



Understanding Our Markov Chain Significance Test: A Reply to Cho and Rubinstein-Salzedo

Maria Chikina, Alan Frieze & Wesley Pegden

To cite this article: Maria Chikina, Alan Frieze & Wesley Pegden (2019) Understanding Our Markov Chain Significance Test: A Reply to Cho and Rubinstein-Salzedo, *Statistics and Public Policy*, 6:1, 50-53, DOI: [10.1080/2330443X.2019.1615396](https://doi.org/10.1080/2330443X.2019.1615396)

To link to this article: <https://doi.org/10.1080/2330443X.2019.1615396>



© 2019 The Authors. Published with license by Taylor & Francis Group, LLC



Accepted author version posted online: 13 May 2019.
Published online: 26 Jun 2019.



Submit your article to this journal [↗](#)



Article views: 222



View related articles [↗](#)



View Crossmark data [↗](#)

Understanding Our Markov Chain Significance Test: A Reply to Cho and Rubinstein-Salzedo

Maria Chikina^a, Alan Frieze^b, and Wesley Pegden^b

^aComputational and Systems Biology, University of Pittsburgh, Pittsburgh, PA; ^bDepartment of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA

ABSTRACT

The article of Cho and Rubinstein-Salzedo seeks to cast doubt on our previous paper, which described a rigorous statistical test which can be applied to reversible Markov chains. In particular, Cho and Rubinstein-Salzedo seem to suggest that the test we describe might not be a reliable indicator of gerrymandering, when the test is applied to certain redistricting Markov chains. However, the examples constructed by Cho and Rubinstein-Salzedo in fact demonstrate a different point: that our test is not the same as another class of gerrymandering tests, which Cho and Rubinstein-Salzedo prefer. But we agree and emphasized this very distinction in our original paper. In this reply, we reply to the criticisms of Cho and Rubinstein-Salzedo, and discuss, more generally, the advantages of the various tests available in the context of detecting gerrymandering of political districtings.

ARTICLE HISTORY

Received April 2019
Accepted April 2019

KEYWORDS

Gerrymandering; Markov chain; Redistricting

1. Introduction

The article *Understanding Significance Tests From a Non-mixing Markov Chain for Partisan Gerrymandering Claims* by Cho and Rubinstein-Salzedo (2019) offers commentary on our previous paper *Assessing Significance in a Markov Chain Without Mixing* (Chikina, Frieze, and Pegden 2017). In 2017, one of us (Pegden) served as an expert witness in the case *League of Women Voters v.s. Pennsylvania*, which ultimately overturned the Pennsylvania Congressional districting. Pegden gave testimony which leveraged the statistical test developed in our paper to make a rigorous claim of gerrymandering in Pennsylvania. Cho also served as an expert witness in this case, hired by the legislature to respond to the plaintiffs' experts' testimony.

Here we take the time to briefly reply to the points raised in the article by Cho and Rubinstein-Salzedo.

Broadly speaking, it seems that Cho and Rubinstein-Salzedo wish to cast doubt on whether one can or should use our test to deduce that a political districting has been gerrymandered. However, rather than present evidence that non-gerrymandered maps can systematically fail our test, they rather demonstrate merely that their preferred test for gerrymandering can sometimes report different answers from ours, on some specific maps. To demonstrate a problem with our test, Cho and Rubinstein-Salzedo would have to either:

1. Demonstrate that random maps can fail our test at a rate in excess of the p -value computed by our test (contradicting our theorem in Chikina, Frieze, and Pegden (2017)), or

2. Demonstrate how a non-partisan process of choosing maps would preferentially select maps which fail our test.

Cho and Rubinstein-Salzedo have done neither of these.

2. Yes Our Test Is Different!

While Cho and Rubinstein-Salzedo acknowledge that “the Supreme Court has yet to accept a particular quantifiable gerrymandering test,” their article nevertheless treats their preferred method for detecting gerrymandering as a gold standard, against which other methods should be judged. In particular, Cho and Rubinstein-Salzedo advocate evaluating districtings by:

1. Using an algorithm to draw random districtings of a state satisfying certain constraints, which is hoped to select maps from a suitable distribution on the entire space of possible districtings, and
2. Comparing the partisan qualities of the given districting to those in the random sample.

Our test on the other hand works as follows:

1. Begin with the districting being evaluated;
2. Carry out a sequence of random changes to the map, while preserving the desired constraints;

CONTACT Wesley Pegden  wes@math.cmu.edu  Department of Mathematical Sciences, Carnegie Mellon University, 5000 Forbes Ave, Wean Hall 7105, Pittsburgh, PA 15203-3815.

¹Even if both tests were identical, this would be no surprise, as both tests are probabilistic tests for which p -values govern their probability of failure. Indeed, since applying the tests involve the use of randomness, repeated applications of our test to a fixed map could even disagree with each other, but this would be a rare event by controlled by our reported p -value.

© 2019 The Authors. Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

3. If the partisan bias of the districting dramatically dissipates in the sequence of random changes, so that the current districting is in the most extreme ε fraction of observed maps with respect to partisanship, the districting is “carefully crafted” with respect to partisan bias, indicating intentional gerrymandering.
4. The theorem from Chikina, Frieze, and Pegden (2017) computes a p -value for this observation, bounding the probability that a randomly chosen districting from our chosen distribution² on maps would exhibit partisan bias as fragile as the given districting.

Cho and Rubinstein-Salzedo provide evidence to demonstrate that these tests are not the same (i.e., they can, in principle produce different answers on specific maps). But a plain reading of our paper (Chikina, Frieze, and Pegden 2017) emphasizes the distinction between what we achieve and random sampling.³

3. Cho and Rubinstein-Salzedo Have Not Explained Why It Is Worse

Though it seems Cho and Rubinstein-Salzedo wish to cast doubt on whether our method should be trusted to infer that a map was intentionally drawn making excessive use of partisan data, they have in fact only worked to show that our test is *different* from global sampling tests (a fact which we ourselves have emphasized).

Of course, the relevant question is not the relationship between our gerrymandering test and the preferred test of Cho and Rubinstein-Salzedo, but instead the relationship between our test and detecting the actual practice of intentionally drawing political lines with excessive use of partisan considerations. In particular, the question is: “Is there some way that nonpartisan districting processes will systematically fail our test, at a rate not captured by our p -value?” Of course, the same question can be asked about the global sampling test advocated by Cho and Rubinstein-Salzedo. Gerrymandering, after all, is an intentional act by human beings; how much we can trust a particular quantitative test designed to detect

it hinges not on how similar it is to the preferred test of Cho and Rubinstein-Salzedo, but on how likely the test is to call a districting gerrymandered, when humans have not actually intentionally drawn a districting in a way which excessively leaned on partisanship.

To be clear, what it means to trust our test is to believe that when a districting fails our test, the failure is one of the following two modes:

Mode 1: The districting was drawn with partisan considerations;

Mode 2: Drawing a map which failed our test was a rare event whose probability is controlled by the p -value computed by our theorem.

For example, in their discussion of their Figure 1, Cho and Rubinstein-Salzedo point out that if we evaluated a districting with a partisanship score 0.16 in the first disconnected subspace of the chain they have constructed, then our test might report the districting as gerrymandered.⁴ The existence of such a map agrees perfectly with the framework of our test: drawing such a map with a nonpartisan districting process is possible, but very unlikely, and would constitute a Mode 2 failure of the test. In general, one cannot argue against the validity of a statistical test like ours with individual counterexamples; our rigorous claims are about the *probabilities* of failure, not *impossibility*.

To suggest that our test may not be a reliable indicator of gerrymandering despite accepting the correctness of our theorem in Chikina, Frieze, and Pegden (2017), one would necessarily have to believe that there can be another mode of failure for our test:

Mode 3: There is some systematic reason other than partisan considerations which leads mapmakers to preferentially draw maps which fail our test.

In their article, Cho and Rubinstein-Salzedo do not suggest what they think this “third mode” of failure for our test could be, nor do they demonstrate that this occurs in practice. In particular, “What non-partisan districting process should preferentially create political districtings for which random changes to the boundary lines have a dramatic, consistent partisan effect?” Indeed, it seems particularly hard to imagine how such systematic biases could be likely to affect our test, but *not* the kind of test favored by Cho and Rubinstein-Salzedo.

To put this in the context of the Pennsylvania trial, Pegden testified there that when applying our method to the Congressional districting of Pennsylvania, we made a sequence of roughly 1 trillion random changes to the Pennsylvania map, preserving traditional districting criteria such as contiguity, compactness, etc., and found that the existing Congressional map was more partisan than 99.999999% of maps produced in the sequence of small random changes. Our theorem asserted statistical significance for this observation at $p < 0.00005$. For some runs of our test, we even observed the remarkable phenomenon that *every single map* produced in the sequence of random

²In their analysis, Cho and Rubinstein-Salzedo fail to distinguish our test from the choice of Markov Chain it is applied to; it can be applied to any reversible Markov chain, which allows one to select which distribution to use in evaluating gerrymandering claims. In particular, if Cho and Rubinstein-Salzedo prefer to apply our test using a Markov Chain which does not have a disconnected state space, they are more than welcome to do so, even if the statistical validity of our test is not affected by this question.

³Instead of quoting from our paper (Chikina, Frieze, and Pegden 2017), Cho and Rubinstein-Salzedo write that “Pegden advocated the CFP theorem as a test for whether a disputed map is an outlier ‘among all possible legal maps’” in the Pennsylvania lawsuit. And indeed, in hours of testimony under oath, Pegden discussed our test in excruciating detail, explaining the precise relationship between ε and p , and the precise nature of the statistical claims our test can make. But the test was not presented as equivalent to a global sampling test. Indeed, as in our original paper, this distinction was emphasized Pegden’s testimony, for example, on pages 749–752 of the trial transcript. This distinction was also emphasized in great detail by Pegden during his cross-examination, as can be found in pages 790–803, beginning with Pegden asking the cross-examining attorney: “Would you like me to describe the difference between a traditional Markov chain analysis and what I do?” (Trial Transcripts 2018).

⁴In actual fact, it does not seem they have succeeded at constructing an example for which our test would produce statistical significance, as for any map in the cluster of size 486, we could optimistically only hope to achieve $p = \sqrt{2/486} \approx .064$ with our theorem.

changes to the Congressional map was fairer than Congressional map itself,⁵ showing the extreme degree to which the districting was optimized with respect to partisan characteristics. Despite wishing to cast doubt on the use of our test to infer legal claims of gerrymandering, Cho and Rubinstein-Salzedo have offered no explanation for what process would prefer to draw maps failing our test as spectacularly as this, short of the intentional and excessive use of partisanship in the districting process.

4. Which Test Is Better?

Even if our test has no “third mode” of failure, one could still ask: are there reasons to *prefer* our test to the kind of global sampling tests advocated for by Cho and Rubinstein-Salzedo? In particular, does it have advantages beyond its validity?

One key advantage of our test is that it can be carried out without any unproven assumptions on the sampling method being used. In particular, there is currently no practical sampling technique known which can provably draw random districtings of a state from a specified distribution with provable bounds on the sampling error. Carrying out a test of the type advocated by Cho and Rubinstein-Salzedo thus involves, for example, heuristic tests of the quality of samples obtained from the sampling algorithm. Our method, by contrast, can be applied in a statistically rigorous way to any reversible Markov Chain, without requiring unproven assumptions on the mixing time of the chain.

However, we would go further, and argue that there are good reasons to afford more trust to gerrymandering detected by our test, separate from any issue of the lack of mathematical rigor underlying global tests. In particular, a global sampling technique of the type advocated for by Cho and Rubinstein-Salzedo is based on the premise that mapmakers gerrymander a map by solving a *global optimization problem*. That is, in the ideal framework to motivate their test, there is some set of constraints on valid districtings, defining a large set of valid possible districtings, and mapmakers gerrymander the map by finding a global optimum (or at least a global outlier) with respect to partisanship. If this was really how gerrymandering worked, and global sampling tests could be carried out rigorously, then we would advocate using global tests exclusively.

But of course, gerrymandering does not work like this. The space of feasible maps is complicated, and just as there is still no practical way for people who are trying to detect gerrymandering to neatly characterize all of this space, there is likewise no practical way for mapmakers trying to gerrymander a map to find a global optimum in this space. In particular, viewed as a planar-graph partitioning problem, any reasonable precise formulation of the gerrymandering problem would be NP-hard.⁶ Even in a hypothetical world where this global optimization

problem could efficiently be solved, the reasonable goal of minimizing disturbances to previous districts (a districting criteria embraced by the US Supreme Court in *Karcher v. Daggett*) would undercut such an effort.

In lieu of solving a global optimization problem, then, mapmakers can instead solve at least a *local* optimization problem—they can at least iteratively change a map to maximize their partisan advantage, until it can hardly be improved. The most reliable signature of a gerrymandered districtings, then, may well be exactly the kind of partisan fragility of a districting that our test is based on. In particular, this kind of reasoning explains why even groups that are at the cutting edge of developing trustworthy *global* samplers are nevertheless sometimes also interested in using those samplers to examine *local redistrictings* of a state (Herschlag et al. 2018).

5. There Is Not One Right Answer

Despite what we feel are strong arguments in favor our test as a reliable indicator of gerrymandering, there is no reason that one test needs to be used to the exclusion of others. Instead, when asking a court to intervene in the process of drawing districting lines, it seems especially important to be able to present multiple kinds of evidence that a districting is gerrymandered. To us, it seems that global tests of the kind advocated for by Cho and Rubinstein-Salzedo can indeed be a strong piece of evidence to bring. In particular, global sampling methods have the advantage that they can provide more information than our test can. Our approach excels at rigorously demonstrating that a districting was intentionally gerrymandered. But it cannot always answer other important questions that might arise in gerrymandering cases, such as: what properties should a neutral districting of a given state have?

Global samplers are especially trustworthy when implemented as by the Mattingly group, where the algorithm is designed to sample from a precise distribution, and its performance can be experimentally verified. This has the advantage of plainly presenting design decisions, and allows consumers of the test (the public and the court, as well as experts) to consider the potential for unintentional biases in the test design. But even when it is applied as by Cho and coauthors as in, for example, Liu, Cho, and Wang (2016), without characterizing a precise stationary distribution the practitioner hopes sampling is being done from, the simple practice of comparing a potentially gerrymandered map to maps drawn by a nonpartisan algorithm is a simple and intuitive approach.

References

- Chikina, M., Frieze, A., and Pegden, W. (2017) “Assessing Significance in a Markov Chain Without Mixing,” *Proceedings of the National Academy of Sciences of the United States of America*, 114, 2860–2864. [50,51]
- Cho, W. K. T., and Rubinstein-Salzedo, S. (2019), “Understanding Significance Tests From a Non-mixing Markov Chain for Partisan Gerrymandering Claims” (to appear), *Statistics and Public Policy*. [50]
- Frieze, A., and Dyer, M. (1985), “On the Complexity of Partitioning Graphs Into Connected Subgraphs,” *Discrete Applied Mathematics*, 10, 139–153. [52]

⁵Note that the first new map in the sequence differs from the initial map at just one voter precinct, the next map by at most 2 precincts, etc., so even tiny changes to the districting consistently decreased its partisan bias.

⁶For example, even the simple decision problem of deciding whether a given planar graph can be partitioned into k connected subgraphs of equal size is NP-complete (Frieze and Dyer 1985).

- Herschlag, G., Kang, H. S., Luo, J., Graves, C. V., Bangia, S., Ravier, R., and Mattingly, J. C. (2018), “Quantifying Gerrymandering in North Carolina,” arXiv no. 1801.03783. [52]
- Liu, Y. Y., Cho, W. K. T., and Wang S. (2016), “PEAR: A Massively Parallel Evolutionary Computation Approach for Political Redistricting Optimization and Analysis,” *Swarm and Evolutionary Computation*, 30, 78–92. [52]
- Trial Transcripts, “League of Women Voters of Pennsylvania v. Commonwealth of Pennsylvania,” Brennan Center, available at <https://www.brennancenter.org/legal-work/league-women-voters-v-pennsylvania/>. The quoted pages are from testimony on December 13, whose transcript is available at https://www.brennancenter.org/sites/default/files/legal-work/LWV_v_PA_TrialTranscript_12.13.17.pdf. [51]