Empirical Validation of Automated Redistricting Algorithms on the Virginia House of Delegates District Map

Madeleine Goertz

International Community School

AP Research

Randall S. Huberman

May 20th, 2021

Empirical Validation of Automated Redistricting Algorithms on the Virginia House of Delegates District Map

Method

Notes, this is still a rough draft.

Choice of Research Method

For this study, I chose to use the experimental design method because it will allow me to isolate the hypothetical impact of the redistricting algorithm from other possible confounding variables. This method also includes the use of a control group, which allows the researcher to establish causation.

Components of Experimental Design

Experimental Units. The experimental units for this study are the complete datasets for each election year in Viriginia. I have one dataset for each of these years: 2015, 2017, 2019. Every row in each dataset corresponds to a precinct, the smallest geographical unit by which votes are tabulated in Virginia. For each precinct, I have the following attributes: total population, population by race, total voting-age population (VAP)(population over the age of 18), VAP by race, total votes for the democratic House of Delegates (HOD) candidate, total votes for the Republican HOD candidate, and the total votes for any other HOD candidate. Additionally, each precinct has a polgon associated with it that represents its geographical shape.

Treatments. The treatments for this study are the three different redistricting algorithm that I'm comparing: Markov chain Monte Carlo (Fifield, Higgins, et al., 2020), Sequential Monte Carlo (McCartan & Imai, 2020), and Random Seed Growth (Chen & Rodden, 2013). I'm using the implementations in the R programming language "redist" package (Fifield, Kenny, et al., 2020). See the Literature Review section for a deeper dive into these algorithms. Broadly, I chose them because they are deterministic. Much of the literature focuses on creating many possible redistricting plans for a commission to choose from, but these three aim to create an "ideal" map.

Response Variables. Broadly, the goal will be to evaluate how "fair" each redistricting plan generated by each algorithm for each year is. Each of the following metrics accounts for a separate aspect of this fairness. These metrics were canonized and made rigorous by Katz et al. (2020).

Partisan Symmetry. The idea of partisan symmetry is that the difference between popular vote share and number of seats controlled should align for both parties. An example would be that if Republicans win 60% of the votes but control 65% of the seats, then in a symmetrical system, Democrats should also be able to control 65% of the seats by winning 60% of the votes.

Efficiency Gap. Still need to read about this.

Chamber Power Balance. Since the redistricting that's occuring is hypothetical and I have precinct-level election results for each of these years, I can simulate what the power distribution in the VA House of Delegates would be if the proposed redistricting plan had been used.

Control Group. The official VA House of Delegates map used in the years 2015-2019 will serve as the control group for this experiment. I will compute the same metrics for this map as I will for my hypothetical redistricting plans.

Principles of Experimental Design

The primary principles of experimental research design are randomization, replication, and local control. This is how I plan to address them.

Randomization. Every experimental unit will receive each treatment, and every experimental unit can be replicated many times without issue, so thereâĂŹs no error from a lack of randomization. Think of each treatment operating within a separate parallel universe.

Replication. There is no need for me to run my trials several times (run the same algorithm on the same data set several times) because these are deterministic algorithms, and the datasets will be immutable.

Local Control. All of the redistricting will be happening in controlled environments, so there will be no way for lurking variables to creep in and confound my results.

References

Chen, J., & Rodden, J. (2013). Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures. *Quarterly Journal of Political Science*, *8*(3), 239–269. https://doi.org/10.1561/100.00012033

Fifield, B., Kenny, C. T., McCartan, C., Tarr, A., Higgins, M., Kawahara, J., & Imai, K. (2020). Redist: Simulation Methods for Legislative Redistricting. Retrieved January 29, 2021, from https://CRAN.R-project.org/package=redist

Fifield, B., Higgins, M., Imai, K., & Tarr, A. (2020). Automated Redistricting Simulation Using Markov Chain Monte Carlo. *Journal of Computational and Graphical Statistics*, *0*(0), 1–14. https://doi.org/10.1080/10618600.2020.1739532

Katz, J. N., King, G., & Rosenblatt, E. (2020). Theoretical Foundations and Empirical Evaluations of Partisan Fairness in District-Based Democracies. *American Political Science Review*, *114*(1), 164–178. https://doi.org/10.1017/S000305541900056X

McCartan, C., & Imai, K. (2020). Sequential Monte Carlo for Sampling Balanced and Compact Redistricting Plans. *arXiv:2008.06131 [cs, math, stat]*. Retrieved January 18, 2021, from http://arxiv.org/abs/2008.06131