

# A Bayesian Statistical Model for the Prediction of the 2016 United States Presidential Election

## Brittany Alexander, Leif Ellingson Ph.D.

### Abstract

Using a combination of polling data and previous election results, Nate Silver successfully predicted the Electoral College distribution in the presidential election in 2008 with 98% accuracy and in 2012 with 100% accuracy. His success was attributed largely to his focus on Bayesian statistical modeling. However, Bayesian statistical modeling for presidential election prediction has not been studied prospectively since Nate Silver's success in 2012. This study is aimed at utilizing Bayesian modeling for predicting the results of two sets of experiments – all 50 of the 2016 Republican Presidential primary elections and the 2016 general Presidential election. In both cases, the data will consist of the most recent polling results. In the primary process, past elections will serve as a prior. In the general election the prior will be based on either a national poll or a poll from a similar state.

### 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

### Political Background

- The GOP nomination process involves assigning delegates to the convention using election results from individual states.
- Most delegates are bound to a single candidate and the candidate with the most delegates wins the nomination.
- 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.

### Nomination Process Methods

- The format was based on the methods discussed by Christensen and Florence (2008)<sup>1</sup> for use as an undergraduate research project on presidential elections.
- Predictions were posted on a blog two days before the election (<http://bit.do/2016Math>)
- Poll data were taken from Pollster.
- The first four voting states were predicted with a t-distribution.
- All other states used a previously voting state as the prior, and poll data as the hypothesis
- The likelihood was generated using a normal distribution and standard error of means
- Normalization was done with second choice data from the state (if available) or national polls.
- When candidates dropped out, their support in the polls was shifted to the most similar candidate.
- Delegate prediction followed the individual regulations for that state

### Nomination Results and Discussion

- 77.6% of the bound delegates to the GOP convention were correctly predicted with this method.
- No other model predicting the delegates of a major party in a presidential nomination process was found.
- It was difficult to find a good prior for some states.
- Exit polls from previous states could be a better prior

### General Election Methods

- A normal posterior was assumed, and the poll data were treated parametrically.
- The calculations were made using a Gaussian Bayesian conjugate prior.
- The poll data were taken from Pollster.
- The calculations were made in Python using Numpy and Scipy.
- A blog was used to post the election predictions 60 hours in advance (<http://bit.do/2016Math>)
- Five different categories were established for use as priors: Midwest Red States (Nebraska data), Southern Red States (Texas data), Northern Blue States (New York data), Western Blue States (California data), and Swing States (National poll data)
- Polls were used if they were taken on or after July 1<sup>st</sup>, and were released before the Friday before Election Day
- Gary Johnson was included in the prediction
- A measurement called the "Other-Johnson Factor" was created to synthesize Johnson's support
- A similar method was used with Evan McMullin in Utah
- The 2016 analysis was finalized on November 5<sup>th</sup>
- The model was applied to the 2008 and 2012 election, using the same method as 2016.
- In 2008 and 2012 only the two major candidates were studied because other support was less than 5% in both years.

### Bibliography

- <sup>1</sup>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>
- <sup>2</sup>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>
- Mayer, W.G. (2003) 'Forecasting Presidential Nominations or, My Model Worked Just Fine, Thank You', *PS: Political Science & Politics*, 36(2), pp. 153–157. doi: 10.1017/S1049096503001999.
- 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>

### Accuracy of Models in Predicting the Winners of States

Race	Real Clear Politics <sup>a</sup>	Princeton Election Consortium <sup>b</sup>	Five Thirty Eight <sup>c</sup>	Tested Model
2008	0.96	0.98	0.98	0.98
2012	0.98	0.98	1	1
2016	0.92	0.90	0.90	0.88

a – [realclearpolitics.com](http://realclearpolitics.com)  
b – [elections.princeton.edu](http://elections.princeton.edu)  
c – [fivethirtyeight.com](http://fivethirtyeight.com)  
d – [2016usaelectionpredictions.blogspot.com](http://2016usaelectionpredictions.blogspot.com)

### General Election Results and Discussion

- The tested model performed well in all predicted elections compared to other models
- A Donald Trump win was not predicted by any of the models in the table above
- The method of using data from other sources in the same election appears to be helpful in addressing limited information

### Conclusion

- Elections are difficult events to predict due to the constantly changing opinions of the American electorate.
- The tested model relied heavily on good polling performance
- The ideal model for presidential election prediction may be complicated
- In the future, I plan to study other ways to improve prediction of American elections

### Acknowledgements

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