initial stake. Mathematically, the system seemed perfect. However, it had one crucial flaw: occasionally the required bets would increase beyond the size of the gambler's - or even the casino's - wallet. Although people might make small profits in the short term, eventually solvency would get in the way of strategy. Despite its popularity, the martingale was a system that no one could afford to pull off successfully.

At the start of the 21st century, another flawed strategy became fashionable. The mathematics was more complicated, and the name less elegant, but the problem was fundamentally the same. 'Collateralized debt obligations' clumped together outstanding loans such as mortgages and allowed investors to earn money by taking on some of the lenders' risk. They were based on the assumption that although one person might have a high chance of defaulting on a loan, it was extremely unlikely that everyone would default at the same time. Like the French gamblers, investors had assumed a rare event wouldn't happen, and bet the bank on that assumption.

They turned out to be wrong. As financial markets descended into chaos in 2008, and economies sank into recession the following year, it became clear that too much faith had been placed in the predictions from mathematical models. 'Wall Street's Math Wizards Forgot a Few Variables', read one headline in the *New York Times*. 'Recipe for Disaster: The Formula That Killed Wall Street', was another in *Wired*.

Before 2008, many people had been in awe of the mathematical tools in finance, happy to believe that hard

science could tame the markets; in 2009 these same ideas were met with skepticism and anger. Seeing the problems created by their field, two prominent financial mathematicians published a 'Financial Modelers' Manifesto'. Emanuel Derman and Paul Wilmott had warned about the limitations of models for years, and wanted to summarize these dangers in the wake of the crisis. Their analysis was followed by a sort of Hippocratic Oath, which began: 'I will remember that I didn't make the world, and it doesn't satisfy my equations.'

Why use models?

Although Derman and Wilmott were preaching caution rather than abandonment, lack of realism is a criticism directed often at mathematical models. Financial markets and biological systems are just too complicated to capture perfectly in a few equations. So what can models possibly contribute to these fields?

Well, quite a lot actually. George Box, a statistician, put it well when he wrote, 'Essentially, all models are wrong, but some are useful.' Despite their limitations, mathematical and statistical methods can help us tackle a number of problems that elude other techniques. When it comes to models in the life sciences, three applications stand out.

Making predictions

In 1897, a British doctor called Ronald Ross showed that it was the *Anopheles* mosquito that transmitted malaria, its bite allowing the parasite to spread from person to person. The discovery won him a Nobel Prize and a knighthood, but what he really wanted was to find a way of stopping the disease.

Ross had long believed that controlling malaria meant controlling the mosquitoes. But his peers were not convinced, pointing out that it would never be possible to remove all the mosquitoes from a region. Ross therefore outlined a mathematical model to demonstrate what might happen if mosquito numbers were reduced. 'We assume a knowledge of the causes', he wrote, 'construct our differential equations on that supposition, follow up the logical consequences, and finally test the calculated results by comparing them with the observed statistics.' The results from the model were clear: in theory disease spread could be prevented without killing every mosquito. Ross had shown that there was a critical mosquito density, below which there would simply not be enough insects to sustain transmission.

*Correspondence: a.kucharski@imperial.ac.uk

Department of Infectious Disease Epidemiology, Imperial College London, London W2 1PG

It has since become apparent that malaria control requires more effort than Ross predicted, but the concept of a critical threshold has become a vital part of ecology and epidemiology. One prominent example is vaccination, which can protect a population even if a few people don't get the jab. Today, we still use models to make predictions about disease control measures, from vaccination strategies to school closures. They are also helping us anticipate the biological effects of other actions, from pollution to overfishing. In this respect, it is good that models are artificial: they allow us to observe what happens when we change a biological system, without interfering with the real world.

Although such models are often simple, they should not be simplistic. The best modeling studies are those that follow Ross' approach: they are open about their assumptions; clear about the consequences of these assumptions; and where possible test their predictions against real observations.

Understanding complex data

As well as producing results that can be compared with data, models can help us analyze the data itself. The advent of genome sequencing has created a rich source of information for researchers, but unraveling the relationships within the data can be challenging. Phylogenetic trees are one way of identifying evolutionary patterns in such datasets. By plotting the points at which each species or variant splits into two distinct branches, the trees allow us to visualize the relationship between different parts of a population. However, even for a few variants, there are a large number of possible trees. By making assumptions about the manner and rate of mutation, we can use models to find the tree that is most likely to capture the observed data.

Phylogenetic trees can help us tackle a range of problems, from understanding the evolution of influenza viruses to mapping the diversity of fishes. When using such techniques, however, it is important to balance complexity with accuracy. Detailed, flexible models will often match the data better than simple, restrictive ones. We must therefore avoid throwing more assumptions into a model than we need to. We can do this by using an 'information criterion', which measures the amount of information that is lost when we use a particular model to describe reality: simplicity and accuracy should be rewarded, and complexity and error penalized.

Finding explanations

Models can help us find patterns, but they can also help explain them. After working on codebreaking and computing during the Second World War, Alan Turing turned his attention to developmental biology. In particular, he was interested in what dictates the shape of

organisms. Using a mathematical model, he found that it was possible to reproduce biological patterns with a 'reaction-diffusion system'. This involves two types of chemical processes: local reactions in which substances are transformed into one another, and diffusion, which makes the chemicals spread out over a surface.

It was a nice theory, but it wasn't until February 2012 that Turing's hypothesis was finally proven experimentally, with researchers showing that a reaction-diffusion system is responsible for the pattern of ridges in the roof of a mouse's mouth. Without Turing's work, we might have taken far longer to find the cause of these stripes. By proposing such mechanisms, models can therefore support - and even guide - experimental work, suggesting possible explanations for observed results, as well as areas for future investigation. However, such research needs modelers to engage with those running experiments - and the science behind this research - as much as it requires biologists to be aware of the merits of mathematical approaches.

Beyond the life sciences

Models have many benefits: they allow us to estimate future outcomes, analyze large amounts of data, and find explanations for observed patterns. Their potential will no doubt continue to increase as computing power does, allowing us to understand complex biological systems from the genetic to population level. The methods will also have applications outside the life sciences: ideas from ecology have recently been used to study networks of financial transactions, for example. No model is perfect, of course, but they can be valuable tools for comprehending - and questioning - our surroundings.

Despite their strengths, however, mathematical methods still meet with hostility. In the recent US election, statisticians Nate Silver and Sam Wang used simple models to predict the results in each state. By averaging across a large number of polls, weighting each according to their perceived reliability, both came to the conclusion that Obama had a good chance of winning. Much of the media disagreed, preferring to stick with the story that the race was too close to call. Pundits called the models a joke, or accused the statisticians of political bias. In these journalists' view, predicting the election was like predicting a coin toss, or a game of roulette: there was an equal chance the support of the electorate would land on the blue of the Democrats or the red of the Republicans. Silver and Wang disagreed, and bet on blue. Thanks to their models, they turned out to be right.

Published: 26 December 2012

doi:10.1186/gb-2012-13-12-180

Cite this article as: Kucharski A: How to be wrong but useful. *Genome Biology* 2012, **13**:180.