Methods used to Analyze 2020 North Carolina State Legislative Redistricting Landscape

Gregory Herschlag and Jonathan C. Mattingly

October 26, 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 proposal from the multiscale forest RECOM algorithm [1, 2] and the single-node flip algorithm [3].

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 two different ensembles generated from different measures that emphasize different policies. Both of the distributions we build our ensembles from respect the county clusters we derived in [4] by algorithmically implementing the ruling Stephenson v. Bartlett, 562 S.E.2d 377 (N.C. 2002). That is to say in both the State House and State Senate, the state is segmented into groups of counties referred to as *county clusters* so that the population of each county cluster can be divided into a number of districts each with a population within 5% of the ideal district population. The county clusters are different for the State House and State Senate as the number of districts, and hence the ideal district populations, are different. Each district is constrained to lay entirely within one county cluster.

Beyond the county cluster requirement both ensembles also satisfy the following constraints:

- The maps minimize the number of split counties.
- Districts traverse counties as few times as possible.
- All districts are required to consist of one contiguous region.
- Except for two exceptions, the deviation of the total population in any district is within %5 of the ideal district population. The two special cases are explained in Section 3.
- Voting tabulation districts (i.e. VTDs or precincts) are not split (see again the two exceptions with population deviation in Section 3)

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 [5], we tuned the distribution so that it yeilds compact redistricting plans by this measure.

Municipality Preservation: We now come to the property which distinguishes the two different distributions on redistricting we present here. The first ensemble makes no attempt to preserve municipalities beyond any effect which favoring compact districts might impart. In the second ensemble, we place a second weighting which favors redistricting plans that fragment fewer municipalities. Since the term used to discuss cities and other municipalities is minor civil division (MCD), we refer to this second ensemble as either the MCD ensemble or the ensemble that preserves municipalities.

3 Mathematical Description of Ensemble Distribution

In designing our two distributions, 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, we consider the labeling ξ of the precincts of the map of NC with the number $\{1, \ldots, d\}$, where d is the total number of districts. So for the i-th precinct, $\xi(i)$ gives the district to which the precinct belongs. If we let $A_j(\xi)$ and $B_j(\xi)$ be respectively the surface area and perimeter (or length of the boundary) of the j-district then our compactness score is

$$J_{\text{compact}}(\xi) = \sum_{j=1}^{d} \frac{A_j(\xi)}{B_j^2(\xi)}.$$

Then the probability of drawing the redistricting ξ is

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

Here Z is a number which makes the sum of Prob(ξ) over all redistricting plans equal to one.

The collection of allowable redistricting plans ξ is defined to be all redistricting plans which satisfy the following conditions:

1. all districts are connected
2. the populations of each district is within %5 of the ideal district population unless the district in the wake county cluster in the senate or the Craven-Carteret county cluster in the house.[1]
3. The number of split counties are minimized.
4. We minimize the occurrence of districts traversing county boundaries.

The second distribution includes a municipality score, $J_{MCD}(\xi)$. This score describes the number of people who have been displaced from a district that could have preserved the voters within their municipality, and is defined as

$$J_{MCD}(\xi) = \sum_{m \in M} \text{pop}_{oust}(\xi, m),$$

where M is the set of all MCDs, and $\text{pop}_{\text{oust}}(\xi, m)$ is the number of displaced people from the municipality m under the redistricting plan ξ. We define pop_{oust} in one way if the population of the municipality is less than the size of a district and another if it is greater.

[1]In the two exceptional clusters, it is impossible to draw districts that preserve precincts and also achieve population balance within 5%. For Wake in the senate, we sample with a deviation of 6% and generate an associated ensemble; past experience has shown that this does not create a partisan effect and we will be confirming this in follow on analyses. In Craven-Carteret, precinct 02 in Craven is the only precinct that connects the bulk of Craven with Carteret and it must be split to achieve population balance between the two districts within this cluster. We have examined the voting patterns when assigning this precinct to the district with the bulk of Craven or with all of Carteret and found minimal effects on the outcome.

If m has a population that is less than the population of a district, we consider the district that holds the most people from the municipality m as the representative district for that municipality. Any person within municipality m, but not within the representative district is considered to have been displaced.

If m has a population that is greater than the population of a district, we consider the number of districts that could fit within m to be $d(m) = \lfloor \text{pop}(m)/\text{pop}_{\text{ideal}} \rfloor$, where $\text{pop}(m)$ is the population of the MCD m and $\text{pop}_{\text{ideal}}$ is the ideal district population. We also consider the remaining population in the municipality that cannot fit within a whole district to be $r(m) = \text{pop}(m) - d(m) \times \text{pop}_{\text{ideal}}$. To determine the displace population, we look at the $d(m)$ districts that contain the largest populations from the municipality m. Hypothetically, everyone in these districts could live in municipality m. Therefore, anyone who is in one of these districts and that does not live in the municipality m could be replaced by someone who does live in the municipality. Thus, we sum the number of people not in m in the $d(m)$ districts that contain the largest populations of m. We also note that the remaining population $r(m)$ could hypothetically be kept intact when drawing a $(d(m)+1)$th district. We therefore look at the number of people in the municipality m who are living in the district with the $(d(m)+1)$th most population of the municipality. If the number of people in m is less than $r(m)$, then we add this difference to the number of ousted people (since each of these people in the municipality could have conceivably been placed in the district).

Formally, we let the $|M| \times d$ matrix, $MCD(\xi)_{m,j}$ represent the number of people who are in the municipality m and the district ξ_j. Then

$$\text{pop}_{\text{oust}}(\xi, m) \begin{cases} \sum_j MCD(\xi)_{m,j} - \max_j(MCD(\xi)_{m,j}) & \text{pop}(m) < \text{pop}_{\text{ideal}} \\ \sum_{j \in D(m)}(\text{pop}(\xi_j) - MCD(\xi)_{m,j}(\xi)) & \text{pop}(m) \geq \text{pop}_{\text{ideal}} \\ \quad + \max(0, MCD(\xi)_{m,N(m)} - r(m)) & \end{cases},$$

where $\text{pop}(\xi_j)$ is the population of district ξ_j, $D(m)$ is the set of district indices that represent the $d(m)$ districts with the largest populations of municipality m, and $N(m)$ represents the district index with the $d(m)+1$ most population of municipality m.

4 Sampling Method

We have chosen our the two distributions from which to draw our ensemble biased primarily biased on how their properties comply with the desired policy and legal considerations. Our primary motivation in choosing our distributions was not the ease of sampling from the distribution. 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 to a measure on redistricting plans that favors plans with a larger number of spanning trees. This alternative distribution satisfies the same constraints, however, it does not consider compactness nor municipal preservation. Furthermore, it can be effectively sampled using a variation on the metropolized multiscale forest RECOM sampling algorithm outlined in [2, 1]. We chose not to utilize this alternative ensemble as our base measure as (i) it does not correlate with the Polsby-Popper score, and (ii) there is some evidence that it introduces biases at the interface between dense and sparse population concentrations

In sampling the interpolating ladder of distributions between the easier-to-sample tree-based measure and our target measure on partitions, 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 a mixture of multiscale forest RECOM proposals and single node flip proposals. We will detail the algorithmic choices and our validation tests in later documents.

5 Construction of Statewide ensembles

Statewide ensembles are created by drawing samples from a number of "sub-ensembles." Because of the county cluster structure, we can sample each county cluster independently of the other county clusters. In the house, we sample the Wake and Mecklenburg county cluster groups separately from the rest of the state as they have many more precincts and districts.

In the senate, we sample the Wake county cluster independently since it must split precincts to achieve the 5% population balance. There are several regions of the state that have multiple options for county clusters and we sample each of the county clustering options separately. We then sample the remainder of the state together.

We combine these sub-ensembles by first choosing which of the county clustering options will be used, treating all options equally. With these fixed, we then choose a map from each of the other sub-ensembles and combine then to produce a statewide map. We used this procedure to create a ensemble of 100,000 maps. These ensembles of statewide maps were used to generate the various figures.

6 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] G. Herschlag, H. S. Kang, J. Luo, S. Bangia, C. V. Graves, R. Ravier, and J. C. Mattingly, "Quantifying gerrymandering in north carolina," *Statistics and Public Policy*, vol. Under review, 2020.

[4] C. Cooper, B. Esselstyn, G. Herschlag, J. Mattingly, and R. Tippett, "Nc general assembly county clusterings from the 2020 census," https://sites.duke.edu/quantifyinggerrymandering/files/2021/08/countyClusters2020.pdf, 2021.

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