Method Used for Measuring Redistricting Bias & Responsiveness

John F Nagle[1]

The method set out here is one used by the author in Election Law Journal 18, #1(2019) 63-77. The purpose of this write-up is to collect the various formulae in more easily readable format with line-by-line instructions to facilitate programming.

1. Obtain the number of Democratic (D) and Republican (R) votes in each block[2] in a state from one or more past statewide elections, such as president or senator.[3]

2. For each map the voter preference index (VPI) for each drawn district is calculated as follows:

   a. The *statewide* D vote share (V_D) is obtained. First separately sum the D votes and the R votes for all blocks in the state. Divide the D total by the two-party D plus R total. This is the two-party vote share for any data base in Step 1.[4]

   b. For each drawn district *n,* sum the D votes in all blocks in the district and separately sum the R votes in all blocks in the district. Divide the D total by the D total plus the R total to get the D vote share (v_{nD}) for the district.

   Each v_{nD} is an estimated D vote share for each district. It is an index of the voting preferences of voters in that district. We call v_{nD} a Voter Preference Index (VPI) for each drawn district.[5]

---

[1] The author thanks Alec Ramsay for suggesting this document and considerable help with it. Alec is a member of Dave's Redistricting DRA2020 team at http://gardow.com/davebradlee/redistricting/launchapp.html.

[2] "Block" means whatever units voting results are reported for. This is typically VTDs/precincts. To use the many census blocks, one would have to disaggregate the VTD data and to use census tracts one would have to aggregate the block data. We will use the typical VTD data as is done In DRA2.2 without aggregation/disaggregation.

[3] Examples that combine more than one past statewide election include Cook's PVI, Hofeller's formula, and my 7s PA data.

[4] In the future, one might also consider adding the Green and Socialist votes to the Democrats and the Libertarians to the Republicans to obtain a better estimate of progressive versus conservative.

[5] Note that a voter preference index may be obtained for each block in Step 1, but summing those does not give the *district* voter preference index because blocks have different numbers of voters.

Exhibit 1[6] shows the VPI for districts in Pennsylvania's recent SCOPA plan using Nagle's 7s data set.[7]

c. Compute a D seat probability for each district n as:

$$P(v_{nD}) = 0.5*(1+\text{prob}((v_{nD}-0.5)/0.04))$$

Unlike an outcome that gives the district either to D or to R, forecasts require probabilities. See Figure 1 below for this particular probability function.

Note: You can implement this formula, using the erf function as:

In FORTRAN: `0.5*(1+erf((v_nD-0.50)/(0.02*8**0.5)))`
In Python: `0.5*(1+erf((v_nD-0.50)/(0.02*sqrt(8))))`

d. Sum the probabilities over all districts to obtain the estimated statewide total number of D seats (S_D).

e. Together, the statewide D vote share (V_D) for the election result and number of seats (S_D) constitute an anchor point (V_D, S_D) on the map's seats/votes curve.
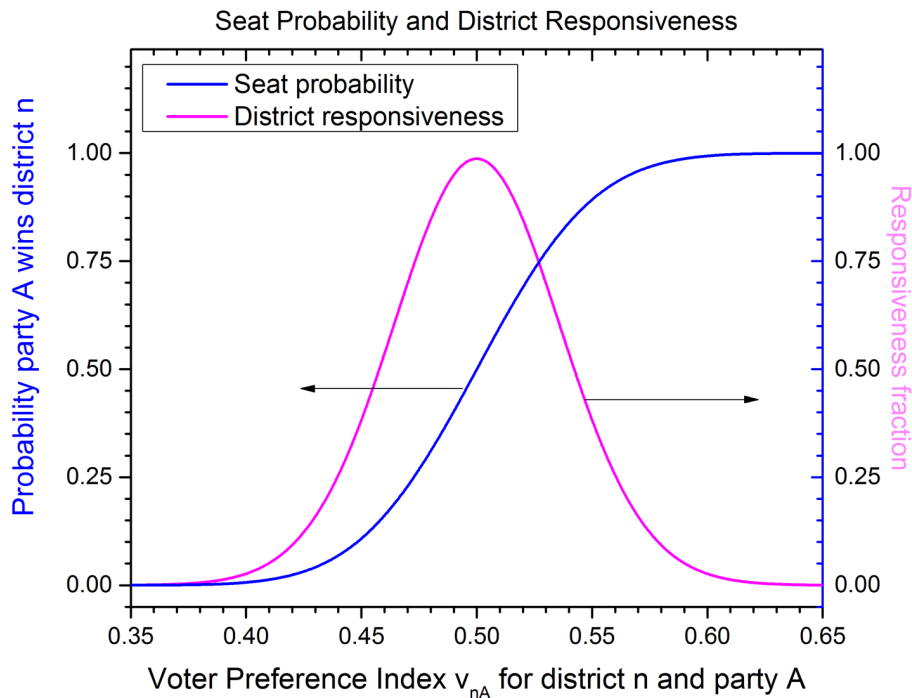
Seat Probability and District Responsiveness

---

[7] Of course, different data sets will have different numbers for the VPI. Fig. 2 in Nagle's paper gives a sense of the uncertainty from using different election results. Of course, combining many elections results generally decreases the uncertainties.

Figure 1

3. Infer a seats/votes S(V) curve for all statewide V (see <u>Figure 2</u> below). A shifted V′ requires a different v_{nD}' in each district. The proportional shift model for estimating the different v_{nD}' assumes equal probability for shifting any vote for a party anywhere in the state when the statewide V′ shifts towards the other party.[8]

   a. To proportionally shift D voters to R voters (decrease V_D), the formula is:

   $v_{nD}' = v_{nD}(V_D'/V_D)$ for $V_D' < V_D$.

   b. To shift R voters to D voters (increase V_D), the symmetric formula is:

   $v_{nR}' = v_{nR}(V_R'/V_R)$ for $V_R' < V_R$.

   Or, since D and R shares have to sum to 1, this can be expressed as:

   $v_{nD}' = 1 - (1 - v_{nD})(1 - V_D')/(1 - V_D)$ for $V_D < V_D'$.

   c. Use Eqs. (2c-d) to compute new D seat probabilities by district ($\mathtt{P(v_{nD}')}$) and the new likely number of D seats won overall ($\mathtt{S_D'}$).

   d. Vary the statewide vote share V to obtain the S(V) curve based on the underlying voting results.[9]

   <u>Figure 2</u> below shows a calculated S(V) curve.[10] Such an inferred S(V) curve graphically illustrates the measures of bias & responsiveness discussed next.

---

[8] This section abstracts the relevant portion of the Appendix in John Nagle's 2015 "Measures of Partisan Bias for Legislating Fair Elections" http://lipid.phys.cmu.edu/papers15/2015eljPub.pdf

[9] The entire range of V is readily calculable by computer.

[10] <u>Exhibit 2</u> shows the discrete points that were actually calculated for the example data set. From these one can easily interpolate the continuous S/V curve shown.
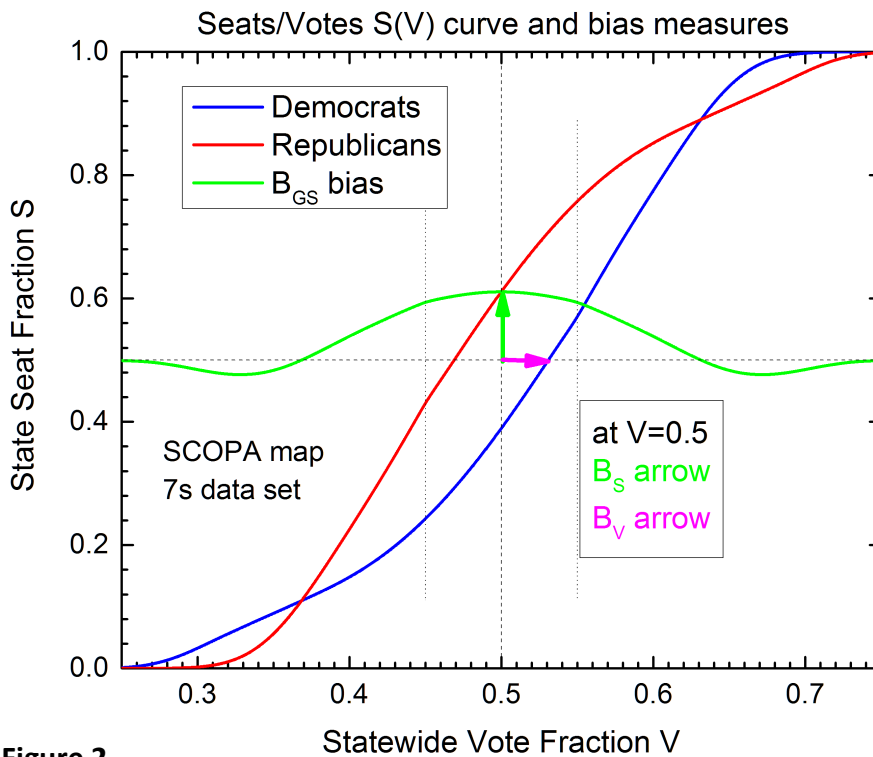
Figure 2

4. Measure the bias of the map in four ways:

    a. <u>A simple seats measure of bias</u>[11] – The number of R seats minus the number of D seats when the two-party vote is evenly split at V = 0.5, divided by twice the number of districts. Graphically, this is the vertical distance denoted 'B$_S$' in <u>Figure 2</u>.[12]

    b. <u>A better seats measure of bias</u>[13] – The number of R seats at vote $1 - V_D$ minus the number of D seats at vote V_D, again divided by twice the number of districts.[14]

---

[11] This is the simple S measure that was mostly reported in the Nagle paper for Pennsylvania because congressional elections in that state average close to $V_D = 0.5$.

[12] Using R – D is just a convention in which positive numbers mean Republican bias and negative ones Democratic.

[13] This is the B$_{GS}$ measure that was reported in Appendix B of the Nagle paper. It is generally better than the simple measure because it takes into account that the most probable vote in some states may be far from 0.5.

[14] This measure is based on the symmetry principle that each party should obtain the same number of seats when they receive the same vote. This measure is the same as the simple measure in (a) when V = 0.5.

c. <u>A simple votes measure of bias</u>[15] – The deviation from half of the vote needed to obtain half the seats. Graphically, this is the horizontal distance denoted 'B$_V$' in <u>Figure 2</u>.

d. <u>A better votes measure of bias</u>[16] – Half the difference between V$_D$ at which the estimated number of D seats is S$_D$ and the vote V$_R$' at which the estimated number of R seats is the same value S$_D$.

5. Measure the responsiveness of the map in several ways:

a. <u>The mathematical responsiveness R(V)</u> – This is the slope of the S/V curve at V.[17]

b. <u>The number of responsive districts (R$_d$)</u> – Sum this responsiveness function:

$$R(v_{nD}) \ = \ 1-4*(P(v_{nD})-V)^2 \qquad \text{over all districts n.}$$

c. In addition to the choice between (a) and (b), one may also consider both these measures of responsiveness at V = 0.5 or, better, at the most likely statewide V$_D$ vote.

<u>Exhibit 3</u> shows the bias and responsiveness metrics for the Pennsylvania SCOPA plan, again using the 7s data set.

The Exhibits can be found at <u>https://github.com/alecramsay/nagle</u>, as can sample Python code that implements this method and calculates these metrics.

---

[15] This is the simple V measure most reported in the Nagle paper.

[16] This is the B$_{GV}$ measure footnoted in Appendix B of my 2019 ELJ paper. Again, it is better because it takes into account that the most probable vote in some states may be far from 0.5. This measure is the same as the simple measure in (c) when V = 0.5.

[17] S and V are expressed as fractions in the S(V) curve. One interpolates what the slope is from the two discrete calculated vote values V$_1$ and V$_2$ that bracket the desired V value in order to calculate R(V) = (S(V$_2$)-S(V$_1$))/(V$_2$-V$_1$).