A Probabilistic Model for Detecting Gerrymandering in Partially-Contested Multiparty Elections

Dariusz Stolicki[1,2(✉)], Wojciech Słomczyński[1,3], and Jarosław Flis[1,4]

[1] Jagiellonian Center for Quantitative Research in Political Science,
Jagiellonian University, ul. Wenecja 2, 31-117 Kraków, Poland
{dariusz.stolicki,jaroslaw.flis}@uj.edu.pl,
wojciech.slomczynski@im.uj.edu.pl
[2] Faculty of International and Political Studies,
Jagiellonian University, Kraków, Poland
[3] Faculty of Mathematics and Computer Science,
Jagiellonian University, Kraków, Poland
[4] Faculty of Management and Social Communication,
Jagiellonian University, Kraków, Poland

Abstract. Classic methods for detecting gerrymandering fail in multiparty partially-contested elections, such as the Polish local election of 2014. A new method for detecting electoral bias, based on the assumption that voting is a stochastic process described by Pólya's urn model, is devised to overcome these difficulties. Since the partially-contested character of the election makes it difficult to estimate parameters of the urn model, an ad-hoc procedure for estimating those parameters in a manner untainted by potential gerrymandering is proposed.

Keywords: Gerrymandering · Partially-contested elections · Pólya's urn model · Dirichlet distribution · Seats-votes relationship

1 Introduction

Most political representative bodies in the world are chosen through multi-district elections, where seats are apportioned among n parties within each of c districts independently, i.e., solely on the basis of the district vote. In such elections, jurisdiction-wide distribution of seats (*the seat distribution*) depends heavily not only on the overall voting result (i.e., a vector of party vote shares), but on the geographical distribution of each party's support over the set of electoral districts (*the vote distribution*). Anomalous vote distributions can lead to skewed electoral results, such as the well-known referendum paradox [11,49,59]. While

such anomalous distributions can arise through natural causes, such as voter self-segregation and other population clustering effects [18,38,42,43,75] (the U.S. electoral college, where two out of five most recent elections involved instances of the referendum paradox favoring the Republican Party, affords a prominent example), they can also be facilitated through deliberate manipulation of electoral district boundaries. Such manipulation, especially when undertaken for the purpose of obtaining an advantage for the party or block of parties controlling the redistricting process, is known as *gerrymandering*.

Gerrymandering is possible under all kinds of voting rules [6], but is most common under the combination of single-member electoral districts and the plurality rule (known in political science as the FPTP system). A classic gerrymander under FPTP is based on a combination of two strategies: assigning as many opposition voters as possible into a small number of districts (*packing*) (obviously, that number needs to be smaller than $c/2$), while spreading out the remainder roughly equally across other districts in such manner that they do not constitute a majority in any of them (*cracking*) [5,27]. When done correctly, this results in a substantial number of opposition votes in the "packed" districts being wasted, while the opposition supporters in other districts are so diluted that they are incapable of securing a plurality in any of them. If there are more than two parties, other strategies also become possible, such as *stacking*, balancing the number of supporters of different opposition parties in such manner that enables the preferred candidate to win with less than majority, but they tend to require more detailed knowledge about voter preferences and their distribution.

Ultimately, however, even as both strategies and objectives of gerrymandering are well-understood, the concept itself, as we will see below, remains difficult to formalize. Even apart from difficulties necessarily involved in discerning intent (and hence distinguishing manipulation from unintentional bias), there is no accepted standard by which a specific vote distribution can be judged "fair" or "natural" [16,36,39]. Without such standard, the concepts of distributional "unfairness" or "anomalousness" are fuzzy at best and meaningless at worst. This obviously makes it more difficult to detect and identify gerrymandering, as resort has to be had to circumstantial or otherwise indirect evidence.

2   Methodological Approaches to Detecting Gerrymandering

Altman et al. [2] distinguish six basic methodological approaches to detecting gerrymandering: *method of stated intent*, which relies on public statements of the authors of the districting plan; *method of totality of the circumstances*, which focuses on the political circumstances (well-known geographical rivalries, past practice, etc.); *method of evaluation of process*, which analyzes the districting process; *methods of inspection*, where gerrymandering is inferred from some qualitative or quantitative characteristics of the districting plan; *method of post-hoc comparisons*, where the districting plan is compared against a random sample of alternative plans; and *method of revealed preferences*, where the districting plan

is compared against alternatives rejected during the districting process. Of those, the first three are purely qualitative and only rarely will suffice to prove gerrymandering, or even systemic bias. In addition, they require extensive extrinsic knowledge about the districting process that cannot be obtained from election results and districting plans alone. The method of revealed preferences, while advocated by [2], also requires such extrinsic knowledge (namely the set of plans that were known to the districting authority but have been rejected).

That leaves us with only two classes of quantitative methods for detecting gerrymandering that can be applied when extrinsic knowledge is unavailable: the methods of inspection and the method of post-hoc comparisons. As noted, in the former we focus on some observable characteristics of the districting plan and compare them against a well-known standard. Most of such methods focus on one of the basic types of plan characteristics: district geometry and the relation between seats and votes. Geometric methods involve tests of district contiguity and of various measures of district compactness [1, 28, 62, 77], trying to formalize the intuition that gerrymandered electoral districts are oddly shaped. Yet the evidence of manipulation provided by such methods is circumstantial at best, as irregularity of shape is neither necessary nor sufficient condition for gerrymandering.

Methods focusing on the seats-votes relation instead start from some assumptions about the desired characteristics of such relation. Such characteristics may include proportionality [8], responsiveness to shifts in voter support (measured by the swing ratio, i.e., the derivative of seats with respect to votes) [56, 73], partisan symmetry (a requirement that seats-votes curves by identical for all competing parties [32, 33, 35, 37, 47, 57]), or the efficiency principle, requiring that the number of wasted votes be equal for all parties [53, 69]. Then each party's seats-votes function (i.e., a function assigning to total vote share v the total seat share s) is tested for deviation from the chosen characteristics. Those methods generally share three principal limitations. The first one is of fundamental nature: most of the methods described above (except the partisan symmetry method) involve *a priori* assumptions that certain form of the seats-votes function is a natural one, but no attempt is made to justify those assumptions, for instance by showing that they arise from some general or accepted distributional assumptions. Without such justification it may well be that those methods generate a large number of false negatives by holding districting plans to a more restrictive standard than mere absence of distributional anomalies. The second problem with methods focusing on the seats-votes relation is more technical: they usually require that the full seats-votes function be known for each party, yet all that is empirically known is a single data point per election. Extrapolation from those data points involves questionable assumptions about how changes in one party's vote share translate into changes in its vote distribution and in other parties' vote shares (like the uniform partisan swing assumption, see [14, 15, 29, 35, 58]). Finally, virtually all methods focusing on the seats-votes relation have been developed with two-party elections in mind and usually lack natural generalization for multiparty elections.

The method of post-hoc comparisons instead compares districting plans with an ensemble of alternative districting plans [18–20, 24, 51, 60]. The problem is

that the full set of correct solutions to the districting problem is in all but simplest cases too numerous to be used for such comparison, so we are reduced to testing the empirical plan against some sample of algorithmically generated random plans. But for proper inference to be drawn from such sample, we need the sample to be drawn from the set of all possible districting plans with some known probability measure, and we are unaware of any algorithm for generating districting plans for which such measure has been analytically determined [2,3].

Finally, all of the methods described above fail in *partially-contested elections*, i.e., those where only some parties (or even none) field candidates in all electoral districts, and other parties field candidates in fewer than all districts (including one-candidate parties that only run in a single district). In such cases, the number of candidates can vary across districts, affecting both vote distributions and seats-votes relationships. In addition, we are no longer free to generate alternative seat allocations by rearranging districts, since we no longer have data about each party's support beyond the districts it contested. It has been already noted by [45] that traditional statistical methods for dealing with missing data cannot be applied to partially-contested elections since failure to contest an election in a district is usually not a random event, but a function of the party's forecasted electoral strength in such district. Yet the methods for dealing with partially-contested elections proposed by, inter alia, [45,50,72,76], are also insufficient when the patterns of electoral contestation are very chaotic, and particularly if the election cannot be described as a mixture of relatively few patterns with multiple districts per each.

We have encountered exactly those problems when analyzing gerrymandering in Polish local election of 2014, which was held under the plurality rule. Due to highly personalized nature of local politics in Poland (especially in the smallest but most numerous class of municipalities, the townships), in 2386 out of 2412 municipalities the election was partially-contested. The chaotic character of electoral contestation patterns is best described by the following selection of facts:

– only 2218 out of 16,971 parties[1] have contested the election in all districts within their respective municipalities,
– if parties were ordered according to the fraction of districts contested within their respective municipalities, a median party would have contested less than half of all districts,
– 4733 parties have contested only a single district,
– there are, on average, 8.26 different district contestation patterns per municipality.

To address the problems described above, we propose a new method for detecting gerrymandering in partially-contested multiparty elections that are conducted under identical rules in multiple jurisdictions with separate districting plans (examples include regional and local elections, but also national elections

---

[1] Under the rules in place for the election, parties are registered at the district level, and every independent candidate is counted as a distinct party, thence the unusually large number of parties.

in which redistricting is done by subnational jurisdictions, as in the case of the U.S. House of Representatives). We proceed on two general assumptions: that voting in each district can be modeled by a stochastic process that is identical (modulo choice of parameters) for all jurisdictions of interest, and that gerrymandering is ultimately an exception rather than a rule, so the parameters of the stochastic model estimated from the set of all jurisdictions are free from the taint of manipulation. We first formulate a general model of vote distribution, then propose a procedure for estimating that model's parameters, and finally use that model to derive a sampling distribution of seat shares against which party seat shares can be compared.

3   Modeling District-Level Vote Distribution

3.1   Definitions and Notation

1. An electoral jurisdiction consists of a finite set of electoral districts D, whose cardinality we denote as $c := |D|$, a finite set of parties $P := \{1, \ldots, n\}$, and a left- and right-total relation $R \subseteq P \times D$ such that $(i, k) \in R$ if the i-th party fields a candidate in the k-th district. It is assumed here that in each district there is exactly one seat to be allocated using the plurality rule and hence each party is able to field only a single candidate.
2. Let $D_i \subseteq D$ be the set of indices of the electoral districts where the i-th party, $i \in P$, fields candidates, i.e., a set of such $k \in D$ that $(i, k) \in R$. Let $c_i := |D_i|$.
3. Let $P_k \subseteq P$ be the set of indices of the parties contesting the k-th district, $k \in D$, i.e., a set of such $i = 1, \ldots, n$ that $(i, k) \in R$. Let $n_k := |P_k|$.
4. Let $\sim$ be an equivalence relation on D identifying districts contested by the same set of parties, i.e., such that $k \sim l$ if and only if $P_k = P_l$. By $[k]_\sim$ we denote an equivalence class of k in D with respect to $\sim$. We call it a *contestation pattern*.
5. The *voting result* in the k-th district is a vector $\mathbf{v_k} := \left(v_{i_1}^k, \ldots, v_{i_{n_k}}^k \right) \in \delta_k$, where δ_k is an n_k-face of the standard $(n-1)$-dimensional simplex Δ_n that includes vertices i_1 to i_{n_k}, v_i^k is the *i-th party's vote share* in the k-th district, and $i_1 < \cdots < i_{n_k}$ are elements of P_k. Note that δ_k can be identified with the standard $(n_k - 1)$-dimensional simplex Δ_{n_k}.
6. Let $v_i := \left(\sum_{k \in D_i} v_i^k w_k \right) / \left(\sum_{k \in D_i} w_k \right)$, where w_k is the number of voters in the k-th district, be the i-th *party's total vote share*.
7. Let $\boldsymbol{D_m}$ be the set of all such districts $k \in \bigcup D$, where the sum is over all electoral jurisdictions of interest, that $n_k = m$.
8. By *quantile mixture* of absolutely continuous probability distributions $\mathcal{M}_1, \ldots, \mathcal{M}_m$ supported on some compact I we understand a probability distribution characterized uniquely by the inverse cumulative distribution function $\Lambda^{-1} : [0, 1] \to I$ given by $\Lambda^{-1}(x) := \frac{1}{m} \sum_{i=1}^m F_i^{-1}(x)$, where F_i^{-1} is the inverse cumulative distribution function of $\mathcal{M}_i$ [44].
9. Where single-district models are discussed (in Sects. 3.2 and 3.3) index k is omitted.

3.2   Overview of Available Models

The problem of modeling voter choice in single-choice electoral systems can be though of as a special case of the problem of modeling preference orderings, which is well known in the social choice theory (see, e.g., [65] and [71]). A number of models has been employed for that purpose, but, since we are only interested in the first choice, we omit the discussion of those that differ only in their treatment of the second and subsequent preferences.

1. Under the *Impartial Culture (IC)* model, each preference ordering (and, therefore, also each choice of the first candidate) is equiprobable and each voter decides independently with fixed probabilities [17]. The voting result **v** follows a multinomial distribution centered at the barycenter of Δ_n with the variance of the square distance from the barycenter $O\left(w^{-1}\right)$. There is extensive evidence for the claim that both the equiprobability and independence assumptions are not satisfied in empirical elections (recounted by, inter alia, [65]).

2. The *multinomial model* is a generalization of the IC model which assigns unequal probabilities to the candidates, but still assumes that each voter makes an independent choice with fixed probabilities described by vector **p**. The voting result **v** follows a multinomial distribution centered at **p** with the variance of the square distance from **p** behaving as $O\left(w^{-1}\right)$. As first noted in [46], this model significantly underestimates the variance of the vote distribution. To avoid that problem, Penrose and others [61,67,74] have proposed *clustered multinomial model*, according to which each district's population consists of κ equally sized clusters of voters who have identical characteristics and instead of randomizing individual voters' choices, we randomize each cluster's choice. Under that model, **v** still follows a multinomial distribution centered at **p**, but its variance increases to $O\left(\kappa^{-1}\right)$ (as $\kappa \ll w$ – Penrose's original estimate for Great Britain was $\kappa \approx 14$).

3. The *Impartial Anonymous Culture (IAC)* model treats each preference profile (and, therefore, each voting result) as equiprobable [31,48]. Accordingly, the voting result **v** follows the uniform distribution on a discrete grid of points within Δ_{n_k}, which, as w approaches ∞, weakly converges to the uniform distribution on Δ_{n_k}.

4. The *Pólya urn model*, first introduced by Eggenberger and P ólya in 1923 [22], has been applied in the field social choice theory by, inter alia, [12, 21,40,66]. Voting is treated as a discrete stochastic process where a ball is drawn from an urn that initially contains α_i balls of the i-th color (where $i = 1, \ldots, n$), and after each draw λ balls of the same color as the one drawn are returned to the urn. The voting result **v** follows the multivariate Pó lya distribution and, as w approaches ∞, converges almost surely to a random variable having the Dirichlet distribution parametrized by vector $(\alpha_1, \ldots, \alpha_n)/\lambda$ [7,41]. Both IC and IAC are special cases of the urn model, with $\alpha_1 = \cdots = \alpha_n = 1$ and $\lambda = 1$ for IAC and 0 for IC.

5. *Spatial models* assume that voter policy preferences are distributed (usually normally) over a multidimensional policy space, that party policy positions are either specified or randomly distributed over the same space and that voters always choose the candidate of the closest party according to some fixed metric [4, 23, 55]. Again, some voting clustering has to be assumed to avoid overestimating homogeneity.

There is considerable evidence that equiprobability and independence assumptions fail to match empirical data, and accordingly both IC and IAC fail as empirical models of electoral behavior [65, 71]. In [71], spatial models are found to be most effective in modeling preference profiles, but in single-choice elections such models involve too many degrees of freedom for estimation unless highly simplifying assumptions are made (for instance, about reduction of the number of dimensions). That leaves only the urn model for our intended applications.

Sociological theory of electoral behavior also provides sound reasons for adopting the urn model. Contagion mechanisms it is used to model translate into an observation that most voters are initially undecided and their political views are shaped through social interactions with others, who include already-committed supporters of the parties and candidates (cf. [13]). Indeed, political parties recognize that direct mobilization of voters through personal interaction is one of the most important tools of electoral campaigning [26]. Even mass media influence on political views, which would seem to support rather fixed-probability models, is indirect and effective primarily when the information communicated by the media is later verified through direct interaction with other members of the community [54]. It is also recognized that such political contagion processes are essentially stochastic, being dependent on the fine structure of social networks [52] which cannot be predicted deterministically.

3.3   Urn Model of Electoral Behavior

A Pólya urn model is usually characterized by two parameters: a vector of initial ball numbers ($\alpha \in \mathbb{R}_+^n$) and the number of additional balls returned after each draw ($\lambda \in \mathbb{R}_+ \cup \{0\}$), but note that by rescaling vector α we can always obtain $\lambda = 1$, thereby reducing our parameter space to $\mathbb{R}_+^n$. In addition, it is often convenient to express α as a product of an n-element vector $\mathbf{p} \in \Delta_n$ and of the concentration parameter $\alpha \in \mathbb{R}_+$.

Definition 1. *Pólya-Eggenberger Urn Model [22, 63].*

Let us consider a countably infinite set of potential voters. Let $X_j \in P$ be the choice of the j-th voter ($j \in \mathbb{N}$, $\mathbb{N} = \{1, 2, 3, \dots\}$).

Voting is a discrete stochastic process where the probability of the $(j + 1)$-th voter choosing the i-th party's candidate is defined by induction as

$$\Pr\left(X_{j+1} = i\right) = \frac{\alpha p_i + |\{k = 1, \dots, j : X_k = i\}|}{\alpha + j}, \tag{1}$$

for $i = 1, \dots, n$.

Intuitively, the attractiveness of the i-th party to the $(j+1)$-th voter is proportional to the sum of the number of voters that already have decided to support it and its initial strength αp_i. In [66] the authors propose that αp_i be interpreted as the number of voters who are committed at the outset to support the i-th party's candidate, but this interpretation raises some issues as αp_i need not be an integer.

Proposition 1. *In the above situation, there exists a random variable* $\mathbf{V} \sim \mathrm{Dir}\,(n; \alpha p_1, \ldots, \alpha p_n)$ *such that* $(\Pr\,(X_j = 1), \ldots, \Pr\,(X_j = n)) \overset{a.s.}{\to} \mathbf{V}$ *as* $j \to \infty$, *where* $\mathrm{Dir}\,(n; \alpha p_1, \ldots, \alpha p_n)$ *(the* Dirichlet distribution*) is a continuous multivariate probability distribution supported on* Δ_n *that has a probability density* f *with respect to the Lebesgue measure on* Δ_n *given by:*

$$f\,(v_1, \ldots, v_n) := \frac{1}{\mathrm{B}\,(\alpha_1, \ldots, \alpha_n)} \prod_{i=1}^{n} v_i^{\alpha_i - 1}, \tag{2}$$

where $\mathbf{v} \in \Delta_n$ *and* $\mathrm{B}\,(\alpha_1, \ldots, \alpha_n)$ *is the multivariate beta function:*

$$\mathrm{B}\,(\alpha_1, \ldots, \alpha_n) := \frac{\prod_{i=1}^{n} \Gamma\,(\alpha_i)}{\Gamma\,(\alpha)}. \tag{3}$$

In the above situation we have for $i = 1, \ldots, n$:

$$V_i \sim \mathrm{Beta}\,(\alpha p_i, \alpha\,(1 - p_i)); \tag{4}$$

$$\mathrm{E}\,(V_i) = p_i; \tag{5}$$

$$\mathrm{Var}\,(V_i) = \frac{p_i\,(1 - p_i)}{\alpha + 1}. \tag{6}$$

For proof of the above proposition see, inter alia, [7] and [41] (Fig. 1).

3.4   Parameter Fitting – The Expectation Vector

Literature on electoral studies recognizes that district-level vote shares depend on two principal factors: overall party popularity, measured by the total vote share vector $\mathbf{v}$, and political geography, i.e., district-specific effects, which are more difficult to model formally. However, as we consider an idealized distribution of vote shares in a non-biased election, in essence approximating an average distribution of district vote shares over the population of non-biased districting plans, we abstract from the effects of political geography altogether.

It would thus appear from (5) that the vector of party total vote shares $\mathbf{v}$ would be the most natural estimate of parameter $\mathbf{p}$. This, however, is only the case if the voting results in all districts in D come from a single distribution, which in turn is equivalent to a condition that the election be fully contested, i.e., that every party j field a candidate in every district k. Otherwise, there must be a different distribution for each equivalence class $[k]_{\sim}$, as each such class is characterized by the presence of a different set of parties. It follows that in
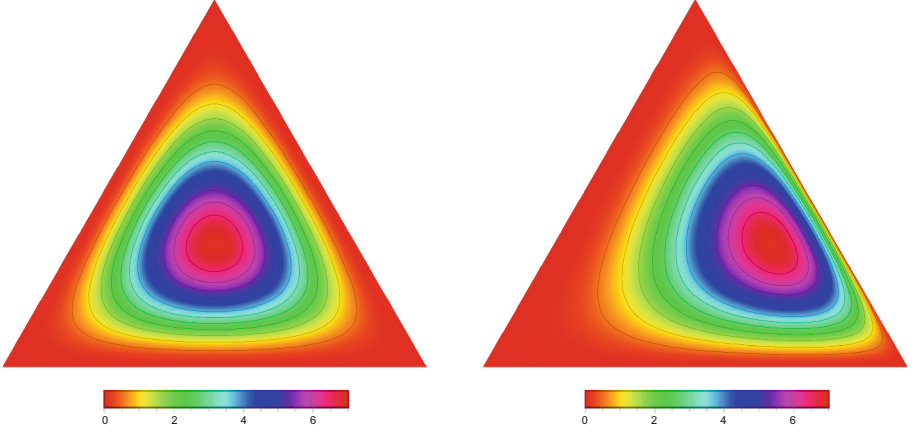
Fig. 1. Density plot of a symmetric Dirichlet distribution on Δ_3 with $\mathbf{p} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $\alpha = 9$ (left) and of an asymmetric Dirichlet distribution on Δ_3 with $\mathbf{p} = (\frac{4}{9}, \frac{1}{3}, \frac{2}{9})$ and $\alpha = 8$ (right).

partially-contested elections, which are of primary interest to us, voting results in D will be distributed according to a direct product of Dirichlet distributions $\mathcal{D}_k := \mathrm{Dir}\,(n_k; \alpha_k \mathbf{p}_k)$, with n_k, $\mathbf{p}_k$ and α_k constant for each equivalence class $[k]_\sim$, and $\mathbf{p}_k$ and α_k being unknown.

We cannot simply assume that $p_i^k = \sum_{j \in [k]_\sim} v_i^j / |[k]_\sim|$ for each $k \in D$, since the empirical vote distribution over equivalence classes $[k]_\sim$ may be tainted by gerrymandering. Instead, we need a theoretical model that is based solely on district contestation patterns (described by relation R) and party total vote shares vector $\mathbf{v}$.

In fitting $\mathbf{p}_k$ to each equivalence class in D with respect to $\sim$, we seek to satisfy the following three natural requirements:

R1 For each district k, $\mathbf{p}_k \in \Delta_{n_k}$, i.e., $\sum_{i \in P_k} p_i^k = 1$.
R2 For any two districts $k, l \in D$, if $k \sim l$, then $\mathbf{p}_k = \mathbf{p}_l$.
R3 For any two parties $i, j \in P$, the order on $\{p_i^k, p_j^k\}$ is identical in every district $k \in D_i \cap D_j$.
R4 For any two parties $i, j \in P$ such that $D_i = D_j$, the order on $\{p_i^k, p_j^k\}$ is identical with the order on $\{v_i, v_j\}$ for each district $k \in D_i$.

In addition, there are three postulates that we seek to satisfy approximately (i.e., to minimize deviation from them):

P5 For each party $i \in P$ its mean expected vote share over districts should be close to its party vote share, i.e., $\sum_{k \in D_i} p_i^k \approx c_i v_i$.
P6 For any two districts $k, l \in D$ if $n_k = n_l$, then $p_i^k \approx p_i^l$ for each party $i \in P_k \cap P_l$.

P7 For each party $j \in P$ and for any two districts $k, l \in D_j$ we have $\varphi_{n_k}\left(p_j^k\right) = \varphi_{n_l}\left(p_j^l\right)$, where $\varphi_m : [0,1] \to [0,1]$, $m \in \mathbb{N}$, is a function mapping a party vote share in a district with m contenders to a standardized value independent of m.

Of those postulates, **P7** clearly requires some additional discussion. The underlying problem consists of comparing vote shares across districts with different number of candidates. Clearly, obtaining 40% of the vote in a district with two candidates is not equivalent to obtaining an identical vote share in a district with ten candidates. In formal terms, this intuition can be expressed as follows: let X_m, $m \in \mathbb{N}$, be a random variable given by $X_m\left(i,k\right) := v_i^k$, where k is drawn from a uniform discrete distribution on $\boldsymbol{D}_m$ and i is later drawn from a uniform discrete distribution on P_k. The distribution of X_m necessarily depends on m, while for vote shares from different districts to be comparable, we need to transform X_m into another random variable with a distribution that is constant with respect to m.

The probability integral transform of X_m is one natural choice of such transformation. Let us consider the cumulative distribution function of X_m. As it is not injective, X_m being discrete, we formally define $\varphi_m : [0,1] \to [0,1]$ as its continuous approximation obtained by integrating the probit-transformed [2] kernel density estimator ψ_m [30] of the distribution of X_m, i.e., $\varphi_m\left(p\right) = \int_0^p \psi_m\left(x\right)\,dx$ for $p \in [0,1]$. This assures that φ_m is invertible, and that φ_m^{-1} is continuous, strictly increasing, and the images of the bounds of its domain are, respectively, 0 and 1. It follows that every linear combination of functions φ_k^{-1}, where $k \in \mathbb{N}$, with positive coefficients summing up to $c > 0$, is also continuous and strictly increasing, and the images of the bounds of its domain are 0 and c. Let $i \in P$. By the intermediate value theorem there exists a unique $q_i \in [0,1]$ such that $\sum_{k \in D_i} \varphi_{n_k}^{-1}\left(q_i\right) = c_i v_i \leq c_i$. Hence the definition $p_i^k := \varphi_{n_k}^{-1}\left(q_i\right)$ would naturally imply **P5**. Parameter q_i has no natural interpretation, however if we assume the distribution of the i-th party's district vote shares to be a quantile mixture of the distributions $\mathcal{D}_k$, where $k \in D_i$, q_i will correspond to the value of such mixture's cumulative distribution function Λ_i for the empirical value of v_i.

Note that model assuming $p_i^k = \varphi_{n_k}^{-1}\left(q_i\right)$ satisfies most of the requirements and postulates specified above:

- **P5** and **P7** are satisfied by definition of q_i and φ_m.
- If $n_k = n_l$, $p_i^k = \varphi_{n_k}^{-1}\left(q_i\right) = \varphi_{n_l}^{-1}\left(q_i\right) = p_i^l$ for any party i and any two districts $k, l \in D_i$, so **P6** is satisfied exactly and therefore implies **R2**.
- **R3** results from the monotonicity of φ_m^{-1}.
- From the monotonicity of φ_m^{-1} we know that the order on $\left\{p_i^k, p_j^k\right\}$ is identical with the order on $\left\{q_i, q_j\right\} = \left\{\Lambda_i\left(v_i\right), \Lambda_j\left(v_j\right)\right\}$. From $D_i = D_j$ it follows that $\Lambda_i = \Lambda_j$. As Λ_i is strictly increasing, the order on $\left\{q_i, q_j\right\}$ is identical to that on $\left\{v_i, v_j\right\}$, as desired under **R4**.

---

[2] We use the probit-transformed kernel density estimator instead of a standard Gaussian density estimation to ensure that the resulting estimator is of bounded support and that φ_m is surjective onto $[0,1]$.

Unfortunately, there is no guarantee that the above model satisfies **R1**. We therefore modify it by renormalizing vector $\mathbf{p}_k$ for each district k. This renormalization ensures that **R1** is satisfied, **R3** is preserved (as renormalization preserves the ordering of $p_1^k, \ldots, p_n^k$), and so are **R2** and **R4** (as the renormalization constant does not vary within $[k]_\sim$). In turn, such renormalization may introduce violations of **P5**, **P6**, and **P7**, but we do not need those postulates to be satisfied exactly.

Note that this method is loosely analogous to the biproportionality method by [9,10,64].

The distribution of p_i^k in all districts $k \in \boldsymbol{D}_m$, $m \in \mathbb{N}$, will be of further interest in the following section (recall that $\boldsymbol{D}_m$ is the set of all districts with m candidates). We denote its cumulative distribution function by Ψ_m.

3.5   Parameter Fitting – The Concentration Parameter

The last parameter of our electoral model is the concentration parameter α_k. Unlike the expected vote shares of the contending parties, α_k is never observable directly, and in most cases we do not have enough data to fit it to empirical voting results using some distribution fitting method that produces a reasonable confidence interval (since such fitting would require large equivalence class $[k]_\sim$). Intuitively, the concentration parameter should depend on at least two further parameters: the number of candidates and the political homogeneity of the jurisdiction under consideration. The latter, in turn, is likely to depend in a complex manner on a large number of factors, such as the population and area of the jurisdiction, historical cleavages, settlement structure, socioeconomic diversity, etc. We do not have a good theoretical model of those relationships that would enable us to estimate α_k and a formulation of such model would go far beyond the scope of this paper.

To circumvent this issue we treat the concentration parameter as another random variable distributed, for each class of districts $\boldsymbol{D}_m$, with a gamma distribution with parameters κ_m and θ_m. To apply our model to a particular class of elections, we still need to estimate those parameters of the distribution of the concentration parameter. We proceed as follows: let Y_m, $m \in \mathbb{N}$, be a random variable given by $Y_m(i,k) := V_i^k$, where V_i^k is the i-th barycenteric coordinate of $\mathbf{V}_k \sim \mathrm{Dir}(n_k; \alpha_k \mathbf{p}_k)$, k is drawn from a uniform discrete distribution on $\boldsymbol{D}_m$ and i is later drawn from a uniform discrete distribution on P_k. Intuitively, it is the theoretical vote share of a random party in a random district in an ideal unbiased election. Under our model, the distribution of Y_m is a compound beta distribution with parameters $(\alpha p, \alpha - \alpha p)$ (see (4)), where $p \sim \Psi_m$ and $\alpha \sim \mathrm{Gamma}(\kappa_m, \theta_m)$. Accordingly, the density of that distribution is given by:

$$f_m(x) = \int_0^1 \int_0^\infty \frac{x^{\alpha p - 1}(1-x)^{\alpha - \alpha p - 1}}{\mathrm{B}(\alpha p, \alpha - \alpha p)} \frac{\alpha^{\kappa_m - 1} e^{-\frac{\alpha}{\theta_m}}}{\Gamma(\kappa_m)\theta_m^{\kappa_m}} \, d\alpha \, d\Psi_m(p). \qquad (7)$$

The function Φ_m is known at this stage (having been estimated in the preceding section), so the only two unknowns in this formula are the gamma distribution

parameters κ_m and θ_m. But note that under our general assumption that gerrymandering is not ubiquitous, the distribution of Y_m should closely approximate the distribution of X_m provided the model is correct. Therefore, we can use that property to obtain κ_m and θ_m. We do that by numerically minimizing, for each $m \in \mathbb{N}$, the total variation distance [34,68] between the distributions of X_m and Y_m.

4   Modeling the Sampling Distribution of Seats

By this point, we have estimated all the parameters necessary to model the ideal unbiased distribution of votes in each electoral district in every jurisdiction, namely, n_k and $\mathbf{p}_k$ for each district k, and κ_m and θ_m for each class of districts $\boldsymbol{D}_m$, $m \in \mathbb{N}$. Of course, not all anomalies in the vote distribution of a party are of interest to us when seeking to detect gerrymanders, but only those that translate into biases in the allocation of seats. To detect such biases, we run a Monte Carlo simulation for each jurisdiction of interest, generating a large number of simulated election results. We proceed as follows:

1. For each district $k \in D$, we generate a single realization of the random variable $\alpha_k \sim \mathrm{Gamma}\left(\kappa_{n_k}, \theta_{n_k}\right)$, which we will denote as $\widehat{\alpha}_k$.
2. For each district $k \in D$, we then generate a single realization of the random vector $\mathbf{V}_k \sim \mathrm{Dir}\left(n_k; \widehat{\alpha}_k \mathbf{p}_k\right)$, which we will denote as $\widehat{\mathbf{V}}_k$.
3. We distribute seats within each district $k \in D$ according to the plurality rule, awarding a single seat to the party with the greatest vote share, i.e., to the one corresponding to the greatest barycentric coordinate of $\widehat{\mathbf{V}}_k$.
4. For each party $i \in P$ we sum seats over districts.

This procedure is repeated 2^{20} times for each electoral jurisdiction. Through this process, we obtain a joint discrete sampling distribution of party seat vectors $\mathcal{S}$ on $\prod_{i=1}^n \{0, \ldots, c_i\}$, and for each party $i \in P$ we denote the marginal sampling distribution of seats by $\mathcal{S}_i$. In the process of estimating the above distributions we do not rely on the empirical distribution of voters among districts, and therefore they are untainted by the possible gerrymandering.

To measure the distance between actual seat count of the i-th party, s_i, and the distribution $\mathcal{S}_i$ obtained above, we introduce a simple measure analogous to the well-known p-value used in statistical hypothesis testing:

$$\pi_i := \min\left(\mathcal{S}_i\left([0, s_i]\right), \mathcal{S}_i\left([s_i, c_i]\right)\right). \tag{8}$$

In other words, π_i is the probability of a party obtaining the number of seats that is equal to or more extreme than its actual number of seats. Low values of π_i are indicative not only of anomalies in the vote distribution, but also of the fact that they translate into a rather improbable deviation from the expected number of seats.

To obtain a jurisdiction-level index, we could simply average the values of π_i over $i \in P$. However, to account for the fact that we are primarily interested in

cases of gerrymandering affecting parties contesting most districts, we weigh π_i by the number of districts c_i. The resulting index,

$$\pi := \frac{\sum_{i=1}^{n} \pi_i c_i}{\sum_{i=1}^{n} c_i}, \tag{9}$$

is our final measure of electoral bias. While not conclusive evidence of gerrymandering, since we still lack proof of intent, as electoral bias can be unintentional and arise due to pecularities of spatial distribution of party voters), it allows us to identify the outlier jurisdictions which can then be analyzed using other, possibly more qualitative methods.

Remark 1. Note that for our primary data set of interest, Polish local elections of 2014, π is quite well approximated by a normal distribution, see Fig. 2. This indicates an absence of pervasive gerrymandering, which agrees with the intuition that gerrymandering (or at least successful gerrymandering) is more difficult in less orderly party systems.
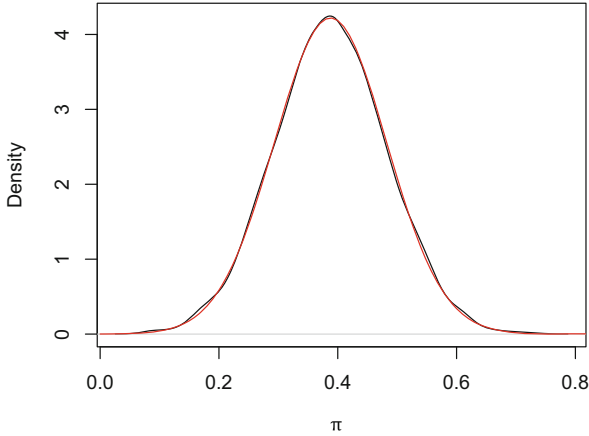

Fig. 2. A kernel density estimate of the empirical density of π for Polish local election of 2014 (black) and a normal density curve with $\mu \approx 0.388$ and $\sigma \approx 0.0945$ (red). (Color figure online)

To conclude, we have seen that classic methods for detecting gerrymandering fail when applied to multiparty partially-contested elections. We propose an alternative method based on a probabilistic model of voting behavior, together with a procedure for estimating the parameters of such model in a manner insulated from the possible taint of gerrymandering. We admit that the method is complex and involves simplifying assumptions and ad-hoc solutions, but they are made necessary due to the complexity of the problem and the limitations of the available data. Ultimately, we are unable to secure any conclusive evidence of gerrymandering, but we do obtain a single index that can be used to identify the suspect jurisdictions for further analysis.

A   Appendix

We have developed two proof-of-concept tests to evaluate the correctness of the proposed method for detecting gerrymandering. First, for any set of jurisdictions to which we seek to apply the proposed method we can test whether the empirical marginal distribution of X_m, $m = 1, \ldots, 11$, agrees with the theoretical marginal distribution expressed by (7). We have run such test for our primary data set of interest, Polish local election of 2014, producing, for the following values of parameters κ_m and θ_m (fitted in accordance with the procedure described in Sect. 3.5), the following total variation distances $d_{\mathbf{TV}}$ between the theoretical density function and the kernel density estimator of the empirical density (Fig. 3):

m	κ_m	θ_m	$d_{\mathbf{TV}}$	m	κ_m	θ_m	$d_{\mathbf{TV}}$
2	10.9656	1	4.97 E − 04	7	15.1239	1	2.51 E − 03
3	9.7069	1	1.42 E − 03	8	16.4486	1	2.73 E − 03
4	10.8787	1	2.02 E − 03	9	18.3641	1	2.93 E − 03
5	12.5535	1	2.22 E − 03	10	18.1292	1	3.24 E − 03
6	13.9117	1	2.27 E − 03	11	24.6423	1	4.33 E − 03

Even if the probabilistic model underlying the test is correct, it remains to be seen if the method is sensitive enough to detect actual gerrymandering (or other instances of electoral bias) and specific enough to keep the level of false positives low. Ideally, we would test the above using an empirical dataset that includes some known instances of gerrymandering, but our main dataset included none. Therefore, we have tested the sensitivity and specificity of the method against an artifical dataset, although one based on empirical data. We have algorithmically created a sample of 1024 districting plans for our home city of Kraków, each having 43 seats (as is the case in reality). Of those, 1020 were generated randomly using a Markov chain Monte Carlo districting algorithm developed by [25] and implemented in R package redist. The remaining four have been generated using two algorithms from [70], designed to produce districting plans gerrymandered in favor of one of the two largest parties (the first algorithm has also been designed to try to keep the districts relatively compact, while the second has been freed of all compactness constraints). Under each of those districting plans, we have calculated simulated election results using the 2014 precinct-level data. We treat each of such simulated elections as a single jurisdiction for which we carry out the procedure described in the article to obtain π. There are 12 distinct outcomes arising in simulated elections. As under all 1024 plans, all seats are won by the two largest parties, those outcomes are uniquely characterized by s_1 (or s_2).
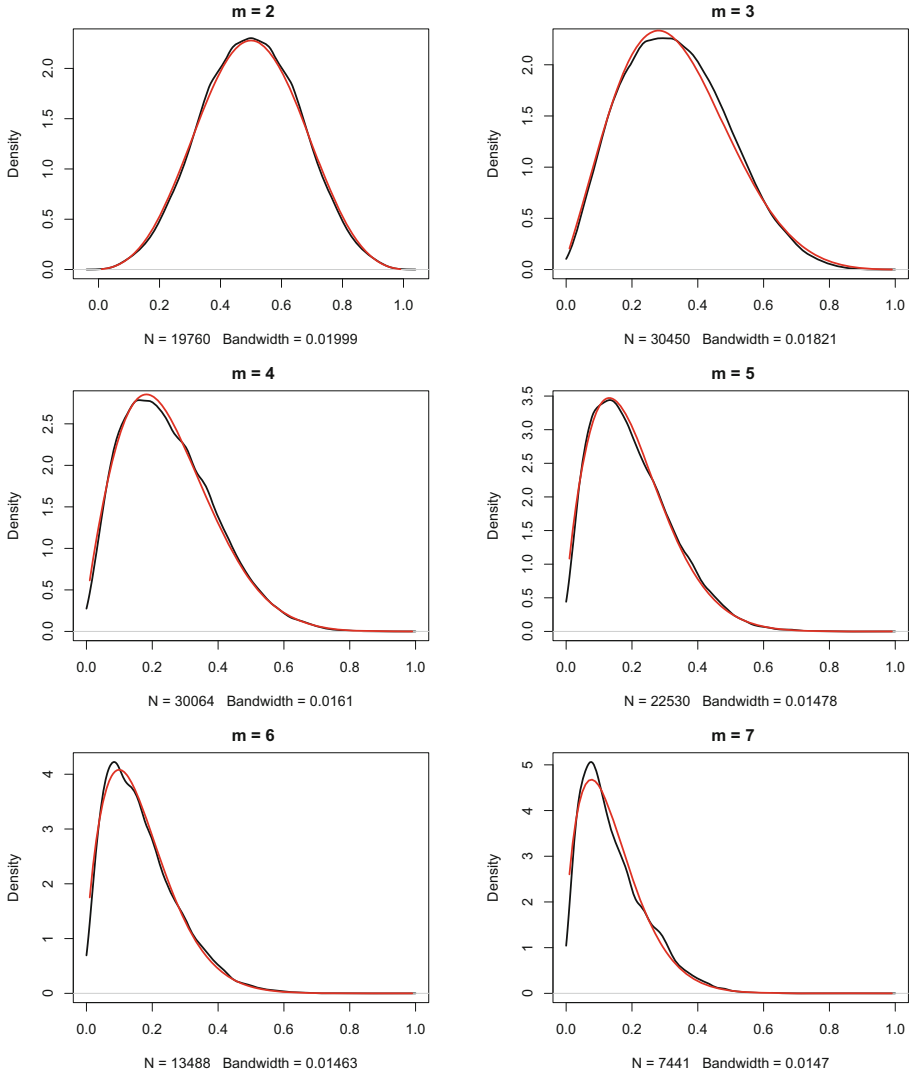
Fig. 3. The theoretical (red) and empirical (black) densities of v_i^k for different values of m. The empirical density is a kernel density estimate with the number of points and the bandwidth given below each plot. (Color figure online)

We list all of them, with corresponding π values, in the table below (the outcomes arising under the four intentionally gerrymandered plans are identified by the bold font):

s_1	s_2	No. of plans	% of plans	π
36	**7**	**1**	**0.10%**	**7.8E − 06**
31	**12**	**1**	**0.10%**	**0.39%**
26	17	1	0.10%	13.67%
25	18	3	0.29%	21.53%
24	19	43	4.20%	31.47%
23	20	226	22.07%	42.94%
22	21	433	42.29%	55.04%
21	22	262	25.59%	44.96%
20	23	48	4.69%	33.32%
19	24	4	0.39%	23.09%
13	**30**	**1**	**0.10%**	**0.48%**
6	**37**	**1**	**0.10%**	**4.4E − 07**

As can be seen from the above table, for all four intentionally gerrymandered plans the value of π are small enough to identify them as suspect, while none of the unbiased plans are so identified.

References

1. Altman, M.: Modeling the effect of mandatory district compactness on partisan gerrymanders. Polit. Geogr. **17**(8), 989–1012 (1998). https://doi.org/10.1016/S0962-6298(98)00015-8
2. Altman, M., Amos, B., McDonald, M.P., Smith, D.A.: Revealing preferences: why gerrymanders are hard to prove, and what to do about it. Technical report, SSRN 2583528, March 2015. https://doi.org/10.2139/ssrn.2583528
3. Altman, M., McDonald, M.P.: The promise and perils of computers in redistricting. Duke J. Const. Law Public Policy **5**(1), 69–159 (2010)
4. Ansolabehere, S., Leblanc, W.: A spatial model of the relationship between seats and votes. Math. Comput. Model. **48**(9–10), 1409–1420 (2008). https://doi.org/10.1016/j.mcm.2008.05.028
5. Apollonio, N., Becker, R.I., Lari, I., Ricca, F., Simeone, B.: The sunfish against the octopus: opposing compactness to gerrymandering. In: Simeone, B., Pukelsheim, F. (eds.) Mathematics and Democracy: Recent Advances in Voting Systems and Collective Choice, pp. 19–41. Springer, Berlin-Heidelberg (2006). https://doi.org/10.1007/3-540-35605-3_2
6. Aras, A., Costantini, M., van Erkelens, D., Nieuweboer, I.: Gerrymandering in three-party elections under various voting rules (2017)

7. Athreya, K.B.: On a characteristic property of Polya's urn. Stud. Sci. Math. Hung. **4**, 31–35 (1969)
8. Bachrach, Y., Lev, O., Lewenberg, Y., Zick, Y.: Misrepresentation in district voting. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, pp. 81–87 (2016)
9. Balinski, M.L., Demange, G.: Algorithms for proportional matrices in reals and integers. Math. Program. **45**(1–3), 193–210 (1989). https://doi.org/10.1007/BF01589103
10. Balinski, M.L., Demange, G.: An axiomatic approach to proportionality between matrices. Math. Oper. Res. **14**(4), 700–719 (1989). https://doi.org/10.1287/moor.14.4.700
11. Barthélémy, F., Martin, M., Piggins, A.: The architecture of the Electoral College, the house size effect, and the referendum paradox. Electoral. Stud. **34**, 111–118 (2014). https://doi.org/10.1016/j.electstud.2013.07.004
12. Berg, S.: Paradox of voting under an urn model: the effect of homogeneity. Public Choice **47**(2), 377–387 (1985). https://doi.org/10.1007/BF00127533
13. Blais, A.: Turnout in elections. In: Dalton, R.E., Klingemann, H.D. (eds.) Oxford Handbook of Political Behavior, pp. 621–635. Oxford University Press, Oxford (2007). https://doi.org/10.1093/oxfordhb/9780199270125.003.0033
14. Brookes, R.H.: Electoral distortion in New Zealand. Aust. J. Polit. Hist. **5**(2), 218–223 (1959). https://doi.org/10.1111/j.1467-8497.1959.tb01197.x
15. Butler, D.E.: Appendix. In: Nicholas, H.G. (ed.) The British General Election of 1950, pp. 306–333. Macmillan, London (1951)
16. Cain, B.E.: Assessing the partisan effects of redistricting. Am. Polit. Sci. Rev. **79**(02), 320–333 (1985). https://doi.org/10.2307/1956652
17. Campbell, C.D., Tullock, G.: A measure of the importance of cyclical majorities. Econ. J. **75**(300), 853 (1965). https://doi.org/10.2307/2229705
18. Chen, J., Rodden, J.: Unintentional gerrymandering: political geography and electoral bias in legislatures. Q. J. Polit. Sci. **8**(3), 239–269 (2013). https://doi.org/10.1561/100.00012033
19. Chen, J., Rodden, J.: Cutting through the thicket: redistricting simulations and the detection of partisan gerrymanders. Election Law J. Rules Politics Policy **14**(4), 331–345 (2015). https://doi.org/10.1089/elj.2015.0317
20. Cirincione, C., Darling, T.A., O'Rourke, T.G.: Assessing South Carolina's 1990s congressional districting. Polit. Geogr. **19**(2), 189–211 (2000). https://doi.org/10.1016/S0962-6298(99)00047-5
21. Coleman, J.S.: Introduction to Mathematical Sociology. Free Press, New York (1964)
22. Eggenberger, F., Pólya, G.: über die Statistik verketteter Vorgänge. Zeitschrift für Angewandte Mathematik und Mechanik **3**(4), 279–289 (1923). https://doi.org/10.1002/zamm.19230030407
23. Enelow, J.M., Hinich, M.J.: The Spatial Theory of Voting: An Introduction. Cambridge University Press, Cambridge (1984)
24. Engstrom, R.L., Wildgen, J.K.: Pruning thorns from the thicket: an empirical test of the existence of racial gerrymandering. Legis. Stud. Q. **2**(4), 465–479 (1977). https://doi.org/10.2307/439420
25. Fifield, B., Higgins, M., Imai, K.: A new automated redistricting simulator using Markov Chain Monte Carlo. Working Paper, Princeton University, Princeton, NJ (2015)

26. Foos, F., John, P.: Parties are no civic charities: voter contact and the changing partisan composition of the electorate. Polit. Sci. Res. Methods **6**(2), 283–298 (2018). https://doi.org/10.1017/psrm.2016.48
27. Friedman, J.N., Holden, R.T.: Optimal gerrymandering: sometimes pack, but never crack. Am. Econ. Rev. **98**(1), 113–144 (2008). https://doi.org/10.1257/aer.98.1.113
28. Fryer, R.G., Holden, R.: Measuring the compactness of political districting plans. J. Law Econ. **54**(3), 493–535 (2011). https://doi.org/10.1086/661511
29. Garand, J.C., Parent, T.W.: Representation, swing, and bias in U.S. Presidential Elections, 1872–1988. Am. J. Polit. Sci. **35**(4), 1011 (1991). https://doi.org/10.2307/2111504
30. Geenens, G.: Probit transformation for kernel density estimation on the unit interval. J. Am. Stat. Assoc. **109**(505), 346–358 (2014). https://doi.org/10.1080/01621459.2013.842173
31. Gehrlein, W.V., Fishburn, P.C.: Condorcet's paradox and anonymous preference profiles. Public Choice **26**(1), 1–18 (1976). https://doi.org/10.1007/BF01725789
32. Gelman, A., King, G.: Estimating the electoral consequences of legislative redistricting. J. Am. Stat. Assoc. **85**(410), 274–282 (1990)
33. Gelman, A., King, G.: A unified method of evaluating electoral systems and redistricting plans. Am. J. Polit. Sci. **38**(2), 514–554 (1994)
34. Gibbs, A.L., Su, F.E.: On choosing and bounding probability metrics. Int. Stat. Rev. **70**(3), 419–435 (2002). https://doi.org/10.1111/j.1751-5823.2002.tb00178.x
35. Grofman, B.: Measures of bias and proportionality in seats-votes relationships. Polit. Methodol. **9**(3), 295–327 (1983)
36. Grofman, B.: Criteria for districting: a social science perspective. UCLA Law Rev. **33**(1), 77–184 (1985)
37. Grofman, B., King, G.: The future of partisan symmetry as a judicial test for partisan gerrymandering after LULAC v. Perry. Election Law J. Rules Polit. Policy **6**(1), 2–35 (2007). https://doi.org/10.1089/elj.2006.6002
38. Gudgin, G., Taylor, P.J.: Seats, Votes, and the Spatial Organisation of Elections. Pion, London (1979)
39. Issacharoff, S.: Gerrymandering and political cartels. Harvard Law Rev. **116**(2), 593–648 (2002). https://doi.org/10.2307/1342611
40. Jamison, D., Luce, E.: Social homogeneity and the probability of intransitive majority rule. J. Econ. Theory **5**(1), 79–87 (1972)
41. Johnson, N.L., Kotz, S.: Urn Models and Their Applications: An Approach to Modern Discrete Probability Theory. Wiley, New York (1977)
42. Johnston, R.: Manipulating maps and winning elections: measuring the impact of malapportionment and gerrymandering. Polit. Geogr. **21**(1), 1–31 (2002). https://doi.org/10.1016/S0962-6298(01)00070-1
43. Johnston, R.J.: Political, Electoral, and Spatial Systems: An Essay in Political Geography. Contemporary Problems in Geography. Clarendon Press; Oxford University Press, Oxford; New York (1979)
44. Karvanen, J.: Estimation of quantile mixtures via L-moments and trimmed L-moments. Comput. Stat. Data Anal. **51**(2), 947–959 (2006). https://doi.org/10.1016/j.csda.2005.09.014
45. Katz, J.N., King, G.: A statistical model for multiparty electoral data. Am. Polit. Sci. Rev. **93**(1), 15–32 (1999). https://doi.org/10.2307/2585758
46. Kendall, M.G., Stuart, A.: The law of the cubic proportion in election results. Br. J. Sociol. **1**(3), 183–196 (1950). https://doi.org/10.2307/588113

47. King, G., Browning, R.X.: Democratic representation and partisan bias in Congressional elections. Am. Polit. Sci. Rev. **81**(4), 1251 (1987). https://doi.org/10.2307/1962588

48. Kuga, K., Nagatani, H.: Voter antagonism and the paradox of voting. Econometrica **42**(6), 1045–1067 (1974). https://doi.org/10.2307/1914217

49. Lepelley, D., Merlin, V., Rouet, J.L.: Three ways to compute accurately the probability of the referendum paradox. Math. Soc. Sci. **62**(1), 28–33 (2011). https://doi.org/10.1016/j.mathsocsci.2011.04.006

50. Linzer, D.A.: The relationship between seats and votes in multiparty systems. Polit. Anal. **20**(3), 400–416 (2012). https://doi.org/10.1093/pan/mps017

51. Mattingly, J.C., Vaughn, C.: Redistricting and the will of the people. Technical report, arXiv: 1410.8796 [physics.soc-ph], October 2014

52. McClurg, S.D.: The electoral relevance of political talk: examining disagreement and expertise effects in social networks on political participation. Am. J. Polit. Sci. **50**(3), 737–754 (2006). https://doi.org/10.1111/j.1540-5907.2006.00213.x

53. McGhee, E.: Measuring partisan bias in single-member district electoral systems: measuring partisan bias. Legis. Stud. Q. **39**(1), 55–85 (2014). https://doi.org/10.1111/lsq.12033

54. Mcleod, J.M., Scheufele, D.A., Moy, P.: Community, communication, and participation: the role of mass media and interpersonal discussion in local political participation. Polit. Commun. **16**(3), 315–336 (1999). https://doi.org/10.1080/105846099198659

55. Merrill, S.: A comparison of efficiency of multicandidate electoral systems. Am. J. Polit. Sci. **28**(1), 23–48 (1984). https://doi.org/10.2307/2110786

56. Niemi, R.G.: Relationship between votes and seats: the ultimate question in political gerrymandering. UCLA Law Rev. **33**, 185–212 (1985)

57. Niemi, R.G., Deegan, J.: A theory of political districting. Am. Polit. Sci. Rev. **72**(4), 1304–1323 (1978). https://doi.org/10.2307/1954541

58. Niemi, R.G., Fett, P.: The swing ratio: an explanation and an assessment. Legis. Stud. Q. **11**(1), 75–90 (1986). https://doi.org/10.2307/439910

59. Nurmi, H.: Voting Paradoxes and How to Deal with Them. Springer, New York (1999). https://doi.org/10.1007/978-3-662-03782-9

60. O'Loughlin, J.: The identification and evaluation of racial gerrymandering. Ann. Assoc. Am. Geogr. **72**(2), 165–184 (1982). https://doi.org/10.1111/j.1467-8306.1982.tb01817.x

61. Penrose, L.S.: On the Objective Study of Crowd Behaviour. H.K. Lewis, London (1952)

62. Polsby, D.D., Popper, R.D.: The third criterion: compactness as a procedural safeguard against partisan gerrymandering. Yale Law Policy Rev. **9**(2), 301–353 (1991)

63. Pólya, G.: Sur quelques points de la théorie des probabilités. Ann. l'inst. Henri Poincaré **1**(2), 117–161 (1930)

64. Pukelsheim, F.: Biproportional scaling of matrices and the iterative proportional fitting procedure. Ann. Oper. Res. **215**(1), 269–283 (2014). https://doi.org/10.1007/s10479-013-1468-3

65. Regenwetter, M., Grofman, B., Tsetlin, I., Marley, A.: Behavioral Social Choice: Probabilistic Models, Statistical Inference, and Applications. Cambridge University Press, Cambridge (2006)

66. Sano, F., Hisakado, M., Mori, S.: Mean field voter model of election to the House of Representatives in Japan. In: APEC-SSS2016, p. 011016. Journal of the Physical Society of Japan (2017). https://doi.org/10.7566/JPSCP.16.011016

67. Stanton, R.G.: A result of Macmahon on electoral predictions. Ann. Discrete Math. **8**, 163–167 (1980). https://doi.org/10.1016/S0167-5060(08)70866-5
68. Steerneman, T.: On the total variation and hellinger distance between signed measures; an application to product measures. Proc. Am. Math. Soc. **88**(4), 684–688 (1983). https://doi.org/10.2307/2045462
69. Stephanopoulos, N.O., McGhee, E.M.: Partisan gerrymandering and the efficiency gap. Univ. Chicago Law Rev. **82**(2), 831–900 (2015)
70. Szufa, S.: Optimal gerrymandering for simplified districts. Working Paper, Jagiellonian Center for Quantitative Research in Political Science, Kraków (2019)
71. Tideman, T.N., Plassmann, F.: Modeling the outcomes of vote-casting in actual elections. In: Felsenthal, D.S., Machover, M. (eds.) Electoral Systems. Paradoxes, Assumptions, and Procedures, pp. 217–251. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-20441-8_9
72. Tomz, M., Tucker, J.A., Wittenberg, J.: An easy and accurate regression model for multiparty electoral data. Polit. Anal. **10**(1), 66–83 (2002). https://doi.org/10.1093/pan/10.1.66
73. Tufte, E.R.: The relationship between seats and votes in two-party systems. Am. Polit. Sci. Rev. **67**(2), 540–554 (1973). https://doi.org/10.2307/1958782
74. Upton, G.: Blocks of voters and the cube 'law'. Br. J. Polit. Sci. **15**(3), 388–398 (1985). https://doi.org/10.1017/S0007123400004257
75. Wildgen, J.K., Engstrom, R.L.: Spatial distribution of partisan support and the seats/votes relationship. Legis. Stud. Q. **5**(3), 423–435 (1980). https://doi.org/10.2307/439554
76. Yamamoto, T.: A multinomial response model for varying choice sets, with application to partially contested multiparty elections. Working Paper, Massachusetts Institute of Technology, Cambridge, MA (2014)
77. Young, H.P.: Measuring the compactness of legislative districts. Legis. Stud. Q. **13**(1), 105–115 (1988). https://doi.org/10.2307/439947