Methods used to Analyze 2020 North Carolina State Congressional Redistricting Landscape

Gregory Herschlag and Jonathan C. Mattingly

November 8, 2021

1 Overview

We place probability distributions on redistricting plans of the state of North Carolina. The distributions embody different policy choices. With each distribution, we produce representative ensembles of maps to serve as benchmarks against which to compare specific maps. The ensembles are generated by using the Metropolis-Hasting Markov Chain Monte Carlo Algoritm in a parallel tempering framework which employees the proposal from the multiscale forest RECOM algorithm [1, 2].

In our analysis, we use historical elections to help situate the behavior of our ensemble under a number of political climates, we do not use any political data in developing our distribution; and hence, the ensemble generated from it. We produce collections of histograms, showing the typical seats awarded to each party under our distribution and a number of different political climates. We also produce *rank ordered boxplots* which show the typical partisan make up of the collection of districts, again under the variety of election climates embodied in different historical votes. These figures and analyses for the North Carolina State House and Senate are included in a companion document. Here we give an overview of the methods used. A more complete description of our methods will be provided in later reports.

2 Ensemble Overview

In this first analysis, we consider a distribution (and hence an ensemble) satisfy the following constraints:

- The maps split no more than 14 counties.
- The maps split no county into more than two districts.
- Districts traverse counties as few times as possible.
- All districts are required to consist of one contiguous region.
- The deviation of the total population in any district is within %1 of the ideal district population.

Compactness: The distributions on redistricting plans are constructed so that a plan with a larger total isoperimetric ratio is less likely that those with lower total isoperimetric ratio. The total isoperimetric ratio of a redistricting plan is simply the sum of the isoperimetric ratios over each district. The isoperimetric ratio is the reciprocal of the Polsby-Poper score; hence, smaller isoperimetric ratio correspond to larger Polsby-Poper scores. As the General Assembly stated in its guidance that the plans should be compact according to the Polsby-Popper score [3], we tuned the distribution so that it yeilds compact redistricting plans by this measure. We further limited our distribution only to include those with Isoparametric score less than 80.

3 Mathematical Description of Ensemble Distribution

In specifying our distribution, we have chosen to define explicit distributions and then use an implementation of the Metropolis-Hastings algorithm to generate the ensemble. We feel this chose promotes transparency because an explicit distribution can better be discussed and critiqued. It also allows us to more explicitly translate the policy considerations into the ensemble.

In order to formally define our distributions, the partition of the precinct adjacency graph in to a spanning forest $\mathcal{T}$ with 14 district trees $\{\mathcal{T}_1, \cdots, \mathcal{T}_{14}\}$ corresponding to each district. Hence $\mathcal{T} = \{\mathcal{T}_1, \cdots, \mathcal{T}_{14}\}$ completely specifies the redistricting.

If we let $A_j(\mathcal{T})$ and $B_j(\mathcal{T})$ be respectively the surface area and perimeter (or length of the boundary) of the j-district then our compactness score is

$$J_{\text{compact}}(\mathcal{T}) = \sum_{j=1}^{14} \frac{A_j(\mathcal{T})}{B_j^2(\mathcal{T})}.$$

Then the probability of drawing the spanning forest $\mathcal{T}$ is

$$\text{Prob}(\mathcal{T}) = \begin{cases} \frac{1}{Z} e^{-w_{\text{compact}} J_{\text{compact}}(\mathcal{T})} & \text{for } \mathcal{T} \text{ which is allowable} \\ 0 & \text{for } \mathcal{T} \text{ which is not allowable} \end{cases}$$

Here Z is a number which makes the sum of $\text{Prob}(\mathcal{T})$ over all spanning forests with 14 trees equal to one.

The collection of allowable spanning forests $\mathcal{T}$ is defined to those which produce redistricting plans which satisfy the following conditions:

1. all districts are connected
2. the populations of each district is within $\%1$ of the ideal district population.
3. No more that 14 counties are split with no county split more once.
4. We minimize the occurrence of districts traversing county boundaries.

4 Sampling Method

We have chosen the distribution from which to draw our ensemble to comply with the desired policy and legal considerations. It is well accepted that not all distributions on possible redistricting plans are equally easy to sample from.

To effectively generate a representative ensemble of maps from these distributions, we use a the well established method of parallel tempering. It allows one to effectively sample form a possibly difficult to sample distribution by connecting it to an easy to sample distribution through a sequence of intermediate "interpolating" distributions.

We connect our desired distributions which includes a compactness to a measure on redistricting plans which is uniform on spanning forests which satisfy the population and county constants. Furthermore, it can be effectively sampled using a variation on the metropolized multiscale forest RECOM sampling algorithm outlined in [2, 1].

In sampling the interpolating ladder of distributions between the easier-to-sample measure and our target measure which includes a compactness score, we use parallel tempering with a classical Metropolis-Hasting sampling scheme to sample each level of the interpolating ladder of distributions. As proposals in the Metropolis-Hasting sampling scheme, we use multiscale forest RECOM proposals. We will detail the algorithmic choices and our validation tests in later documents.

5 Election Data Used in Analysis

The historic elections we consider are from the years 2008, 2012, 2016, and 2020. We only consider statewide elections. In our visualizations we have selected a representative subset to improve to readability of the plots We will use the following abbreviations: AG for Attorney General, USS for United States Senate, CI for Commissioner of Insurance, LG for Lieutenant Governor, GV for Governor, TR for State Treasure, SST for Secretary of State, AD for State Auditor, CA for Commissioner of Agriculture, and PR for United States President. We add to these abbreviations the last two digits of the year of the election. Hence CI08 is the vote data from the Commissioner of Insurance election in 2008.

References

[1] E. A. Autry, D. Carter, G. Herschlag, Z. Hunter, and J. C. Mattingly, "Multi-scale merge-split markov chain monte carlo for redistricting," 2020.

[2] E. Autrey, D. Carter, G. Herschlag, Z. Hunter, and J. C. Mattingly, "Metropolized forest recombination for monte carlo sampling of graph partitions," 2021.

[3] N. Legislature, "2021 joint redistricting committee proposed criteria," https://ncleg.gov/documentsites/committees/Senate2021-154/2021/08-09-2021/20212021.