

MARCH 19, 2021

Stanford statisticians and *Washington Post* data scientists build more honest prediction models

A new statistical model built on Stanford research generates more nuanced predictions for complicated events. The Washington Post ran this model during the 2020 presidential election and plans to use it for future elections.

BY TAYLOR KUBOTA

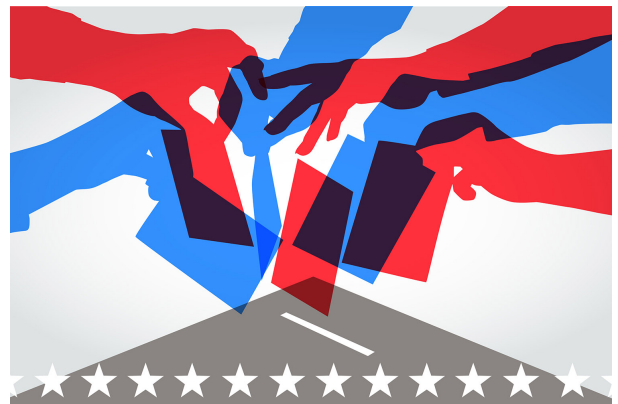
On Nov. 3, 2020 – and for many days after – millions of people kept a wary eye on the presidential election prediction models run by various news outlets. With such high stakes in play, every tick of a tally and twitch of a graph could send shockwaves of overinterpretation.

A problem with raw tallies of the presidential election is that they create a false narrative that the final results are still developing in drastic ways. In reality, on election night there is no “catching up from behind” or “losing the lead” because the votes are already cast; the winner has already won – we just don’t know it yet. More than being merely imprecise, these riveting descriptions of the voting process can make the outcomes seem excessively suspicious or surprising.

“Predictive models are used to make decisions that can have enormous consequences for people’s lives,” said Emmanuel Candès (<https://profiles.stanford.edu/emmanuel-candes>), the Barnum-Simons Chair in Math and Statistics in the School of Humanities and Sciences (<https://humsci.stanford.edu/>) at Stanford University. “It’s extremely important to understand the uncertainty about these predictions, so people don’t make decisions based on false beliefs.”

Such uncertainty was exactly what *The Washington Post* data scientist Lenny Bronner (BS ’16, MS ’17) aimed to highlight in a new prediction model that he began developing for local Virginia elections in 2019 and further refined for the presidential elections, with the help of John Cherian (BS ’17, MS ’17), a current PhD student in statistics at Stanford whom Bronner knew from their undergraduate studies.

“The model was really about adding context to the results that were being shown,” said Bronner. “It wasn’t about predicting the election. It was about telling readers that the results that they were seeing were not reflective of where we thought the election was going to end up.”



(<https://news-media.stanford.edu/wp-content/uploads/2021/03/18114925/prediction.jpg>)

A prediction model developed by *The Washington Post* for the presidential elections in 2020 applied Stanford statistics research. The model highlighted the uncertainties that exist in voting result forecasts. (Image credit: Getty Images)

This model is the first real-world application of an existing statistical technique developed at Stanford by Candès, former postdoctoral scholar Yaniv Romano and former graduate student Evan Patterson. The technique is applicable to a variety of problems and, as in the *Post's* predication model, could help elevate the importance of honest uncertainty in forecasting. While the *Post* continues to fine-tune their model for future elections, Candès is applying the underlying technique elsewhere, including to data about COVID-19.

Avoiding assumptions

To create this statistical technique, Candès, Romano and Evan Patterson combined two areas of research – quantile regression and conformal prediction – to create what Candès called “the most informative, well-calibrated range of predicted values that I know how to build.”

While most prediction models try to forecast a single value, often the mean (average) of a dataset, quantile regression estimates a range of plausible outcomes. For example, a person may want to find the 90th quantile, which is the threshold below which the observed value is expected to fall 90 percent of the time. When added to quantile regression, conformal prediction – developed by computer scientist Vladimir Vovk – calibrates the estimated quantiles so that they are valid outside of a sample, such as for heretofore unseen data. For the *Post's* election model, that meant using voting outcomes from demographically similar areas to help calibrate predictions about votes that were outstanding.

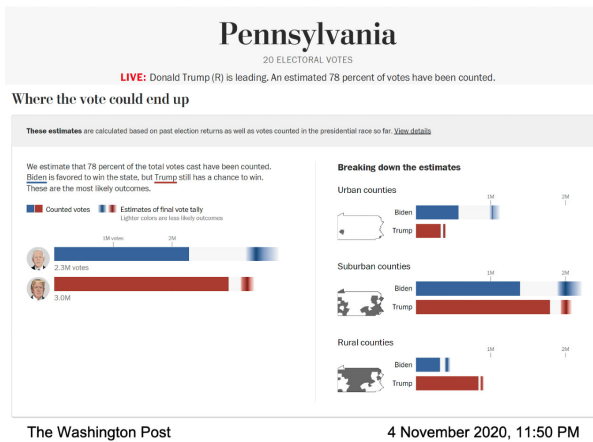
What's special about this technique is that it begins with minimal assumptions built into the equations. In order to work, however, it needs to start with a representative sample of data. That's a problem for election night because the initial vote counts – usually from small communities with more in-person voting – rarely reflect the final outcome.

Without access to a representative sample of current votes, Bronner and Cherian had to add an assumption. They calibrated their model using the vote tallies from the 2016 presidential elections so that when an area reported 100 percent of their votes, the *Post's* model would assume that any changes between that area's 2020 votes and its 2016 votes would be equally reflected in similar counties. (The model would then adjust further – reducing the influence of the assumption – as more areas reported 100 percent of their votes.) To check the validity of this method, they tested the model with each presidential election, beginning with 1992, and found that its predictions closely matched the real-world outcomes.

“What's nice about using Emmanuel's approach to this is that the error bars around our predictions are much more realistic and we can maintain minimal assumptions,” said Cherian.

Visualizing uncertainty

In action, the visualization of the *Post's* live model was painstakingly designed to prominently display those error bars and the uncertainty they represented. The *Post* ran the model to forecast the range of likely election outcomes in different states and county types; counties were categorized according to their demographics. In every case, each nominee had their own horizontal bar that filled in solid – blue for Joe Biden, red for Donald Trump – to show known votes. Then, the rest of the bar contained a gradient that represented the likeliest outcomes for the outstanding votes, according to the model. The darkest area of the gradient was the most likely result.



(https://news-media.stanford.edu/wp-content/uploads/2021/03/18115146/thumbnail_penn_1)

Screenshot of *The Washington Post* election model, showing the voting prediction for Pennsylvania on Nov. 4, 2020. (Image credit: Courtesy of *The Washington Post*)

“We talked to researchers about visualization of uncertainty and we learned that if you give someone an average prediction and then you tell them how much uncertainty is involved, they tend to ignore the uncertainty,” said Bronner. “So we made a visualization that is very ‘uncertainty forward.’ We wanted to show, this is the uncertainty and we’re not even going to tell you what our average prediction is.”

As election night wore on, the darkest part of Biden’s gradient in the total vote visualization was further to the right side of the bar, which meant the model predicted he would end up with more votes. His gradient was also wider and spread asymmetrically toward the higher vote side of the bar, which meant the model predicted there were many scenarios, with decent odds, where he would win *more* votes than the likeliest number.

“On election night, we noticed that the error bars were very short on the left side of Biden’s bar and very long on the right side,” said Cherian. “This was because Biden had a lot of upside to potentially outperform our projection in a substantial way and he didn’t have that much downside.” This asymmetric prediction was a consequence of the particular modeling approach used by Cherian and Bronner. Because the model’s forecasts were calibrated using results from demographically similar counties that had finished reporting their votes, it became clear that Biden had a good chance to significantly outperform the 2016 Democratic vote in suburban counties, while it was extremely unlikely he would do worse.

Of course, as the vote counting headed toward the finish, the gradients shrunk and the *Post*’s uncertain predictions looked increasingly certain – a nerve-racking situation for data scientists concerned with overstating such important conclusions.

“I was particularly worried that the race would come down to one state, and we would have a prediction on our page for days that ended up not coming true,” said Bronner.

And that worry was well founded because the model did strongly and stubbornly predict a Biden win for several days as the final vote tallies crept in from not one state, but three: Wisconsin, Michigan and Pennsylvania.

“He ended up winning those states, so that ended up working well for the model,” added Bronner. “But at the time it was very, very stressful.”

Following their commitment to transparency, Bronner and Cherian also made the code to their election model public, so people can run it themselves. They’ve also published technical reports on their methods (available for download here (https://s3.us-east-1.amazonaws.com/elex-models-prod/2020-general/write-up/election_model_writeup.pdf)). The model will run again during Virginia state elections this year and the midterm elections in 2022.

“We wanted to make everything public. We want this to be a conversation with people who care about elections and people who care about data,” said Bronner.

Forcing honesty

The bigger picture for Candès is how honest and transparent statistical work can contribute to more reasonable and ethical outcomes in the real world. Statistics, after all, are foundational to artificial intelligence and algorithms, which are pervasive in our everyday lives. They orchestrate our search results, social media experience and streaming suggestions while also being used in decision-making tools in medical care, university admissions, the justice system and banking. The power – and perceived omnipotence – of algorithms troubles Candès.

Models like the one the *Post* used can address some of these concerns. By starting with fewer assumptions, the model provides a more honest – and harder to overlook – assessment of the uncertainty surrounding its predictions. And similar models could be developed for a wide variety of prediction problems. In fact, Candès is currently working on a model, built on the same statistical technique as the *Post*'s election model, to infer survival times after contracting COVID-19 on the basis of relevant factors such as age, sex and comorbidities.

The catch to an honest, assumption-free statistical model, however, is that the conclusions suffer if there isn't enough data. For example, predictions about the consequences of different medical care decisions for women would have much wider error bars than predictions regarding men because we know far less about women, medically, than men.

This catch is a feature, though, not a bug. The uncertainty is glaringly obvious and so is the fix: We need more and better data before we start using it to inform important decisions.

“As statisticians, we want to inform decisions, but we're not decision makers,” Candès said. “So I like the way this model communicates the results of data analysis to decision makers because it's extremely honest reporting and avoids positioning the algorithm as the decision maker.”

To read all stories about Stanford science, subscribe to the biweekly *Stanford Science Digest* (<http://eepurl.com/dLmCng>).

Media Contacts

Taylor Kubota, Stanford News Service: (650) 724-7707, tkubota@stanford.edu (<mailto:tkubota@stanford.edu>)
