TEXAS TECH UNIVERSITY
Honors College™

A Bayesian Statistical Model to Predict American Presidential Elections

Brittany Alexander, Leif Ellingson Ph.D.

TEXAS TECH UNIVERSITY
College *of* Arts & Sciences™

Abstract

A previous Bayesian model used to predict the 2008, 2012, and 2016 United States Presidential Elections using only poll data resulted in nearly identical electoral college predictions to FiveThirtyEight, and 95.329% relative accuracy to the FiveThirtyEight Polls Plus model in terms of root mean square error of the predictions of the two major candidates. The previous model used poll data from either other states or national polls and used the Gaussian conjugate prior. The previous model used the Gaussian conjugate prior calculation based on the mean and variance of poll data. In an attempt to increase the accuracy of the model and further test the viability of this method, multiple models with minor differences are used on the same data used as the previous model. The definition of swing states and regions used to define the priors used are more formalized than the previous model. The new models now pool the polls together from other states in the regions and uses the pooled estimates as the prior instead of relying on poll data from one state. The new models compare the beta and Gaussian conjugate prior and use three different methods to reassign undecided voters, and either pooling all the polls together and make one calculation with the conjugate prior and or update the posterior with every poll. A set of models with a noninformative beta (1,1) prior was also used. Only the iterative Gaussian model outperformed the previous model, and the proportional normalization was the best.

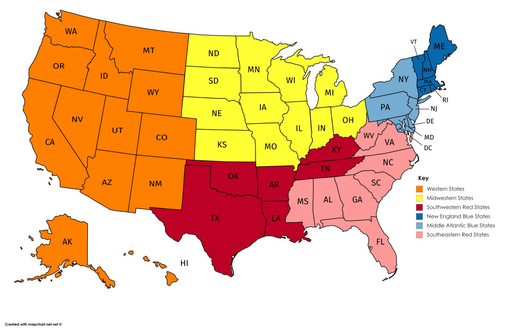
General Methodology

- The models used data from Pollster
- Only polls conducted from July 1st to the Friday before the election with sample size, date, and the support of the two major candidates was included
- Minor candidates were excluded from the prediction and the voting results were adjusted to exclude minor candidates
- Bayesian models except the noninformative Beta model used a pooled collection of polls from either other states in the group
- Swing states were defined as a state that had been won by both the Democratic and Republican in the past four presidential elections
- States were broken up into groups based on regions and the political lean of a state
- Undecided and minor candidate voters were reassigned either based on the poll results (proportionally), based on past votes, or were split 50-50

Specific Methodology of the Models

- Model 1 Prior Polls: Takes average of the other states in a group and uses that to predict a state
- Model 2 Polling Average: Takes average of the polls for each state
- Model 3 Beta Conjugate Prior: Uses poll data from other states as the prior
- Model 4 Gaussian Iterative : Updates using the normal approximation to the binomial distribution for each poll, uses poll data for other states as the prior
- Model 5 Gaussian Pooled People: Pools all the responses together into a single giant poll, uses poll data for other states as the prior
- Model 6 Gaussian Pooled Polls: Uses the average and standard deviation of polls, uses poll data for other states as the prior
- Model 7 Beta Noninformative Prior: Uses a Beta (1,1) noninformative prior

Map of Regions



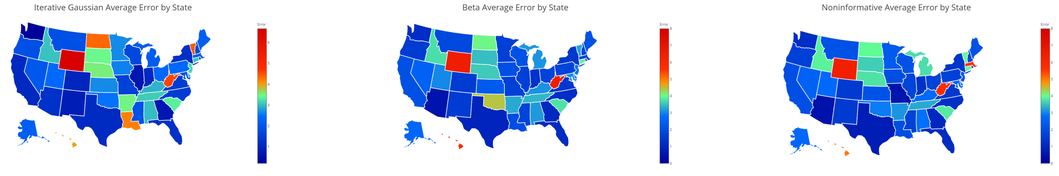
Comparison to Original Model

- Only the Gaussian Iterative Model, the noninformative model, and the polls average model performed better than the original model, but the other models were very close
- The new methods didn't usually change the performance
- Unlike the original model. the new models correctly predicted Ohio in 2016

Mathematical Background

- Bayes' Theorem is a way to use prior information to find an estimate of a probability
- Bayes' Theorem: $P(H|E) = \frac{P(H)P(E|H)}{P(E)}$ where H is the hypothesis and E is the evidence (also called the prior). $P(E|H)$ is called the likelihood. $P(H|E)$ is called the posterior probability.
- Analysis of multiple hypotheses can be done with Bayes' Rule: $P(H|E) \propto P(H)P(E|H)$
- If the distribution of the prior is known a conjugate prior can be used

Average Error of the Models with Proportional Normalization



	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Average Error	5.9183	2.500653	2.783663	2.559477	2.526797	2.80915	2.65817

Results and Discussion

- All models were worse than the FiveThirtyEight Polls Plus Model in terms of RMSE, but performed the same at predicting winners
- Polls from other similar states are not good predictors of the results in Partisan states
- State level fundamental modelling would likely have better results
- The independence assumption needs to be removed and the polls need to be adjusted for bias in future research
- In the future, I plan to study other ways to improve prediction of American elections

Political Background

- The president is decided by the electors in the electoral college. A majority is needed to elect the president. Most electors are chosen based on the winner of that state.
- Most Bayesian models use hierarchical modelling based on economic and political data

Bibliography

[1]Christensen, W. F., & Florence, L. W. (2008), January 1). Predicting Presidential and Other Multistage Election Outcomes Using State-Level Pre-Election Polls. *The American Statistician*, *62*(1), 1-10. doi:10.1198/000313008x267820

Five Thirty Eight. (2016). [Data] available from http://fivethirtyeight.com

[2]Gelman, A., & King, G. (1993). Why Are American Presidential Election Campaign Polls So Variable When Votes Are So Predictable? British Journal of Political Science, 23(04), 409. doi:10.1017/s0007123400006682

Huffington Post. (2016). Pollster [Data] available from http://elections.huffingtonpost.com/pollster

Popkin, S. L. (1991). The reasoning voter: communication and persuasion in presidential campaigns. Chicago: University of Chicago Press.

Princeton Election Consortium (2016). [Data] Available from http://election.princeton.edu

Real Clear Politics. (2016). Real Clear Politics Averages [Data] available from http:www.realclearpolitics.com

Acknowledgements

I would like to thank the Honors College Undergraduate Research Scholars Program supported by CH and Helen Jones Foundations.