

Édouard Lucas:

The theory of recurrent sequences is an inexhaustible mine which contains all the properties of numbers; by calculating the successive terms of such sequences, decomposing them into their prime factors and seeking out by experimentation the laws of appearance and reproduction of the prime numbers, one can advance in a systematic manner the study of the properties of numbers and their application to all branches of mathematics.



MCMC Ensemble Methods

Daryl DeFord

MIT/Tufts
Metric Geometry and Gerrymandering Group

MGGG/LCCR Expert Witness Workshop
Radcliffe IAS Harvard
December 8, 2018



Outline

- ① Introduction
- ② Ensemble Methods
- ③ Scoring Plans
- ④ MCMC
- ⑤ Postprocessing
- ⑥ Results



Why ensembles?



Outliers

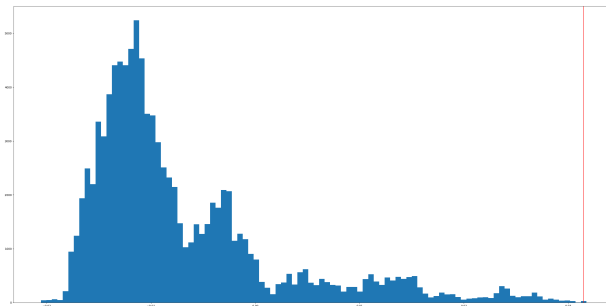


Figure: Mean–Median score using senate 2016 election data on 1,000,000 plans.



Which ensembles?



References

- Jonathan Mattingly (Duke)
 - Redistricting Drawing the Line ([arXiv: 1704.03360](#)
Gregory Herschlag, Robert Ravier, Jonathan C. Mattingly)
 - Evaluating Partisan Gerrymandering in Wisconsin ([arXiv: 1709.01596](#)
Sachet Bangia, Christy Vaughn Graves, Gregory Herschlag, Han Sung Kang, Justin Luo, Jonathan C. Mattingly, Robert Ravier)
- Kosuke Imai (Princeton/Harvard)
 - A New Automated Redistricting Simulator Using Markov Chain Monte Carlo ([Preprint](#) Benjamin Fifield, Alexander Tarr, Michael Higgins, and Kosuke Imai)
 - redist: Markov Chain Monte Carlo Methods for Redistricting Simulation ([R package](#))



MCMC for redistricting

- 1 Start with an initial plan
- 2 Propose a modification of the current plan
- 3 Accept using MH criterion
- 4 Repeat



MCMC for redistricting

- 1 Start with an initial plan
- 2 Propose a modification of the current plan
- 3 Accept using MH criterion
- 4 Repeat

Why?



MCMC for redistricting

- 1 Start with an initial plan
- 2 Propose a modification of the current plan
- 3 Accept using MH criterion
- 4 Repeat

Why?

- Control over sampling distribution
- Possibility of local sampling
- Ergodic Theorem



Procedural Outline

- Operationalize Laws
- Proposals
- Try to sample (MCMC)
- Tune score functions
- Adjust temperature
- Winnow ensemble
- Analyze



MORAL:

Computational Redistricting is
NOT a solved problem!



Operationalizing Laws

- Contiguity
- Population Balance
- Compactness
- County Splitting
- VRA compliance



Population Balance

-

$$\sqrt{\sum \left(\frac{p_i}{\bar{p}} - 1 \right)^2}$$

-

$$\sum \left| \frac{p_i}{\bar{p}} - 1 \right|$$

-

...



Compactness

•

$$\sum \frac{\text{Perimeter}(D_i)^2}{\text{Area}(D_i)}$$

•

$$\sum_{k=1}^n \sum_{i,j \in D_k} p_i p_j d(i,j)^2$$

•

...



County Splits

The number of counties split two ways times the sum over split counties of the squareroot of the fraction of counts VTDs in 2nd largest intersection of a district with the county **plus** a large constant times the number of triply split counties times the sum over triply split counties of the squareroot of the fraction of county VTDs not in 1st or 2nd largest intersection of a district with the county.



VRA Compliance

$$\sqrt{H(44.48 - D_1)} + \sqrt{H(36.2 - D_2)}$$



MORAL:

Computational Redistricting is
NOT a solved problem!



Proposals

- Boundary edge flip
 - Uniformly select an edge between two districts and change one endpoint
- Modified Swendsen–Wang
 - Activate edges, form clusters, then partition



Energy Functions

- Weight plans proportional to $e^{-\beta \sum w_i \text{scores}(D_i)}$
- Varying β controls the strictness of the constraints
- Varying w_i changes the relative strengths of the scores
- Exploit vs. Explore



Temperature Variations

- Simulated Annealing
 - Start at a high temperature and then gradually
 - 40,000 - 60,000 - 20,000
- Parallel Tempering
 - Run several chains at different temperatures and swap
- Simulated Tempering
 - Extra MH step to update β value



Winnowing (Individual Districts)

- Strict 1% population bound
- Strict compactness bound
- Strict VRA bounds
- No triply split counties



Validation

- Longer runs
- Longer annealing
- Different starting points
- Different winnowing
- Other weight choices
- MCMC Heuristics



Validation

- Longer runs
- Longer annealing
- Different starting points
- Different winnowing
- Other weight choices
- MCMC Heuristics

:)



Summary

- North Carolina 24,000 plans
 - Outlier Analysis
 - “Signature of gerrymandering”
- Complete enumeration
 - Converges to target distribution
 - Local walks to perturb for fairness



MORAL:

Computational Redistricting is
NOT a solved problem!



Next steps

- Direct comparison of methods
- Sensitivity Analysis
- More complete description of space of plans
- ...



The End

Thanks!

