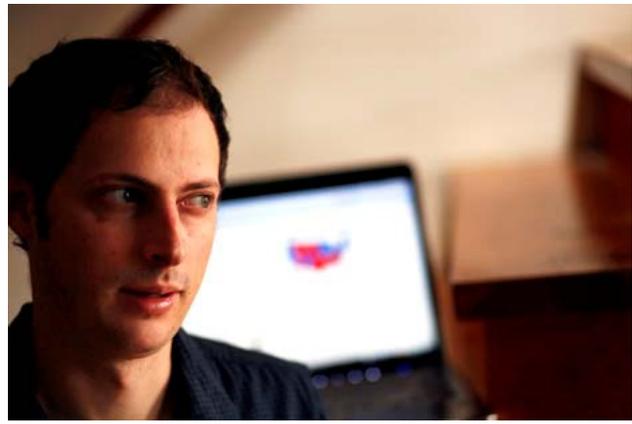


The New Yorker

JANUARY 25, 2013

What Nate Silver Gets Wrong BY

GARY MARCUS AND ERNEST DAVIS



Can Nate Silver do no wrong? Between elections and baseball statistics, Silver has become America's secular god of predictions. And now he has a best-seller, "The Signal and the Noise," in which he discusses the challenges and science of prediction in a wide range of domains, covering politics, sports, earthquakes, epidemics, economics, and climate change. How does a predictor go about making accurate predictions? Why are certain types of predictions, like when earthquakes will hit, so difficult? For any lay reader wanting to know more about the statistics and the art of prediction, the book should be essential reading. Just about the only thing seriously *wrong* with the book lies in its core technical claim.

Broadly speaking, prediction consists of three parts: dynamic modelling, data analysis, and human judgments. One of the most valuable parts of the book is the way Silver describes the interactions between these different elements of predictions, including cases in which mathematical modelling can't fully replace human judgment. Weather predictions, for example, that combine judgment with computation are between ten and twenty-five per cent more accurate than those that rely on computer programs alone. In Major League Baseball, most teams ultimately rely on a combination of hard-nosed stats and old-school human scouts. Silver also does a great job of rejecting the notorious (but ridiculous) [prediction](#) made by Chris Anderson, in *Wired*, in 2008, that Big Data would make the development of scientific theories and models "obsolete"; as Silver notes, raw data, no matter how extensive, are useless without a model: "Numbers have no way of speaking for themselves.... Data-driven predictions can succeed—and they can fail. It is when we deny our role in the process that the odds of failure rise."

Silver's one misstep comes in his advocacy of an approach known as Bayesian inference. According to Silver's excited introduction, Bayes' theorem is nominally a mathematical formula. But it is really much more than that. It implies that we must think differently about our ideas.

Lost until Chapter 8 is the fact that the approach Silver lobbies for is hardly an innovation; instead (as he ultimately acknowledges), it is built around a two-hundred-fifty-year-old theorem that is usually taught in the first weeks of college probability courses. More than that, as valuable as the approach is, most statisticians see it as only a partial solution to a very large problem.

A Bayesian approach is particularly useful when predicting outcome probabilities in cases where one has strong prior knowledge of a situation. Suppose, for instance (borrowing an old example that Silver revives), that a woman in her forties goes for a

mammogram and receives bad news: a “positive” mammogram. However, since not every positive result is real, what is the probability that she actually has breast cancer? To calculate this, we need to know four numbers. The fraction of women in their forties who have breast cancer is 0.014, which is about one in seventy. The fraction who do not have breast cancer is therefore $1 - 0.014 = 0.986$. These fractions are known as the prior probabilities. The probability that a woman who has breast cancer will get a positive result on a mammogram is 0.75. The probability that a woman who does not have breast cancer will get a false positive on a mammogram is 0.1. These are known as the conditional probabilities. Applying Bayes’s theorem, we can conclude that, among women who get a positive result, the fraction who actually have breast cancer is $(0.014 \times 0.75) / ((0.014 \times 0.75) + (0.986 \times 0.1)) = 0.1$, approximately. That is, once we have seen the test result, the chance is about ninety per cent that it is a false positive. In this instance, Bayes’s theorem is the perfect tool for the job.

This technique can be extended to all kinds of other applications. In one of the best chapters in the book, Silver gives a step-by-step description of the use of probabilistic reasoning in placing bets while playing a hand of Texas Hold ’em, taking into account the probabilities on the cards that have been dealt and that will be dealt; the information about opponents’ hands that you can glean from the bets they have placed; and your general judgment of what kind of players they are (aggressive, cautious, stupid, etc.).

But the Bayesian approach is much less helpful when there is no consensus about what the prior probabilities should be. For example, in a notorious series of experiments, Stanley Milgram showed that many people would torture a victim if they were told that it was for the good of science. Before these experiments were carried out, should these results have been assigned a low prior (because no one would suppose that they themselves would do this) or a high prior (because we know that people accept authority)? In actual practice, the method of evaluation most scientists use most of the time is a variant of a technique proposed by the statistician Ronald Fisher in the early 1900s. Roughly speaking, in this approach, a hypothesis is considered validated by data only if the data pass a test that would be failed ninety-five or ninety-nine per cent of the time if the data were generated randomly. The advantage of [Fisher’s approach](#) (which is by no means perfect) is that to some degree it sidesteps the problem of estimating priors where no sufficient advance information exists. In the vast majority of scientific papers, Fisher’s statistics (and more sophisticated statistics in that tradition) are used. Unfortunately, Silver’s discussion of alternatives to the Bayesian approach is dismissive, incomplete, and misleading. In some cases, Silver tends to attribute successful reasoning to the use of Bayesian methods without any evidence that those particular analyses were actually performed in Bayesian fashion. For instance, he writes about Bob Voulgaris, a basketball gambler,

Bob’s money is on Bayes too. He does not literally apply Bayes’ theorem every time he makes a prediction. But his practice of testing statistical data in the context of hypotheses and beliefs derived from his basketball knowledge is very Bayesian, as is his comfort with accepting probabilistic answers to his questions.

But, judging from the description in the previous thirty pages, Voulgaris follows instinct, not fancy Bayesian math. Here, Silver seems to be using “Bayesian” not to mean the use of Bayes’s theorem but, rather, the general strategy of combining many different kinds of information.

To take another example, Silver discusses at length an important and troubling paper by John Ioannidis, “[Why Most Published Research Findings Are False](#),” and leaves the reader with the impression that the problems that Ioannidis raises can be solved if statisticians use Bayesian approach rather than following Fisher. Silver writes:

[Fisher’s classical] methods discourage the researcher from considering the underlying context or plausibility of his hypothesis, something that the Bayesian method demands in the form of a prior probability. Thus, you will see apparently serious papers published on how toads can predict earthquakes... which apply frequentist tests to produce “statistically significant” but manifestly ridiculous findings.

But NASA’s [2011 study of toads](#) was actually important and useful, not some “manifestly ridiculous” finding plucked from thin air. It was a thoughtful analysis of groundwater chemistry that began with a combination of naturalistic observation (a group of toads had abandoned a lake in Italy near the epicenter of an earthquake that happened a few days later) and theory (about ionospheric disturbance and water composition).

The real reason that too many published studies are false is not because lots of people are testing ridiculous things, which rarely happens in the top scientific journals; it’s because in any given year, drug companies and medical schools perform thousands of experiments. In any study, there is some small chance of a false positive; if you do a lot of experiments, you will eventually get a lot of false positive results (even putting aside [self-deception, biases toward reporting positive results, and outright fraud](#))—as Silver himself actually explains two pages earlier. Switching to a Bayesian method of evaluating statistics will not fix the underlying problems; cleaning up science requires [changes to the way in which scientific research is done and evaluated](#), not just a new formula.

It is perfectly reasonable for Silver to prefer the Bayesian approach—the field has remained split for nearly a century, with each side having its own arguments, innovations, and work-arounds—but the case for preferring Bayes to Fisher is far weaker than Silver lets on, and there is no reason whatsoever to think that a Bayesian approach is a “think differently” revolution. “The Signal and the Noise” is a terrific book, with much to admire. But it will take a lot more than Bayes’s very useful theorem to solve the many challenges in the world of applied statistics.

Gary Marcus and Ernest Davis are professors at New York University.

Marcus has also written for [newyorker.com](#) about [the facts and fictions of neuroscience](#), [moralmachines](#), [the future of Web search](#) and what needs to be done to [clean up science](#).

Photograph by Beth Rooney/The New York Times/Redux.