

Understanding Significance Tests from a Non-Mixing Markov Chain for Partisan Gerrymandering Claims

Wendy K. Tam Cho*

Departments of Political Science, Statistics, Mathematics, and Asian American Studies,
the College of Law, and the National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

Simon Rubinstein-Salzedo

Euler Circle, Palo Alto, CA

Abstract

Recently, [Chikina, Frieze and Pegden \(2017\)](#) proposed a way to assess significance in a Markov chain without requiring that Markov chain to mix. They presented their theorem as a rigorous test for partisan gerrymandering. We clarify that their ε -outlier test is distinct from a traditional global outlier test and does not indicate, as they imply, that a particular electoral map is associated with an extreme level of “partisan unfairness.” In fact, a map could simultaneously be an ε -outlier and have a typical partisan fairness value. That is, their test identifies local outliers but has no power for assessing whether that local outlier is a global outlier. How their specific definition of local outlier is related to a legal gerrymandering claim is unclear given Supreme Court precedent.

Keywords: Markov Chain Monte Carlo; Redistricting; Simulation

*This research has been funded in part by the National Science Foundation (Grant No. SES-1725418/1728902) and the Guggenheim Foundation, as well as multiple computing allocation grants on the Blue Waters sustained-petascale computing resources, which is supported by the NSF (Grants OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and the National Center for Supercomputing Applications.

1 Introduction

An important problem in probability and statistics, with extensive applications, is determining how to sample from complicated probability distributions that may not be explicitly describable. The standard method for sampling from an unknown distribution is the set of Markov Chain Monte Carlo (MCMC) techniques, including, for example, the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970), the Gibbs Sampler (Geman and Geman, 1984), and coupling from the past (CFTP) (Propp and Wilson, 1996).

A sequence of random variables that take on values in a state space is called a *Markov chain* if the probability of the next step depends only on the current state,

$$\mathbb{P}(x_n \mid x_{n-1}, \dots, x_1) = \mathbb{P}(x_n \mid x_{n-1}). \quad (1.1)$$

If the Markov chain is *irreducible*, i.e. every state is accessible in a finite number of steps from any other state, and *positive recurrent*, i.e. the expected return time to any state is finite, then its stationary distribution, π , is the unique probability distribution satisfying $\sum_i \pi(i)P_{ij} = \pi(j)$, where $P_{ij} = \mathbb{P}(x_n = j \mid x_{n-1} = i)$. If we can construct such a Markov chain that has the unknown distribution as its stationary distribution, and we run this chain for a sufficiently long time to achieve *mixing* (i.e., the chain has approached a state of equilibrium), then each state visited by the chain is close to a representative sample of the underlying distribution.

A current area of interest, where Markov chains could prove useful, is in the area of redistricting. It is conjectured that some electoral maps are gerrymandered. These suspicions are fueled by several observations. First, many of these maps were devised by partisan and self-interested legislators. Second, there are large discrepancies between the proportion of seats won by a certain party and the proportion of the statewide vote won by the same party. Despite the U.S. system not being one of proportional representation (PR), many intuitive and normative notions of fairness are violated when there is a large deviation between the results and PR.

The Supreme Court has declared that partisan gerrymandering is unconstitutional. However, it has yet to identify a manageable standard for adjudicating partisan gerrymandering.

dering. That is, the Court does not know how to distinguish a partisan gerrymander from a constitutional electoral map. It has stated that partisan information may be used in the construction of a map, and that “[t]he central problem is determining when political gerrymandering has gone too far.” That is, “the issue is one of how much is too much.” (*Vieth v. Jubelirer*, 541 U.S. 267 (2004)). The implication is that maps that use partisan information *excessively* are unconstitutional. This has led to the idea that one tactic for challenging the constitutionality of electoral maps is to demonstrate that a disputed map has a partisan outcome that is “not typical” of the set of all maps that could have been drawn. In statistics, one way to help us understand what is typical and atypical is to construct the underlying population distribution, which may be done for redistricting by sampling the space of legally viable maps.

In redistricting, a state is allotted some number of districts, say d . Each state has $N > d$ precincts, and a district is a certain partition of precincts into districts.¹ Districts are almost always required by law to be contiguous and equi-populous. One way to define equi-populous is that for a certain $\varepsilon > 0$, for each pair i and j of districts, if p_i and p_j denote the populations of districts i and j , then $\left| \frac{p_i}{p_j} - 1 \right| < \varepsilon$. The law also requires compliance with the Voting Rights Act, which ensures minority representation. In addition to these legal constraints, the Court has articulated a number of “traditional districting principles,” for which it values adherence (e.g. compactness and the preservation of political subdivisions and communities of interest).

The state space of all maps that satisfy the Court’s criteria is the extremely complicated state space from which we would like to be able to sample uniformly. A simpler (but much larger) space is the space of maps with just contiguous districts, i.e. partitions of the precincts into d districts, S_1, \dots, S_d , where the only requirement is that each precinct is contained in exactly one of the S_i ’s, and each S_i is a connected subset of the entire electoral jurisdiction. This space may not be the space of all legal and viable maps if there are other legal criteria that must be taken into account, but it may provide a useful starting point. For instance, an idea for sampling legal districts is to begin with some partition into

¹A precinct is the lowest level of geographic aggregation where election data is available. One may use smaller geographic units (e.g. census blocks) for redistricting, but then the election results must be estimated, which is not ideal for partisan gerrymandering claims since the estimation induces another source of uncertainty and error.

connected regions (e.g. the challenged map), and then use an MCMC model to explore the space of legal districts. This approach is taken in several papers ([Bangia et al., 2017](#); [Herschlag, Ravier and Mattingly, 2017](#); [Mattingly and Vaughn, 2014](#); [Fifield et al., 2017](#)).

2 The Chikina–Frieze–Pegden Test

Recently, [Chikina, Frieze and Pegden \(2017\)](#) (CFP) proposed another approach for assessing whether a disputed map might have used partisanship excessively. They do not attempt to produce a representative sample of legal maps, but rely instead on the existence of a reversible Markov chain to provide the basis for an outlier test. They prove the following theorem.²

Chikina, Frieze, and Pegden’s Theorem. *Let $\mathcal{M} = (M_0, M_1, \dots)$ be a reversible Markov chain with stationary distribution π . If M_0 is distributed according to π , then for any fixed ε and k , the probability that M_0 is an ε -outlier among M_0, M_1, \dots, M_k is at most $\sqrt{2\varepsilon}$.*

They seek to apply this theorem in the redistricting context. In their paper, they discuss how this test might be applied to the redistricting of congressional districts in Pennsylvania. They provide the Pennsylvania example to illustrate how their test might work, but do not delve deeply into how they propose to translate their mathematical theorem into a formal test for gerrymandering in a court of law. However, we can gain an accurate sense for how CFP intend to translate their theorem into the legal context through the expert witness report of one of the authors, Pegden, who testified under oath in Pennsylvania state court in a case challenging the Pennsylvania map as a partisan gerrymander ([Pegden, November 27, 2017](#)). In particular, Pegden advocated the CFP theorem as a test for whether a disputed map is an outlier “among all possible legal maps.” Pegden writes on the first page of his expert report, “In my analysis, I find that the present Congressional districting of Pennsylvania is indeed a gross outlier with respect to partisan bias, among the set of *all possible districtings of Pennsylvania* (emphasis added).” The word “all” in this statement

²A Markov chain is reversible if there exists a probability distribution, π , such that $\pi_i \mathbb{P}(X_{n+1} = j \mid X_n = i) = \pi_j \mathbb{P}(X_{n+1} = i \mid X_n = j)$, for all states i and j .

makes the comparison set the space of all possible electoral maps, which implies that the CFP theorem is able to identify a global outlier with respect to some partisan metric for maps.

To make the adaptation of the CFP theorem to redistricting, they suppose that we have a function $\omega : X \rightarrow \mathbb{R}$, which can be taken to be an arbitrary function. Here, X is the set of electoral maps. The function ω is used to determine what constitutes an outlier, since it is straightforward to order elements in \mathbb{R} , but an arbitrary state space may not have a convenient structure. That is, it may be simple to order elements in \mathbb{R} , but how to order electoral maps is less clear. For a real number $\varepsilon > 0$, we say that a real number, α_0 , is an ε -outlier among a sequence of $k + 1$ numbers, $\alpha_0, \alpha_1, \dots, \alpha_k$, if

$$\#\{i : 0 \leq i \leq k : \alpha_i \leq \alpha_0\} \leq \varepsilon(k + 1).$$

Less formally, α_0 is an outlier if we rarely encounter a number as small as α_0 in the observed sequence, $\alpha_0, \alpha_1, \dots, \alpha_k$. In the partisan gerrymandering context, ω would be some measure of “partisan unfairness.” Pegden uses a mean-median test as well as a variance ratio test for his analysis for the court, but the measure is immaterial to the test and could just as well be, for example, the efficiency gap, the number of seats, or measures derived from the seats-votes curve (Grofman and King, 2007; Stephanopoulos and McGhee, 2015).

To translate back to the Markov chain framework, given the Markov chain \mathcal{M} , we say that M_0 is an ε -outlier among M_0, M_1, \dots, M_k if and only if $\omega(M_0)$ is an ε -outlier among $\omega(M_0), \omega(M_1), \dots, \omega(M_k)$. That is, a particular map, M_0 , is an ε -outlier among the set of observed maps if its measure of partisan unfairness, α_0 , is an ε -outlier among the associated partisan metrics of those maps. CFP present the ε -outlier test as a rigorous way to detect gerrymandering.

2.1 The Redistricting State Space

We examine the proposal to employ the ε -outlier test as a way to detect partisan gerrymandering. While the Supreme Court has yet to accept a particular quantifiable gerrymandering test, they are clear that “judicial action must be governed by *standard*, by *rule*.”

Here, the rule or test would help determine “when political gerrymandering has gone too far” (*Vieth v. Jubelirer*, 541 U.S. 267 (2004)).

The CFP test appears promising because the theorem is concerned with identifying “outliers,” and, in particular, what CFP call “ ε -outliers.” The Court’s language—“how much is too much”—seems consistent with the idea that if something can be deemed an “outlier,” then there is a logical way to understand that as “too much.” The Court has, in addition, referred to “an extremity of unfairness,” which also seems to translate nicely to the statistical notion of an outlier.

To examine the relationship between the CFP ε -outlier and the legal definition of a partisan gerrymander, we begin with an observation about the state space that characterizes the redistricting application. Let us write \mathcal{P} for the set of precincts and $[d] = \{1, 2, \dots, d\}$ for the set of districts. Precincts $p_i, p_j \in \mathcal{P}$, $i \neq j$, are geographically adjacent if and only if they share a border of positive length. The state space X is the set of electoral maps, which are functions $f : \mathcal{P} \rightarrow [d]$ satisfying various constraints like contiguity and equipopulousness.³ Furthermore, we only consider functions up to relabeling of the districts.

To transition from one state to the next, for each state $f \in X$, we compute a proposal distribution θ_f on X , depending on f . If $M_n = f$, then in order to determine M_{n+1} , we sample $g \in X$ according to θ_f . The CFP chain accepts the proposal and moves, $M_{n+1} = g$, if this results in a legally viable map, and rejects the proposal otherwise.

The CFP chain does not allow movement into infeasible regions at all. That is, the CFP chain allows movement from one map to another only when the movement of one geographic unit results in another feasible map (where feasible is defined as being a map that satisfies the legal criteria imposed on electoral maps). For any redistricting application, the feasible regions vary in size in some unknown manner. The CFP chain would not be able to move throughout the state space, though CFP state that their theorem “applies even if the chain is not irreducible (in other words, even if the state space is not connected), although in this case, the chain will never mix ([Chikina, Frieze and Pegden, 2017](#), p. 2861).” In other words, the CFP chain would not be able to visit all states, which CFP acknowledge, but

³The specific criteria for “legal maps” can vary from state to state. We make no claims about how a legal map should be defined. Instead, here, “legal maps” refers only to maps that must satisfy some set of constraints.

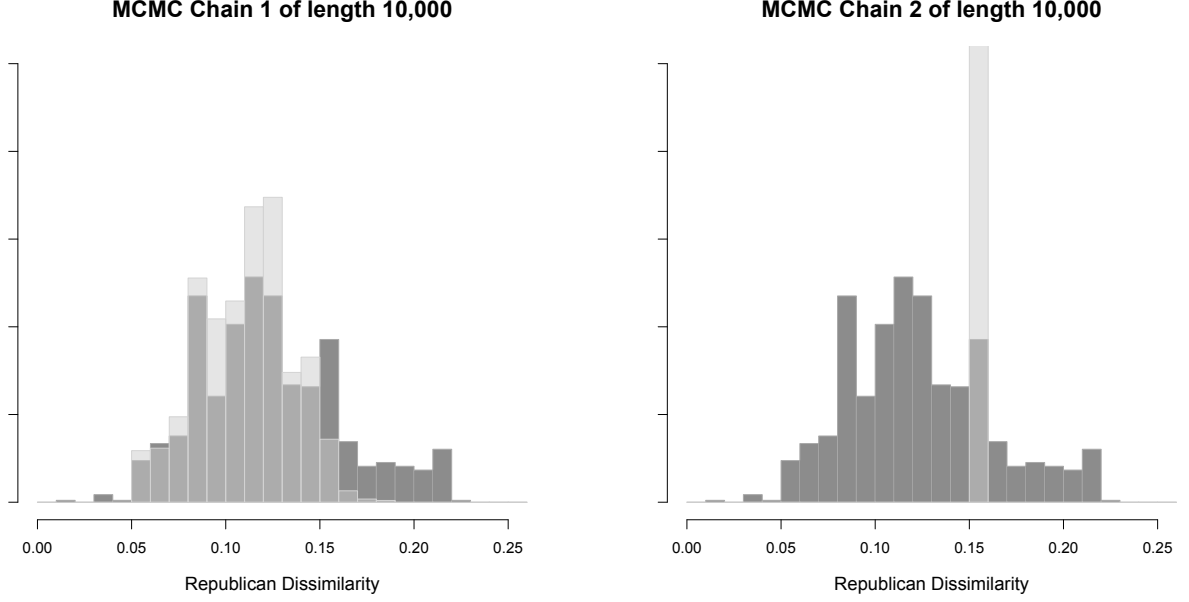


Figure 1: The dark gray histograms show the underlying population. The light gray histograms display the sampling behavior of the Markov chains.

their theorem is valid even for a disconnected state space.

Figure 1 shows two Markov chains where movement is defined by the shift of one geographic unit to an adjacent district. The underlying data come from 25 precincts from the state of Florida.⁴ This data set is well-suited for our purposes because it is both large enough to be non-trivial and small enough that it can be examined in depth. The number of ways to partition 25 precincts into 3 districts without constraints is a Stirling number of the second kind, $S(25, 3) = 141,197,991,025$. At the same time, the number of possibilities is small enough that we can enumerate the entire set of feasible maps. Since we have the underlying population distribution, we know the correct answer for these data. If we impose a contiguity constraint, the number of valid partitions reduces by several orders of magnitude to 117,688. If we further impose a population constraint that requires the population deviation from the ideal population to be less than 10%, the number of valid partitions drops to 927.⁵

⁴This data set is available in the *redist* CRAN package (Fifield et al., 2017).

⁵This data set is much smaller than an actual redistricting instance, but it exhibits the phenomenon we wish to highlight that also occurs for the much larger actual redistricting problems. Using small data sets with a known answer is a common way to test algorithms intended for larger data instances where the answer is unknown since the small data sets allow us to identify where problems might occur.

The behavior of two different Markov chains is shown in Figure 1. Both chains are of length 10,000. They were each started from a different (randomly generated) feasible map. The dark gray bars show metrics for the underlying population of feasible maps.⁶ The light gray bars show metrics from the states visited by the Markov chain. The chain on the left traverses more of the space than the chain on the right. It visits 486 feasible maps. The chain on the right moved between only two of the 927 feasible maps because these two maps are disconnected in the state space from all of the other maps. That is, no other feasible map, outside of these two maps, can be reached within the population and contiguity constraints with a single movement of one geographic unit to an adjacent district.

While the chain on the left visited a little more than half of the feasible maps, we can see that there is a portion of the state space that was never visited (the maps with larger values on the partisan metric). This portion of the space is disconnected from the visited space and cannot be reached no matter how long the chain is run. In this instance, if we were to begin the chain at the map in this disconnected subspace with a partisan metric of 0.16 and use the CFP test to explore whether this map is an extreme statistical outlier, the result would be highly significant because virtually all other maps that can be reached by this chain would have a smaller value on the partisan metric. At the same time, while a map with a metric of 0.16 may be a CFP ε -outlier, it is not an outlier in the space of all feasible maps. It has an outlying value in the space that was traversed by the chain, but we do not know how the space that was traversed by the chain is distributed in the entire state space. The 0.16 map would be (correctly) identified as an ε -outlier, but it is not a global outlier.

The plot on the right side of Figure 1 further exemplifies that a Markov chain can traverse only a very small portion of the state space. Though this chain also ran for 10,000 steps, it explored only a very small portion of the overall state space, only 2 maps in total.

⁶The metric is “Republican dissimilarity.” This measure was provided in the data set. We do not make any claims about this measure or about its relationship to the concept of partisan fairness. Note that the measure is unimportant for the point we are making. If we are able to sample from the space of all feasible maps, we can recover the distribution of *any* map metric since we have recovered a representative sample of maps. The Court has made no pronouncements about the proper partisan metric. We make no claims here either. We make no substantive claims whatsoever about whether partisan gerrymandering has occurred. We are not exploring what might be a proper metric. Those arguments are completely orthogonal to the exercise here.

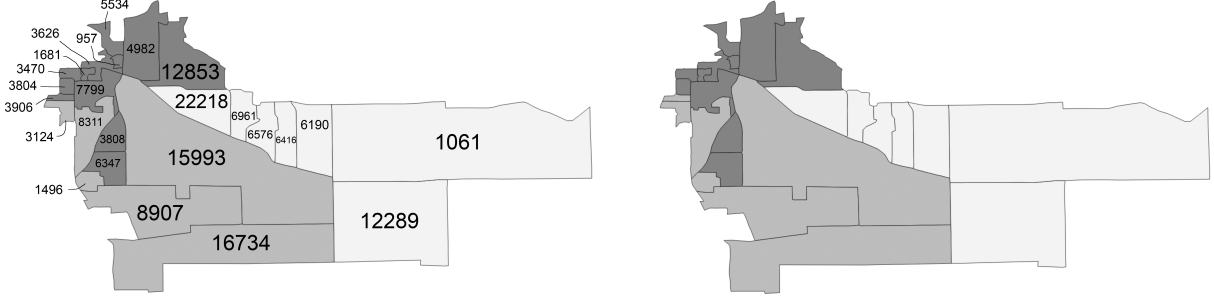


Figure 2: Two feasible maps, connected to one another in the state space, but isolated from the rest of the feasible state space. The map on the left shows the population counts for each precinct.

Figure 2 shows the two maps it visited and allows us to verify that if either of these maps is the starting point, it is only possible to reach the other map with one unit moves. All other possible maps are unreachable because either the population or the contiguity constraint is violated by such a move. The CFP test can assess whether a unit is an ε -outlier only within the space that its Markov chain traverses. Moreover, the extent of the traversed space varies, depending on the starting point of the chain, and it may be quite small. Further, and importantly, as this example illustrates, all of the traversed space may lie in the middle portion of the underlying distribution.

In this case, CFP would (again correctly) identify neither of these two maps as an ε -outlier. We are not arguing that the CFP theorem produces an improper p -value for a *particular* phenomenon. We point out only that there is little relationship between an ε -outlier and a global outlier. If two isolated maps had a metric near 0.23, which *is* a global outlier, CFP would still (correctly) identify them as *not* ε -outliers. A map can be an ε -outlier and not a global outlier. Likewise, a map can be a global outlier without being an ε -outlier. Neither one implies the other.

If the underlying distribution is unknown, it is also unknown where in a distribution a Markov chain may be traversing. Identifying that a state is an ε -outlier in the CFP sense, meaning that it is an outlier in some connected subspace traversed by some Markov chain, does not provide any information about whether that state is an outlier in the global space of all possible redistricting maps. Because we have no idea what part of the underlying distribution of all maps is being explored by the CFP chain, the CFP chain can compute

a p -value in a valid way within the space it explores, but the entirety of the explored space can still be comprised of maps that all have typical partisan metrics in the global space.

To be sure, in our first example in Figure 1, it would be unusual to begin with the map at the right tail of the light gray histogram. But, has the Supreme Court asked a question that can be translated as “is it unusual to have started a Markov chain with a map that is on the tail of the light gray histogram?” The Court asks whether a map is a partisan gerrymander. Whether this could be the same question is a legal issue that must be determined by the Court.

2.2 Detecting Gerrymandering with the CFP Test

CFP apply their theorem to detect whether the current Pennsylvania congressional map is an outlier among “all possible electoral maps” (Pegden, November 27, 2017). They claim that this is a way to *rigorously detect gerrymandering*. In Pegden’s report for the court, he states that,

when I report that Pennsylvania’s 2011 Congressional districting is gerrymandered, I mean not only that there is a partisan advantage for Republicans and that districtings with less partisan bias were available to mapmakers, but indeed that among the entire set of available districtings of Pennsylvania, the districting chosen by the mapmakers was an extreme outlier with respect to partisan bias, in a statistically rigorous way (Pegden, November 27, 2017, p. 2).

He does not qualify this statement to say that his test can identify a map as an extreme outlier in partisan bias even when the identified partisan bias value is a perfectly typical partisan bias value for the set of all maps. He states that his p -value identifies the phenomenon of partisan gerrymandering, a legal phenomenon that is defined by the Supreme Court and is not subject to redefinition outside the Court.

Moreover, regardless of what is proposed or by whom, and whether the test is mathematically or statistically rigorous or not, whether the proposal is adopted by a court of law is a legal question. It is the court who makes the determination of whether they wish to adopt any particular test as a judicially manageable standard for adjudicating partisan redistricting claims. Their determination respects mathematical rigor, but is necessarily and heavily guided by existing case law and precedent. To make this determination, the

court does not need to understand all of the mathematical details, but it *must* understand precisely what is being measured.

To be clear, we take no issues with the mathematics behind the CFP theorem or its proof. We make, however, an important observation that mathematics and the law are not harmonious or helpful to the other unless there is common understanding and effective communication. One way in which communication may not be effective is when both sides use the same words, but attribute different meaning to those words. Here, the word gerrymandering is used by both mathematicians as well as the legal community. The legal community understands gerrymandering only as it is defined by the Supreme Court. The Supreme Court has stated that partisan gerrymandering is unconstitutional. The question is not constitutionality. The question is whether there is a judicially manageable standard for measuring when partisan unfairness has reached “an extremity of unfairness.”

To understand if CFP is a rule or standard that the court wishes to adopt, we must first have a very clear understanding of the rule. At first blush, it appears that CFP are claiming that they have devised an outlier test that is similar to what could be achieved with a mixed MCMC chain. The only seeming difference is that since their Markov chain is not required to mix, they do not assess whether an observation is an outlier from observing its location in the sample distribution. Instead, they provide a test where a p -value can be assigned to an outlier hypothesis without a mixed chain and without a sample of the underlying distribution. Though both a mixed MCMC chain and CFP present ways to define an “outlier test,” these two outlier tests are in fact distinct and quite different from one another.

An outlier test from a mixed MCMC chain allows us to make statements about whether a particular unit is a global outlier in the space of all feasible redistricting maps. That is, given the distribution of all possible partisan metrics, a particular value lies in the tail of the distribution. However, an ε -outlier, as CFP have defined it, can lie anywhere in the distribution. An ε -outlier is not a *global* outlier, as we traditionally understand global outliers.

It appears that CFP wish to use their theorem to attribute a *different* “global phenomenon” to their ε -outliers. In particular, their theorem is based on the idea that, in the

global space, there cannot be a large proportion of local outliers. “The $\sqrt{\varepsilon}$ test is based on the fact that no reversible Markov chain can have too many local outliers (CFP, p. 2861).” Note that this is an important, legally distinct, and non-traditional definition of a “global outlier.” While one might define an outlier in this fashion, CFP’s definition choice is subtle and not transparent. This is a crucial point because whether or how the CFP definitions might serve as the basis for a legal standard is a separate question from the correctness of the mathematical foundations of the CFP theorem.

Determining the appropriateness of the CFP test in the legal realm requires both a nuanced understanding of the properties of an ε -outlier as well as a nuanced understanding of the law. CFP wish to argue that maps that are ε -outliers are “carefully crafted,” meaning that the map drawer could have chosen a large number of other maps that are very similar to the current map, but chose not to for partisan reasons. The Court has never proposed this “carefully crafted” phenomenon as partisan gerrymandering and has not evaluated whether it would accept this argument if only some unknown portion of the set of all maps has been examined. It *may* accept this argument, but 1) this test does not emanate from case law, and 2) for the Court to properly evaluate it, legal scholars must appropriately understand the CFP test.

What the CFP test precisely identifies is not clear unless one has a reasonably deep understanding of mathematics. Realizing whether the CFP test is consistent with the Supreme Court’s legal framework requires a reasonably deep understanding of the law. Though both disciplines are deeply enshrouded in logic, they use different languages, and so special care must be exercised to avoid miscommunication.

References

- Bangia, S., C. Vaughn Graves, G. Herschlag, H. S. Kang, J. Luo, J. C. Mattingly and R. Ravier. 2017. “Redistricting: Drawing the Line.” *ArXiv e-prints*.
- Chikina, Maria, Alan Frieze and Wesley Pegden. 2017. “Assessing significance in a Markov chain without mixing.” *Proceedings of the National Academy of Sciences* 114(11):2860–2864.

- Fifield, Benjamin, Michael Higgins, Kosuke Imai and Alexander Tarr. 2017. “A New Automated Redistricting Simulator Using Markov Chain Monte Carlo.” Working Paper.
- Geman, S. and D. Geman. 1984. “Stochastic Relation, Gibbs Distributions, and the Bayesian Restoration of Images.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:721–741.
- Grofman, Bernard and Gary King. 2007. “The Future of Partisan Symmetry as a Judicial Test for The Future of Partisan Symmetry as a Judicial Test for Partisan Gerrymandering after LULAC v. Perry.” *Election Law Journal* 6(1):2–35.
- Hastings, W. K. 1970. “Monte Carlo sampling methods using Markov chains and their applications.” *Biometrika* 57(1):97–109.
- Herschlag, G., R. Ravier and J. C. Mattingly. 2017. “Evaluating Partisan Gerrymandering in Wisconsin.” *ArXiv e-prints*.
- Mattingly, J. C. and C. Vaughn. 2014. “Redistricting and the Will of the People.” *ArXiv e-prints*.
- Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller and Edward Teller. 1953. “Equation of State Calculations by Fast Computing Machines.” *The Journal of Chemical Physics* 21(6):1087–1092.
- Pegden, Wesley. November 27, 2017. “Pennsylvania’s Congressional districting is an outlier: Expert Report.” Expert Report filed in *League of Women Voters of Pennsylvania et al., v. the Commonwealth of Pennsylvania et al.*, in the Commonwealth Court of Pennsylvania (Civ. No. 261 MD 2017).
- Propp, James Gary and David Bruce Wilson. 1996. “Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics.” *Random Structures and Algorithms* 9(1–2):223–252.
- Stephanopoulos, Nicholas O. and Eric M. McGhee. 2015. “Partisan Gerrymandering and the Efficiency Gap.” *University of Chicago Law Review* 82.