DRAFT - The Ensemble Method and its use for Outlier Analysis in Redistricting[1]

Jonathan C. Mattingly and Gregory Herschlag

Duke University

What follows is the description of ensemble methods for creating and using a normative collection of maps. The ensemble of maps translates the specific redistricting design criteria into an understanding of what a redistricting plan should look like if the specified criteria are followed without ulterior motive. The criteria are formalized in a distribution on redistricting plans from which a sufficiently rich ensemble is drawn to capture the properties of the distribution. If the criteria are non-partisan then the ensemble gives a non-partisan normative collection against which other plans can be compared for their partisan and non-partisan properties.[2]

Highlights

(1) Ideally, maps should be sampled from a known distribution on redistricting plans which can be adapted to public policy goals.
(2) Sampling from a known distribution makes explicit the preferences used to create the ensemble of alternative redistricting plans and is an important step to minimizing unintended biases in the ensemble of maps created. The ensemble method is not simply creating a collection of alternative redistricting plans and then identifying outlier statistics.
(3) Regardless of what algorithm is used to generate the alternative redistricting plans (say sequential Monte Carlo, ReCom or other Markov Chain Monte Carlo method), we strongly advocate using the Metropolis-Hastings algorithm to adjust the output so the generated maps are sampled from the desired, policy driven distribution on redistricting plans. Failure to correct the output of Monte Carlo by use of Metropolis-Hastings increases the chance that unintended biases exist in the ensemble of maps produced.
(4) We advocate highlighting the structure of a particular map of interest through comparison with an ensemble to explain rather than just quoting probabilities or metrics.

Outline of Ensemble Method

We outline the use of Ensemble Methods to perform an outlier analysis of a particular map. The basic steps are as follows. Each is elaborated below:

(1) Establish the criteria and guidance for adhering to the criteria and any relative prioritization of the criteria.
(2) Create an explicit description of the distribution on redistricting plans from which the ensemble shall be drawn. Efforts should be made to not introduce biases to the distribution

---

[1] Version : April 13th, 2021.

[2] Here we have not discussed the methods based on the theorems originated by Chikina, Frieze, and Pegden. Those do not produce a normative ensemble but rather address if a particular plan is unusual with respect to several metrics by employing a sampling algorithm starting from the particular plan of interest. Its disadvantage is that it does not provide positive information about what is typical. Rather it only assesses if the particular plan is an outlier. It does however come with strong theoretical guarantees. The two methods work together; each favoring a slightly different perspective. The discussion of designing the distribution is relevant to both methods.

which are not stated explicitly. Describing the distribution explicitly is important for transparency, reproducibility, and ensuring that the choices made are open to being discussed.

(3) Draw an ensemble of plans from the distribution as faithfully as possible using currently available methods. A sufficient number of redistricting plans should be sampled from the distribution to ensure that the quantities reported are stable. Evidence of stability (e.g. convergence) should be presented to show the statistics are robust.

(4) Use the ensemble to report non-partisan statistics of interest under the ensemble. Examples include a measure of compactness, a measure of county or municipal preservation, typical BVAP percentages, and the expected strength of minority coalitions. These might be presented across the whole state or localize in a geographic region of interest.

(5) Use a representative collection of historical elections to present visualizations and statistics characterizing the typical partisan structure and performance of the ensemble of redistricting. This might include the typical partisan seat count or the typical partisan make-up of the collection of districts or the typical range of a particular partisan metric.

(6) Situate a particular map of interest in the distributions and visualizations reported in (4) and (5). Decisions might be made not to accept a proposed plan which deviates significantly from what is typical of the ensemble as reported in (4) and (5).

Expansion of above items

(1) Establish the criteria and guidance:
   a. This guidance might come by explicitly stating a particular distribution on redistricting plans, but more likely than not it would be stipulated by stating particular policy preferences. Examples include
      i. Limit parameters such as population deviation to a particular range. Example: The population deviation from the idea should be no more than 5%.
      ii. Sometimes a target value and a maximal standard deviation around the value could be specified. Target values might not be stated explicitly but rather in terms of a few historical maps whose properties in some regard are viewed as acceptable.
      iii. Sometimes a specific range is not appropriate as it depends on the particular geometry or population distribution of a state. In such cases, one might state that a quantity should be minimized to the extent possible while still achieving the other priorities listed (especially if indicated to be of greater importance).

(2) Create an explicit description of the distribution to be sampled:
   a. We emphasize that for the highest degree of transparency, we advocate explicitly stating the distribution from which the maps are to be sampled. Preferably this should be stated without reference to a particular sampling algorithm as algorithms that do not sample from an explicit distribution can have hidden implicit biases.
      i. As an example, consider an algorithm that grows a district starting from a randomly chosen precinct. It likely will make a difference if the random seed precinct is chosen uniformly over all precincts (and hence more likely to be in a city) or if a county is chosen first and then a precinct (and hence less likely, in many settings, to start in a city.)

b. One possibility is to say plans should be drawn with equal probability from the plans which are within some bounds. In some cases, this may lead to many plans being chosen just under the threshold.
c. Alternatively, one can specify that a particular quantity of interest should have a marginal distribution which is of the form of a Gaussian Distribution or Exponential distribution or another distribution of the Exponential family. The exact distribution of the marginal distributions may deviate because of the interactions with the state's geometry, spatial population distribution, and other constraints. Yet, such specifications give an effective way, long used in Bayesian statistics, to specify preferences about the distribution of plans.
d. It may be that initial attempts to translate the specifications and criteria into an explicit distribution may not achieve a distribution with the desired properties. This may be due to interactions between the different criteria and with the geometry and population distribution of the state. Hence, it is expected that the determining of a specific distribution on plans will be an iterative process where the most important policy imperatives may be given preference over other criteria.

e. The specification of the distribution should be as close to the policy goals as possible, without reference to specific algorithmic choices which do not have a clear policy interpretation. This helps prevent the introduction of unintended biases.

(3) With the distribution specified, it is then possible for many different actors to sample the distribution using whichever algorithm they feel is best.
   a. The algorithmic methods continue to evolve over recent years.
   b. It is possible, though increasingly unlikely given recent advances, that even with the improving state of the sampling one might not be able to convincingly sample from the distribution chosen to be most appropriate from policy considerations. In such a case, any compromise made will be made in the open where they can be assessed.
   c. The exact number of plans generated should be dictated by the number needed to sufficiently sample the quantities of interest. For example, if the partisan seat count histograms and the rank order marginal plots are to be produced, several independent chains should be run to show that the results are stable and that enough plans have been generated (see (5) below for an explanation of these plots).

(4) Non-partisan criteria. In most instances, there are numerous options. We present several examples:
   a. When comparing compactness across plans, options to consider include average, maximum and minimum compactness scores across all districts; comparisons that try to match districts geographically to create regionally localized hectographs; and creating summary statistics after ordering the districts from most to least compact.
   b. Controlling and assessing the splitting of counties and municipalities also presents many choices. In the previous redistricting cycle, for example, double district traversals across boundaries were undesirable for country boundaries while acceptable for municipal boundaries.
   c. Specificity in these statistics will further clarify redistricting goals and norms. For example, county preservation may include merely looking at the number of split counties. In this case, one would only report the number of split counties.

Alternatively, county preservation may consider how counties are split. In this case, one may prefer to split a county by keeping 95% of its residents in one district and 5% in another over a split that cuts the county down the middle; in this case, we can examine the extent of how counties are split.

(5) In determining what partisan criteria to measure, one should take care in choosing measures.
   a. Many measures can detect extreme behavior in retrospect but can be "gamed" if they are the known criteria ahead of time. Of the metrics to look at, we recommend partisan seats won under the ensemble over a variety of plausible vote counts. We also recommend not just looking at the partisan winners of a map, but also the margins of victory. This can be achieved through the use of rank-ordered marginal plots like used in Rucho. Some examples are linked here:
      i. https://sites.duke.edu/quantifyinggerrymandering/2019/12/02/court-chooses-to-stick-with-relatively-unresponsive-maps/
      ii. https://sites.duke.edu/quantifyinggerrymandering/2018/02/11/marginal-box-plots-summarizing-what-is-typical/
      iii. https://sites.duke.edu/quantifyinggerrymandering/2019/06/26/the-fix-is-in-the-votes-dont-matter/
   b. We are generally skeptical of measures or scores which only look at global vote counts or margins without any information of the region's political geography.
   c. One of the strengths of the ensemble method is that it refrains from invoking any grand philosophical principles about what is correct in an election but outside of the specification coming to our electoral system. Instead of invoking principles such as proportional representation or symmetries or competitiveness, it builds a normative ensemble which is informed by state geography, the details of the population distribution and the regulations, and legislative and legal norms around constructing a representative plan (which often vary from state to state).

Revision History

First version publicly released April 13th, 2021