

Statistics win in US election

Neuroscientist Sam Wang says that predicting vote winners is easy if you know the maths.

Brendan Maher

14 November 2012

Aside from the main outcome of last week's US presidential election, the vote tallies also promised to vindicate or vilify a cadre of blogging statisticians. For at least the past three election cycles, some bloggers have predicted the winner of the presidential election in each state with an accuracy that seemed to border on wizardry. Their secret? Aggregating dozens of national and state polls conducted throughout the election campaign, and applying statistics.

Much attention was paid to statistician Nate Silver, who writes the popular [FiveThirtyEight blog](#) for *The New York Times*. But neuroscientist Sam Wang, of Princeton University in New Jersey, not only [matched Silver's accuracy](#) on the presidential race, but also outdid him by correctly calling the result of two closely contested Senate races that Silver missed. *Nature* talked to Wang about his hobby, how it relates to his research and whether he thinks there will ever be surprises in political races again.

What did you do to celebrate the success of your predictions?

The funny thing about this is that the quality of the information is so high that election night itself is a little anticlimactic. I was at a viewing party with friends, and given how well states are correlated with one another, once I saw that New Hampshire had been called [in favour of Barack Obama], I was pretty sure that the polls were as accurate as they had been in previous years. So, my uncertainty ended at about 9.30 p.m.

How did you get into this area?

In 2004 I was watching politics, which I'm very interested in. I noticed to my frustration this incessant noise of polls, where there would be a poll from Ohio, a poll from Pennsylvania and maybe another from Pennsylvania that contradicted the first one. And each poll was accompanied by a breathless news story saying how suddenly [Senator John] Kerry was in the lead or suddenly [George W.] Bush was in the lead. It bothered me as a scientist, because people were essentially giving colour commentary on individual data points.

What I wanted was some measure that could tell me the temperature of a race at any given moment. So I cooked up a meta-analysis that took all the data and calculated a probabilistic model of where the race was likely to be. I turned that into what I would call an electoral thermometer that could monitor the race and its ups and downs over time.

Does your work as a neuroscientist inform your hobby?

Somewhat. Like many neurophysiologists, I use statistical methods to try to extract signal out of noisy data sets. And that's a common problem in neuroscience, as is analysis of time-series data.

Are you using special statistical tools to generate the predictions?

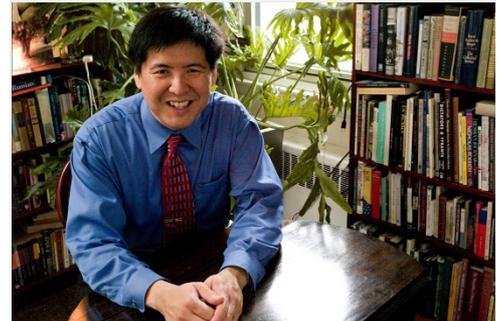
I would say that it is simple basic maths such as median-based statistics for robustness, calculation of probabilities, some Bayesian statistical analysis — but really, things that are fairly simple and within the reach of any hobbyist.

What goes into the model besides poll data? Are economic indicators predictive?

Economic data play no part in the analysis. I think that the best measure of opinion is pollsters who are experts in their craft, who call people and ask what their opinions are. And so the calculation I do is purely based on state polls. I called 50 out of 50 races [correctly, including the District of Columbia] and the 51st race, Florida, is essentially a tie. Out of 10 close Senate races, I called 10 of them correctly. So, that's one benchmark that suggests the method works well and doesn't need any econometric voodoo.

Are the polls weighted based on inherent biases?

No, that's more like putting a finger on the scale... People who are interested in polls, both professionals and hobbyists, often get



Laura Straus

Neuroscientist Sam Wang's statistical modelling correctly predicted the outcome of the recent US presidential election in 49 states plus the District of Columbia.

wrapped up in the details. One of the things I try to demonstrate while doing this is that those details often don't matter and it's better to step back a few paces and look at the whole picture. It's unsentimental but it works amazingly well.

Nate Silver took a big share of the spotlight for blogging statisticians. Was that an annoyance or a blessing?

He certainly made this whole activity more high profile. In the 2008 election, he gave colour commentary and made poll aggregation interesting. There were a number of us doing it in 2004, and our approach generally was to put up the poll aggregates and not talk about them. He made the innovation of giving a play-by-play — he's got a background in sports — and he made it fun.

What might change the ability of this kind of method to predict winners?

The profession of polling is changing. As people become less accessible through landline phones, it becomes more of a challenge to reach people — either by mobile phone or through the Internet. How well [pollsters] succeed will determine the quality of the data feed.

The other thing that is going to come up is whether the data feed continues to be of high quality. There may be a day that these data feeds become dominated by partisan organizations, or other organizations that would seek to control the flow of information. If that happens, then there's a question as to the integrity of the entire data set. Generally speaking, it all depends on having a source of high-quality data.

Do you see it as a victory for maths?

I do think that, in principle, it should get journalists and pundits to think twice about dismissing people who have a good quantitative understanding of political races. I think it was a good showcase for the kinds of contributions that poll aggregation can make.

Nature | doi:10.1038/nature.2012.11825