Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

ELECTORAL DAVID-VS-GOLIATH: PROBABILISTIC MODELS OF SPATIAL DISTRIBUTION OF ELECTORS TO SIMULATE DISTRICT-BASED ELECTION OUTCOMES

Adway Mitra

Centre of Excellence in Artificial Intelligence
Indian Institute of Technology, Kharagpur
Kharagpur, INDIA

ABSTRACT

In district-based elections, electors cast votes in their respective districts. In each district, the party with maximum votes wins the corresponding "seat" in the governing body. The election result is based on the number of seats won by different parties. In this system, locations of electors across the districts may severely affect the election result even if the total number of votes obtained by different parties remains unchanged. A less popular party may end up winning more seats if their supporters are suitably distributed spatially. In this paper, we frame the spatial distribution of electors of a multi-party system in a probabilistic setting, and consider different models to simulate election results, while capturing various properties of realistic elections. We use Approximate Bayesian Computation (ABC) framework to estimate model parameters. We show that our model can reproduce the results of elections held in India and USA, and also produce counterfactual scenarios.

1 INTRODUCTION

Elections are conducted by almost all democratic countries to choose representatives for governing bodies, such as parliaments. A common democratic system is the district-based system in which the country is spatially divided into a number of regions called districts (or constituencies). There is a seat in the governing body corresponding to each district. The residents of each district elect a representative from a set of candidates, according to any voting rule. In many countries, these candidates are representatives of political parties, and electors may cast their votes in favour of the parties rather than individual candidates.

The election results are understood in terms of the number of seats won by different parties, rather than the total number of votes obtained by them. If the relative popularity of the different parties is spatially homogeneous across all the districts, then the most popular party may win all the seats. But this is very rarely the case. One reason for this may be the individual popularity of candidates may vary. But a more complex reason is the spatial variation of demography across the country, since the popularity of different parties often varies with demography (Brooks et al. 2006). Demography varies spatially as people usually prefer to choose residences based on social identities, such as race, religion, language, caste, profession and economic status. This process is sometimes called "ghettoization", where people with similar social identities huddle together in pockets (Dawkins 2004; Dawkins 2007). Such ghettoization plays a very important role in district-based elections if different political parties represent the interests of different social groups. Even if a political party is not popular overall, it can win a few seats if its supporters are densely concentrated in a small number of districts, which forms strongholds of the party. On the other hand, a party which is overall quite popular, may fail to win many seats if its supporters are spread all over without concentration. Also, electors often vote according to the advice of local community leaders and other local factors (Braha and De Aguiar 2017), which causes "polarization" of voters in favour of one/two parties inside each district.

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

Due to these spatial effects, district-based election system doesn't guarantee that the seat distribution of parties is an accurate representation of their relative popularity, leading to questions of fairness (Katz et al. 2020). Since the process of partitioning the country into districts is exogenous to the election, the robustness and comprehensiveness of the results are also questionable. Many countries have the malpractice of "gerrymandering" in which parties having executive powers try to redefine the districts with the aim of maximizing their seats in upcoming elections. For the two-party system of USA, this problem has been studied thoroughly, including recent quantitative analysis by (Chen et al. 2013). Our work is focused on India which has one of the most complex electoral processes in the world with many parties and a highly heterogeneous society where social identities are deeply interlinked with politics.

It is important to lay out a framework that can be used to explore alternative policies for district-based elections to improve its stability, fairness and overall satisfaction of electors. We intend to achieve this by enabling policy-makers to explore election outcomes through realistic simulations, under different conditions related to the overall popularity of the different parties, spatial distribution of supporters of these parties etc. In this work, we look to explore the space of electoral outcomes under any given vote share of the parties by considering different probabilistic models for the spatial distribution of voters across the districts, which are capable of capturing the phenomena like ghettoization and local polarization as discussed above. We demonstrate our results on synthetic data, and also fit the model on real data based on elections in India and USA, which requires parameter estimation. However, the proposed models are "generative-only": from which samples can be drawn but analytical computation of likelihood function is infeasible. So we take the help of likelihood-free inference techniques under the realm of Approximate Bayesian Computation (ABC). For this purpose, we design summary statistics of the election results which are both useful for the ABC techniques but also useful to understand election results. We also modify the ABC Rejection algorithm to make a focused search over the parameter space.

2 RELATED WORKS

2.1 Spatial Bias and Gerrymandering

A significant amount of literature exists in computational social science regarding district-based elections to study how spatial bias can create a difference between overall popularity of parties and the number of district seats won by them. (Chen et al. 2013) point out that this can happen due to either intentional manipulation (gerrymandering) or unintentional effects of political geography in the context of USA. (Chen et al. 2013; DeFord et al. 2020) use the concept of "re-districting through simulations", to observe how election results may change if districts are drawn differently.

Many more works focus on gerrymandering - altering the districts to favour a particular party. (Erdélyi et al. 2015) introduce and examine different algorithms of manipulating elections by splitting and merging districts under a two-round voting scheme. (Lewenberg et al. 2017; Lasisi 2018) consider a setting where a subset of an initial set of districts may be retained and the rest merged, and suggest heuristic algorithms to maximize the number of districts won by a particular party. (Borodin et al. 2018) introduce geometric constraints such as contiguity in defining districts, and explores the relationship between vote share, spatial distribution of voters and the number of districts won in a two-party setting. (Stoica et al. 2020) consider the redrawing of districts with the aim of improving the number of seats of the less popular parties, by utilizing the geometric heterogeneity of vote share. (Lev and Lewenberg 2019) consider a game-theoretic setting in which the electors are rational agents who may relocate to another district to facilitate their preferred outcome. Finally, (Bachrach et al. 2016) define *misrepresentation ratio* caused by spatial effects in a two-party system, and gives theoretical bounds on this quantity using simulated election results.

2.2 Likelihood-free Inference

Every stochastic process involves one or more parameters related to probability distributions. Fitting such models to observations requires us to estimate these parameters. However, well-known parameter estimation approaches need to evaluate the likelihood function, i.e. the probability that the model, under a given parameter setting, will be able to generate the observed data. If the stochastic process is complicated, then analytically calculating this probability may not be tractable. In such a situation, we use *likelihood-free inference*, by either approximating the likelihood function or by directly estimating the posterior distribution by drawing samples. This approach is known as *Approximate Bayesian Computing (ABC)*.

One of the earliest approach to likelihood-free inference was the ABC rejection algorithm (Beaumont et al. 2002; Pritchard et al. 1999). This algorithm samples candidate parameter values from a prior distribution, use these values to run the process simulation, and accept them only if the simulated outcomes are close enough to the observed values. Using the accepted values of the parameters, a posterior distribution over the parameter space, conditioned on the observations, can be calculated. Because comparing the outcomes in full details is often difficult and usually futile, (Wood 2010) suggested that only some *summary statistics* of the outcomes and observations can be computed and compared. Such summary statistics may be either provided by experts of the process, or estimated from the data using neural networks (Åkesson et al. 2020).

One major problem of this particular approach is that most of the samples will be rejected, so that the algorithm will have to run very long. (Engblom et al. 2020) improvise the algorithm to navigate the parameter space more smartly, so that we can move rapidly towards the acceptable parameters. Another body of works tries to predict whether a sample will be acceptable or not, without actually simulating, by training a classifier such as logistic regression (Thomas et al. 2021) or by constructing a *synthetic likelihood* (Gutmann and Corander 2016) for the summary statistics and accepting samples on the basis of such likelihood (Cranmer et al. 2015). Neural networks have been used to learn a parametric approximation of the posterior distribution of the parameters (Papamakarios and Murray 2016; Papamakarios et al. 2016). In some recent works, the rejection process is replaced with regression to map the each observation to a parameter value, using something like a neural network (Lueckmann et al. 2019; Jiang et al. 2017; Åkesson et al. 2020). The data needed to train such neural networks can be obtained by running the simulation for a range of parameter values.

3 MODELS AND ANALYSIS

In this section, we first define the notations of our setting, and then proceed to discuss a series of models for spatial distribution of electors.

3.1 Notation

Let the total number of districts be S. There is one seat in the parliament corresponding to each district. The total number of electors is N, and each elector must register themselves in one district. Let Z_i denote the district in which elector i registers themselves. But each district has a fixed number of electors, denoted by $\{n_1, n_2 \ldots, n_S\}$. Clearly, $n_1 + n_2 + \cdots + n_S = N$. Now, there are K political parties, and the numbers of their supporters are $\{v_1, v_2, \ldots, v_K\}$, such that $v_1 + v_2 + \cdots + v_K = N$. The relative vote shares of these parties can be considered as a K-dimensional discrete distribution, θ. In the subsequent analyses, we consider all the above quantities except Z_i to be fixed and known, unless otherwise stated.

In the electoral setting, let the number of votes polled by the different parties at any district s be denoted by $\{V_{s1}, V_{s2}, \ldots, V_{sK}\}$. Clearly, $\sum_{k=1}^{K} V_{sk} = n_s$ and $\sum_{s=1}^{S} V_{sk} = v_k$. In any district s, the winner W_s is that party which receives the highest number of votes in that district, i.e. $W_s = argmax_k(V_{s1}, V_{s2}, \ldots, V_{sK})$. In each district, the "winning margin" P_s is the fraction of votes won by the winning party, i.e. $P_s = \frac{V_{sW_s}}{n_s}$. The number of seats M_k won by any party k is the number of districts where it is the winner, i.e. $M_k = \sum_{s=1}^{S} I(W_s = k)$ (here I denotes the indicator function). In the analyses below, some or all of Z, V, W, P and M are considered as random variables.

3.2 District-wise Model (DM)

In the first model, we consider that each district s has its own relative popularity of the parties, $\{\theta_s\}$ which is related to the overall popularity θ. θ_s for each district is drawn from a Dirichlet distribution with parameter vector θ, which is fixed and known. Each elector in the district then votes by sampling by θ_s, and the winner will be the mode of θ_s. To constrain the total number of votes obtained by each party $(v_1,\ldots,v_K)$, we deactivate the choice of each party once the number of votes it gets from the different districts reaches the stipulated value v_k. This count is maintained by book-keeping variable m_k. If we denote by X_{si} the vote of the i-th voter in district s, then its distribution is expressed by by Equation (1) as follows:

$$\theta_s \sim Dirichlet(\theta); prob(X_{si} = k) \propto \theta_k I(m_k < v_k) \tag{1}$$

3.3 District-wise Polarization Model (DPM)

Next we consider the effect of local polarization, where in each district the voters choose a party, based on local popularity. If n_{sk} electors in district s have already expressed support for party k, a new elector in that district will choose k based on n_{sk}, but will also account for its country-wide popularity θ_k. This model is a realistic representation of the voting behavior in many countries, where people often make a trade-off between the local candidate and the top leadership of a party before choosing to vote for it. This model is based on the famous Chinese Restaurant Process (CRP) (Pitman 1995), and expressed by Equation (2). Once again, we use the book-keeping variables to keep track of the total number of votes obtained by each party as the process proceeds.

$$prob(X_{si} = k) \propto (\gamma_s n_{sk} + (1-\gamma_s)\theta_k) I(m_k < v_k) \tag{2}$$

Here γ_s is the *polarization parameter* specific to district s. A high value of γ_s indicates that electors in that district tend to choose the locally popular party, with less influence of the overall popularity of the parties indicated by θ, and it creates the possibility of diversity across the districts. If γ_s is low in all districts, then the proportion of votes will reflect θ everywhere, and almost all districts will have the same winner.

3.4 Elector Community Model (ECM)

This model is based Hierarchical Dirichlet Process (HDP) (Teh et al. 2005) for grouped data. The HDP first considers a measure P, which follows stick-breaking or GEM distribution. Next, for every data group i, a measure Q_i is created from P using a stick-breaking process. Finally, n_i samples are drawn from Q_i, as the data-points associated with the group i. In this case, we can identify each group as a district, and n_s as the number of electors in district s. The base distribution H can be considered as the overall vote share θ, and Q_s is the vote share of the K parties specific to the district s. Accordingly n_s votes are polled for the different parties, as $\{V_{s1},\ldots,V_{sK}\}$ by sampling from the distribution Q_s, and the winners are calculated.

This model becomes more interesting and suitable for the voting scenario when we consider the Chinese Restaurant Franchise (CRF) representation of the HDP, which is obtained by marginalizing over P and Q. In our setting, the electors within each district s first form communities among themselves (which we denote by C) according to Equation (3), and then all the members of a community vote for the same party (denoted by D) according to Equation (4). This is a common feature in the elections of many countries, as people vote according to the influence of their social communities rather than by individual choice. The communities are not uniformly sized, rather there are a few big and many small communities, due to the self-reinforcing ("rich getting richer") nature of Equation (3). Each community tends to vote for a party which is already popular in other communities. This is also a realistic feature of elections in many countries, where people have a tendency to vote for that party whom they consider the strongest.

- *if n_{sj} electors have joined community j :*

$$\textit{for elector i: } prob(C_{si} = j) = \frac{n_{sj}}{i - 1 + \alpha_s} \tag{3}$$

- *if j is a new community : $prob(C_{si} = j) = \frac{\alpha_s}{i-1+\alpha_s}$*
- *if v_k communities have voted for party k :*

$$\textit{for community c: } prob(D_c = k) = \frac{v_k + \beta.\theta}{c - 1 + \beta} \tag{4}$$

Here, α and β are two parameters of the model. High value of α indicates formation of many small communities within each district, while high value of β creates high polarization *across districts*, a situation where a small number of parties account for most of the votes. Once again, to make sure the parties get votes according to pre-specified θ, we include book-keeping variables in Equation (4).

3.5 Party-wise Concentration Model (PCM)

Now we consider a model for the distribution of support of each party across the districts. The effect of this model is to create local concentrations of support in favour of different parties, which helps them to be effective in district-based elections. It is also a realistic phenomena, because support to political party is often based on social identities, and in most countries people choose residential areas based on social identities. For this model, we once again use the Chinese Restaurant Process model as in the District-wise Polarization Model. But this time we make the process two-step: each person i is first assigned to a party X_i, then (s)he is assigned a district Z_i based on concentration of support for that party.

The model is governed by Equation (5). In this model, $\{\eta_1, \ldots, \eta_K\}$ are the *concentration parameters*. High value of the parameter η_k encourages voters of party k to concentrate in a few districts, instead of spreading out uniformly. If all parties have low value of η_k, then once again the vote distribution in all districts will mirror θ, and the most popular party overall will win all seats. Concentration of votes is particularly beneficial to parties which are less popular overall, it allows them to create local strongholds where they can win, even if they are non-existent elsewhere.

$$prob(X_i = k) \propto \theta_k I(m_k < v_k); \quad prob(Z_i = s | X_i = k) \propto (\eta_k V_{sk} + (1 - \eta_k)U(1, K))I(l_s < n_s) \tag{5}$$

m_k, l_s are book-keeping variables to make sure that the total number of votes obtained by each party and the capacity of each district is maintained.

4 PARAMETER ESTIMATION BY APPROXIMATE BAYESIAN COMPUTATION

Clearly, these models have many parameters. To explain and analyze the results of actual elections using the models, we need to estimate these parameters. For parameter estimation, well-known approaches like maximum-likelihood and Bayesian estimation are intractable, due to the lack of a closed-form expression of the likelihood function, especially because of the book-keeping variables. So we look to likelihood-free inference techniques using Approximate Bayesian Computation (ABC).

For most ABC approaches, we need a low-dimensional representation of the data-points using which we wish to compute the posterior approximately. One well-known way of getting such a low-dimensional representation is by using summary statistics defined by the user, specific to the problem at hand. In this case, we too define the following summary statistics, which can be easily calculated from $\{V_{s1}, \ldots, V_{sK}\}$, i.e. the number of votes obtained by each party in each district. The summary statistics we considered here are as follows: i) Number of districts "won" by each party according to voting rule (K-dim) ii) The mean fraction of votes won by each party across all districts (K-dim) iii) The mean and standard deviation of the winning margin (MWM, SWM) across all districts (2-dim)

4.1 ABC Explore-Exploit Rejection Algorithm

The simplest approach for this problem of parameter estimation is the ABC Rejection algorithm (Pritchard et al. 1999; Cranmer et al. 2015; Thomas et al. 2021). Here we sample candidate values for model parameters (denoted by ψ) from a suitable prior distribution, and use them to run the simulation and get the result x, from which we calculate the summary statistics $S(x)$. Next, we compare $S(x)$ with the summary statistics $S(x_0)$ computed from the observed value x_0. If they are close enough to each other, then we accept the sample of the parameters, otherwise we reject it.

The big problem with this approach is that it is very slow, as most samples are rejected. Once a sample is accepted, we may search in the neighborhood of the accepted sample rather than sampling again from the prior, but then we may get stuck at a local optima in the parameter space. So we use the *explore-exploit* approach, where we first draw a limited number of samples from the prior and choose the best few among them as seeds (*explore* phase), and then we draw more samples around them, by using Gaussian distribution (*exploit* phase). We accept those samples for which the simulation summary statistics are close enough to $S(x_0)$. The process is repeated until we have a large enough set of samples. We then find that sample which creates the simulation summary statistics which is closest to the observed data, and use it as the optimal estimate ψ_{OPT}. We call this as ABC Explore-Exploit Rejection, which is a modified version of SLAM algorithm (Engblom et al. 2020).

4.2 Regression-Rejection Hybrid Approach

Since rejection-based ABC algorithms are slow, there have been attempts to augment them with supervised learning. Here, we try to estimate the model parameters by posing this as a regression problem. We first generate a training set by drawing N samples of the model parameters $\{y_i\}_{i=1}^N$, and also the popularity proportion θ of the parties, the number of electors N and number of districts S. Once we have a fairly diverse but representative set of samples, we run the models on them, to get the outputs V. Using N, S, θ, V we construct the feature vectors $\{x_i\}_{i=1}^N$, along with the parameters as corresponding output values $\{y_i\}_{i=1}^N$. Now, given any feature vector x and a model parameter estimate y', we can train a logistic regression classifier to predict whether the correct parameter value y (by which the model can produce the output encoded in x from the input encoded in x) is greater than y' or not. Depending on the classifier's prediction, we can shift y' either above or below its current value, midway within the feasible region (for example γ of DPM and η of PCM must lie in $(0,1)$). The feasible region also shrinks with each update of y'. We continue this process till y' converges. For every test value y', we train the logistic regression classifier on the training $\{x_i, y_i\}_{i=1}^n$ set by re-labeling all samples based on $sign(y_i - y')$. If the model has multiple parameters (ECM and PCM), we do this separately for all parameters.

The above approach was found to return parameter values which are close to the optimal solution, but often not close enough. Hence, we use a hybrid approach combining regression and rejection. We replace the "explore" phase of ABC Explore-Exploit Rejection algorithm with the regression-based approach, to get an estimate called y'_0. The candidate set of the parameters can then be found using the *exploit* phase of the same algorithm.

5 ANALYSIS OF MODELS THROUGH SIMULATION

In this section, we illustrate through simulations various aspects of our models on synthetic data in a 2-party system. Keeping the number of districts $S = 100$ and number of electors $N = 1000000$, we vary the popularity proportion θ, and observe the results under different settings of the parameters of our models. For each setting, we carry out 100 simulations. The number of seats won by the different parties is noted in each simulation, and the mean number of seats won by each party across these simulations is reported. The minimum and maximum values of these numbers are also noted and indicated in Tables 1, 2, 3 as 75 ± 3 (suggesting a range of 72 and 78). In later tables, we omit this range as it is small/negligible. In

	Number of seats won by Party 1			St. Dev. Party 1 voteshare		
$\gamma\|\theta_1$	0.55	0.6	0.7	0.55	0.6	0.7
-	57 ± 2	63 ± 2	74 ± 2	0.32	0.31	0.3
0.25	100	100	100	0.01	0.01	0.01
0.50	100	100	100	0.01	0.01	0.01
0.75	80 ± 3	96 ± 2	100	0.06	0.06	0.06
0.90	61 ± 3	72 ± 3	83 ± 3	0.15	0.16	0.18
0.99	55 ± 3	60 ± 3	70 ± 2	0.30	0.30	0.31

Table 1: Synthetic results in a two-party election under different parameter settings (shown in left column) of District-wise Model (first row), and District-wise Model Polarization Model (rows 2-6), showing number of seats won by party 1 (popularity θ_1 indicated for each column) and the st. dev. of its votes across districts.

	Number of seats won by Party 1			St. Dev. Party 1 voteshare		
$(\alpha,\beta)\|\theta_1$	0.55	0.6	0.7	0.55	0.6	0.7
$\{1,0.25\}$	55 ± 3	61 ± 3	72 ± 3	0.36	0.35	0.33
$\{1,0.75\}$	55 ± 3	61 ± 3	72 ± 3	0.39	0.37	0.34
$\{1,0.99\}$	55 ± 3	60 ± 3	70 ± 3	0.42	0.4	0.37
$\{5,0.25\}$	61 ± 4	71 ± 4	88 ± 2	0.18	0.18	0.17
$\{5,0.75\}$	58 ± 6	69 ± 8	83 ± 5	0.23	0.21	0.21
$\{5,0.99\}$	54 ± 9	61 ± 9	71 ± 8	0.37	0.37	0.33
$\{10,0.25\}$	67 ± 3	80 ± 3	95 ± 2	0.14	0.14	0.14
$\{10,0.75\}$	60 ± 12	75 ± 9	88 ± 7	0.2	0.18	0.19
$\{10,0.99\}$	56 ± 12	66 ± 15	73 ± 12	0.36	0.26	0.31

Table 2: Synthetic results in a two-party election under different parameter settings (shown in left column) of Elector Community Model, showing number of seats won by party 1 (popularity θ_1 indicated for each column) and the st. dev. of its votes across districts.

Tables 1,2,3 we also report the variance in the vote share of the more popular party across the districts and across all the 100 simulations.

The results presented in Tables 1, 2, 3 show that the party which is more popular overall, often has an extra advantage in terms of seats won, i.e. the proportion of seats won by them usually exceeds their proportion of popularity (vote share). However, it is possible for the smaller party to offset this disadvantage in certain circumstances. In case of DM, the seats won by both parties are roughly proportional to their vote shares, though this model does not account for any spatial concentration of support. In case of DPM, high value of γ causes the supporters of both parties to concentrate and form *stronghold districts*, and the seat-share reflects the vote share. In case of ECM, we see that increasing α, i.e. having many small communities helps the larger party, but increasing β, i.e. concentration of support benefits the less popular party as its seat-share approaches and may even exceed ($\alpha = 10, \beta = 0.99$) its vote-share. In case of PCM, we see that the bigger party gains by having less concentration of its supporters, but the smaller party gains with greater concentration. An anomalous situation arises when the bigger party is too concentrated while the smaller party is spread all over ($\eta_1 = 0.99, \eta_2 = 0.5$) - the bigger party may lose the election in terms of seats even with 55% popularity. This is precisely the situation analyzed by works like (Chen et al. 2013) in the context of US presidential elections in a two-party system, where the Democratic candidate may lose despite higher vote-share (e.g. US Presidential Election, 2016), as their supporters are concentrated in dense cities while the Republican supporters are spread across the sparse countrysides.

5.1 Simulation of 3-party elections

Next, we explore a three-party system which has the potential to become more intriguing. We consider 4 values of θ, i.e. the relative popularity or vote share of the three parties - i) $\theta = (0.5, 0.4, 0.1)$ where there are two strong parties and a much weaker one, ii) $\theta = (0.4, 0.3, 0.3)$ where there is one strong party and two equally popular parties, iii) $\theta = (0.4, 0.35, 0.25)$ with two almost equally popular parties and a slightly less popular one, and iv) $\theta = (0.37, 0.33, 0.3)$ where there are three parties with nearly equal popularity.

	Number of seats won by Party 1			St. Dev. Party 1 voteshare		
$(\eta_1,\eta_2)\|\theta_1$	0.55	0.6	0.7	0.55	0.6	0.7
$\{0.5,0.5\}$	100	100	100	0.02	0.02	0.02
$\{0.5,0.7\}$	94 ± 2	100	100	0.02	0.02	0.02
$\{0.5,0.99\}$	67 ± 1	77 ± 2	90 ± 1	0.05	0.05	0.04
$\{0.7,0.5\}$	97 ± 2	100	100	0.06	0.07	0.07
$\{0.7,0.7\}$	89 ± 2	99 ± 1	100	0.06	0.07	0.7
$\{0.7,0.99\}$	68 ± 2	75 ± 3	90 ± 2	0.09	0.09	0.08
$\{0.99,0.5\}$	52 ± 3	61 ± 3	80 ± 1	0.34	0.34	0.32
$\{0.99,0.7\}$	52 ± 3	62 ± 3	80 ± 3	0.34	0.34	0.32
$\{0.99,0.99\}$	58 ± 3	65 ± 3	80 ± 2	0.34	0.34	0.33

Table 3: Synthetic results in a two-party election under different parameter settings (shown in left column) of Partywise Concentration Model, showing number of seats won by party 1 (popularity θ_1 indicated for each column), and the st. dev. of its votes across districts.

θ	$\gamma=0.8$			$\gamma=0.9$		
	nA	nB	nC	nA	nB	nC
$(0.5,0.4,0.1)$	54	41	5	-	-	-
$(0.4,0.3,0.3)$	44	29	29	-	-	-
$(0.4,0.35,0.25)$	43	35	22	-	-	-
$(0.37,0.33,0.3)$	39	33	28	-	-	-
$(0.5,0.4,0.1)$	78	22	0	60	35	5
$(0.4,0.3,0.3)$	66	17	17	48	26	26
$(0.4,0.35,0.25)$	61	32	7	47	34	19
$(0.37,0.33,0.3)$	50	30	20	41	33	26

Table 4: Synthetic results in a 3-party election by the District-wise Model (rows 1-4), and under different parameter settings of the District-wise Model Polarization Model (rows 5-8), showing number of seats won by the parties corresponding to popularity θ indicated in the left column.

In Table 4 we show the results for the District-wise Model (DM) and District-wise Polarization Model (DPM) under two values of the concentration parameter γ: 0.8 and 0.9. In case of DM, we find that the parties win seats in proportion to their popularity, with a slight additional advantage to the most popular party. In case of DPM, it is found that low values of concentration causes almost all seats to go to the most popular party, while high concentration causes the seat share to approach θ. With moderately high values of concentration, as shown in Table 4, we find potentially interesting results. The numbers reported in the table are the mean of 100 simulations. These results are also quite robust, with a variance of only about 3 seats for each party.

In case of the Elector Community Model, we show 2 settings for each parameter: $\alpha \in \{20,50\}$ and $\beta \in \{0.5,0.8\}$. The resultant 4 parameter combinations are shown in Table 5. It is found that increasing α, that tends to create many small communities, gives advantage to the most popular party, while increasing β causes the seat share towards θ, just like the γ parameter of DPM. The figures reported in Table 5 are the mean of 100 simulation runs, and the variance is also quite large - about 12 seats, especially when both α and β are high. This means that in case of close margins, there are some simulation runs where a less popular party ends up winning more seats than a more popular one.

Next, we come to the most interesting case of Partywise Concentration Model. Here we consider two concentration values of each party: 0.5 (low) and 0.99 (high). Each party can have its own concentration, independent of the others. When all three parties have low concentration, the most popular party tends to win almost all the seats, and when all three parties have high concentration the seat share is similar to θ. However, other combinations are most fascinating, which are shown in Table 6. For the most popular party, low concentration is generally better than high concentration. In fact, the most popular party is likely to lose the election if it is concentrated and the second party is not, if the difference between their vote shares is low. For the second party, its performance depends on the other parties. If the first party's support is concentrated, then it is beneficial for the second party to be diffused. But if the first party's support is diffused, then the second party's best chance of maximizing its seats is by concentrating its support. For

θ	$\alpha = 20, \beta = 0.5$			$\alpha = 20, \beta = 0.8$		
	nA	nB	nC	nA	nB	nC
$(0.5, 0.4, 0.1)$	70	30	0	67	31	2
$(0.4, 0.3, 0.3)$	64	18	18	56	22	22
$(0.4, 0.35, 0.25)$	57	34	9	51	34	15
$(0.37, 0.33, 0.3)$	49	31	20	39	33	28

θ	$\alpha = 50, \beta = 0.5$			$\alpha = 50, \beta = 0.8$		
	nA	nB	nC	nA	nB	nC
$(0.5, 0.4, 0.1)$	78	22	0	51	48	1
$(0.4, 0.3, 0.3)$	70	15	15	66	17	17
$(0.4, 0.35, 0.25)$	62	34	4	56	37	7
$(0.37, 0.33, 0.3)$	53	29	18	48	28	24

Table 5: Synthetic results in a 3-party election under different parameter settings (mentioned along the columns) of Elector Community Model, showing number of seats won by the parties corresponding to popularity θ indicated in the left column.

θ	$\eta = (0.99, 0.50, 0.50)$			$\eta = (0.50, 0.99, 0.50)$		
	nA	nB	nC	nA	nB	nC
$(0.5, 0.4, 0.1)$	53	47	0	68	32	0
$(0.4, 0.3, 0.3)$	50	25	25	72	28	0
$(0.4, 0.35, 0.25)$	45	55	0	67	33	0
$(0.37, 0.33, 0.3)$	45	51	4	66	34	0

θ	$\eta = (0.50, 0.99, 0.99)$			$\eta = (0.50, 0.50, 0.99)$		
	nA	nB	nC	nA	nB	nC
$(0.5, 0.4, 0.1)$	69	31	0	100	0	0
$(0.4, 0.3, 0.3)$	52	24	24	72	0	28
$(0.4, 0.35, 0.25)$	55	30	15	78	2	20
$(0.37, 0.33, 0.3)$	47	29	24	67	4	29

θ	$\eta = (0.99, 0.99, 0.50)$			$\eta = (0.99, 0.50, 0.99)$		
	nA	nB	nC	nA	nB	nC
$(0.5, 0.4, 0.1)$	56	42	2	51	48	1
$(0.4, 0.3, 0.3)$	43	26	31	42	31	27
$(0.4, 0.35, 0.25)$	44	35	21	42	41	18
$(0.37, 0.33, 0.3)$	36	32	32	37	37	26

Table 6: Synthetic results in a 3-party election under different parameter settings (mentioned along the columns) of Partywise Concentration Model, showing number of seats won by the parties corresponding to popularity θ indicated in the left column.

the third party, concentration seems to be the best option always, except the case where both the other parties are also concentrated. In the latter case, the third party having diffuse support may help to win more seats than its vote share.

6 ANALYSIS OF ELECTIONS IN INDIA AND USA

Finally, we attempt to fit our models to actual election results. We use the hybrid regression-rejection algorithm discussed earlier to find the optimal parameters to fit each model to each election. We also explore alternative results to these elections, under the same proportion of popularity (i.e. vote share) of the contestant parties, but different parameter settings.

First, we consider the Indian province of Delhi National Capital Region (NCR), whose local assembly has 70 seats. Around 9 million people (on average) participate in the elections, with roughly equal distribution of electors across the 70 districts. Since 2013, 5 elections have taken (local and national) in which 3 main political parties (anonymized here as A,B,C) have competed, along with small parties and independent candidates. The overall vote-shares (popularity proportions) of the parties have varied across the elections. The data has been collected from (Election Commission of India).

In Table 7, we show the expected results according to our models, and compare them with the actual results for the local assembly elections of 2013, 2015 and 2020. Due to lack of space, only two parameter settings are shown: default and optimal. In default settings, we assume maximum polarization and concentration (0.99) for all parties. Optimal settings are estimated using the hybrid regression-rejection algorithm discussed in the previous section. We find that in all cases, PCM and ECM are able to recreate the actual results under the optimal parameter settings. The same parameters may be used to simulate the national election results (2014, 2019) too.

Next, we consider the presidential elections in the United States of America in 2016 and 2020. The results are obtained from (Federal Election Commission a) (2016) and (Federal Election Commission b) (2020). We considered the two main parties - Democratic (D) and Republican (R), while neglecting other candidates. We considered 56 states or districts (Washington D.C. has 1, Maine 2 and Nebraska 3

	Party A	Party B	Party C	MWM	SWM
2013θ	0.30	0.33	0.25	NA	NA
Proportional	21	23	18		
DM	23	27	18	0.58	0.02
DPM(0.89)	**25**	**35**	10	0.37	0.07
DPM(0.99)	21	**24**	18	0.55	0.18
ECM($\{16, 0.24\}$)	**28**	**34**	8	0.39	0.07
ECM($\{1, 0.99\}$)	21	23	18	0.77	0.19
PCM($\{0.55, 0.89, 0.84\}$)	**27**	**34**	9	0.36	0.09
PCM($\{0.99, 0.99, 0.99\}$)	**24**	**28**	18	0.51	0.11
Actual	**28**	**34**	8	0.39	0.06

2015θ	0.54	0.32	0.10	NA	NA
Proportional	38	22	7	NA	NA
DM	44	23	3	0.7	0.18
DPM(0.86)	**68**	2	0	0.54	0.09
DPM(0.99)	37	23	4	0.71	0.21
ECM($\{30, 0.21\}$)	**67**	3	0	0.55	0.07
ECM($\{1, 0.99\}$)	38	23	7	0.87	0.17
PCM($\{0.74, 0.89, 0.68\}$	**67**	3	0	0.54	0.05
PCM($\{0.99, 0.99, 0.99\}$	**49**	19	2	0.64	0.14
Actual	**67**	3	0	0.55	0.07

2020θ	0.54	0.39	0.05	NA	NA
Proportional	37	27	4	NA	NA
DM	42	27	1	0.73	0.16
DPM(0.87)	**60**	10	0	0.55	0.08
DPM(0.99)	38	28	4	0.73	0.18
ECM($\{36, 0.57\}$)	**62**	8	0	0.55	0.06
ECM($\{1, 0.99\}$)	**40**	25	4	0.85	0.17
PCM($\{0.72, 0.80, 0.72\}$)	**62**	8	0	0.55	0.08
PCM($\{0.99, 0.99.0.99\}$)	**43**	27	0	0.66	0.13
Actual	**62**	8	0	0.55	0.06

Table 7: Elections in Delhi-NCR, India: The actual and model-predicted performances of 3 top parties in past 3 assembly elections (2013, 2015, 2020), based on their popularity proportions θ (vote share). For each model, results are shown with the default parameters as well as optimal settings as computed by Hybrid Regression-Rejection Algorithm. In each case, the number of seats won by each party is compared with seats proportional to their vote share, and the cases where a party gains seats are highlighted

	Party D	Party R	MWM	SWM
2016θ	0.51	0.49	NA	NA
Proportional	29	27		
PCM($\{0.99, 0.02\}$)	22	**34**	0.64	0.08
PCM($\{0.99, 0.99\}$)	28	**28**	0.7	0.14
Actual	22	**34**	0.6	0.08

	Party D	Party R	MWM	SWM
2020θ	0.52	0.48	NA	NA
Proportional	29	27		
PCM($\{0.95, 0.5\}$)	28	**28**	0.59	0.06
PCM($\{0.99, 0.99\}$)	**33**	23	0.7	0.12
Actual	28	**28**	0.6	0.08

Table 8: US Presidential Elections 2016 and 2020: The actual and model-predicted performances of 2 main parties, based on their adjusted popularity proportions θ (ignoring smaller parties). For the PCM model, results are shown with the default parameters as well as optimal settings as computed by Hybrid Regression-Rejection Algorithm. In each case, the number of districts/states won by each party (out of 56) is compared with seats proportional to their vote share

Congressional districts). The total number votes cast in favour of these two parties were considered for all the states/districts, to estimate the number of electors n_s in each district, and their overall popularity proportion θ. We simulated both the elections using the models. However, in 2016 Presidential elections, the party with lower overall popularity won more districts/states due to variation of spatial concentrations of the electors, and this effect can be captured only by the Partywise Concentration Model (PCM). The results of PCM simulations under optimal and default parameter settings are shown in Table 8. It needs to be noted that unlike the elections in India, the different states/districts have widely varying electorate size, leading to uncertainties in the simulations even under same parameter settings. For example, the same number of votes can be utilized to win one big state like California, or several small states/districts. Hence, we run simulations 10 times for each parameter settings and choose the most likely results. In the US electoral system, each states/districts has a certain number of *electoral votes*, but we do not utilize this aspect in our simulations as the number of electoral votes in each district is not proportional to its population. We note that in 2016, there was a huge difference in the level of concentration of the two parties, unlike 2020. Also, with equally high level of concentration (0.99), 2016 could have seen a close result while 2020 could have seen a major Democratic victory.

7 CONCLUSION

In this paper, we explored the effects of spatial distribution of electors on district-based elections. Specifically, these models allow us to explore the space of possible results of elections under fixed number of popular votes of the parties. We considered 4 stochastic models for such spatial distribution, based on the observations by political scientists that i) an individual's vote is influenced by overall popularity of parties, ii) people tend to vote for parties based on their social identity, and iii) people with similar social identity tend to live geographically close. We demonstrated that our models can reproduce observed election results in India and USA using likelihood-free inference for parameter estimation, and can create counterfactual scenarios too. This work is the first step in the direction of building a simulation-based framework for exploration of district-based elections. Such a framework may be used to recommend alternative electoral processes, which will be i) robust to gerrymandering of district boundaries, ii) reduce the skew between vote share and seat share, iii) minimize the number of people whose choices are not represented at any level. It may be noted that while this work may be used by political parties to optimize their electoral strategies, it does not enable them or any other malicious agents to influence electoral processes unethically. This is because this work does not provides any way to re-organize the districts (gerrymandering), and also the simulations by this model are conditional on the popularity proportion (vote share) of the different parties, which are not known before any election. Hence this work cannot be used to bias the outcomes of future elections. The utility of this work is purely in diagnostic purposes, to perform various "what-if" analyses.

REFERENCES

Åkesson, M., P. Singh, F. Wrede, and A. Hellander. 2020. "Convolutional Neural Networks as Summary Statistics for Approximate Bayesian Computation". *arXiv preprint arXiv:2001.11760*.

Bachrach, Y., O. Lev, Y. Lewenberg, and Y. Zick. 2016. "Misrepresentation in District Voting". In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*. July 9th-16 th, New York, USA, 81–87.

Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. "Approximate Bayesian Computation in Population Genetics". *Genetics* 162(4):2025–2035.

Borodin, A., O. Lev, N. Shah, and T. Strangway. 2018. "Big City vs. the Great Outdoors: Voter Distribution and how it Affects Gerrymandering". In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. July 13th-19th, Stockholm, Sweden, 98–104.

Braha, D., and M. A. De Aguiar. 2017. "Voting Contagion: Modeling and Analysis of a Century of US Presidential Elections". *PloS one* 12(5):1–30.

Brooks, C., P. Nieuwbeerta, and J. Manza. 2006. "Cleavage-based Voting Behavior in Cross-National Perspective: Evidence from Six Postwar Democracies". *Social Science Research* 35(1):88–128.

Chen, J., J. Rodden et al. 2013. "Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures". *Quarterly Journal of Political Science* 8(3):239–269.

Cranmer, K., J. Pavez, and G. Louppe. 2015. "Approximating Likelihood Ratios with Calibrated Discriminative Classifiers". *arXiv preprint arXiv:1506.02169*.

Dawkins, C. J. 2004. "Measuring the Spatial Pattern of Residential Segregation". *Urban Studies* 41(4):833–851.

Dawkins, C. J. 2007. "Space and the Measurement of Income Segregation". *Journal of Regional Science* 47(2):255–272.

DeFord, D., M. Duchin, and J. Solomon. 2020. "A Computational Approach to Measuring Vote Elasticity and Competitiveness". *Statistics and Public Policy* 7(1):69–86.

Election Commission of India. "Election Results - Full Statistical Reports". https://eci.gov.in/statistical-report/statistical-reports/, accessed 5.03.2020.

Engblom, S., R. Eriksson, and S. Widgren. 2020. "Bayesian Epidemiological Modeling over High-Resolution Network Data". *Epidemics* 32(100399):1–9.

Erdélyi, G., E. Hemaspaandra, and L. A. Hemaspaandra. 2015. "More Natural Models of Electoral Control by Partition". In *Proceedings of the 4th International Conference on Algorithmic Decision Theory*. September 27th-30th, Kentucky, USA, 396–413.

Federal Election Commission. "Federal Elections 2016". https://www.fec.gov/resources/cms-content/documents/federalelections2016.pdf, accessed 15.04.2021.

Federal Election Commission. "Official 2020 Presidential General Election Results". https://www.fec.gov/resources/cms-content/documents/2020presgeresults.pdf, accessed 15.04.2021.

Gutmann, M. U., and J. Corander. 2016. "Bayesian Optimization for Likelihood-free Inference of Simulator-based Statistical Models". *The Journal of Machine Learning Research* 17(1):4256–4302.

Jiang, B., T.-y. Wu, C. Zheng, and W. H. Wong. 2017. "Learning Summary Statistic for Approximate Bayesian Computation via Deep Neural Network". *Statistica Sinica*:1595–1618.

Katz, J. N., G. King, and E. Rosenblatt. 2020. "Theoretical Foundations and Empirical Evaluations of Partisan Fairness in District-based Democracies". *American Political Science Review* 114(1):164–178.

Lasisi, R. O. 2018. “Improved Manipulation Algorithms for District-Based Elections”. In *Proceedings of the 31st International Conference of the Florida Artificial Intelligence Research Society*. May 21st-23 rd, Florida, USA, 44–49.
Lev, O., and Y. Lewenberg. 2019. ““Reverse Gerrymandering”: Manipulation in Multi-Group Decision Making”. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Jan 27th-Feb 1st, Honolulu, Hawaii, USA, 2069–2076.
Lewenberg, Y., O. Lev, and J. S. Rosenschein. 2017. “Divide and Conquer: Using Geographic Manipulation to Win District-based Elections”. In *Proceedings of the 16th International Conference on Autonomous Agents and Multi Agent Systems*. May 8th-12th, Sao Paulo, USA, 624–632.
Lueckmann, J.-M., G. Bassetto, T. Karaletsos, and J. H. Macke. 2019. “Likelihood-free Inference with Emulator Networks”. In *Proceedings of the 3rd Symposium on Advances in Approximate Bayesian Inference*. Jan-Feb 2019, Virtual Event, 32–53.
Papamakarios, G., and I. Murray. 2016. “Fast ε-free Inference of Simulation Models with Bayesian Conditional Density Estimation”. In *Proceedings of International Conference on Advances in Neural Information Processing Systems*. Dec 5th-10th, Barcelona, Spain, 1028–1036.
Papamakarios, G., D. Sterratt, and I. Murray. 2016. “Sequential Neural Likelihood: Fast Likelihood-Free Inference with Autoregressive Flows”. In *Proceedings of 22nd International Conference on Advances in Neural Information Processing Systems*. April 16th-18th, Okinawa, Japan, 837–848.
Pitman, J. 1995. “Exchangeable and Partially Exchangeable Random Partitions”. *Probability theory and related fields* 102(2):145–158.
Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. 1999. “Population Growth of Human Y Chromosomes: a Study of Y Chromosome Microsatellites”. *Molecular Biology and Evolution* 16(12):1791–1798.
Stoica, A.-A., A. Chakraborty, P. Dey, and K. P. Gummadi. 2020. “Minimizing Margin of Victory for Fair Political and Educational Districting”. In *Proceedings of the 19th International Conference on Autonomous Agents and Multi Agent Systems*. May 9th-13th, Auckland, New Zealand, 1305-1313.
Teh, Y. W., M. I. Jordan, M. J. Beal, and D. M. Blei. 2005. “Sharing Clusters among Related Groups: Hierarchical Dirichlet processes”. In *Proceedings of the 18th International Conference on Advances in Neural Information Processing Systems*. December 5th-8 th, Vancouver, Canada, 1385–1392.
Thomas, O., R. Dutta, J. Corander, S. Kaski, and M. U. Gutmann. 2021. “Likelihood-free Inference by Ratio Estimation”. *Bayesian Analysis* TBA(TBA):1–31.
Wood, S. N. 2010. “Statistical Inference for Noisy Nonlinear Ecological Dynamic Systems”. *Nature* 466(7310):1102–1104.

AUTHOR BIOGRAPHIES

ADWAY MITRA is an assistant professor in the Centre of Excellence in Artificial Intelligence at Indian Institute of Technology Kharagpur. His research interests include probabilistic modeling of complex spatio-temporal processes and agent-based simulation. He is focusing on hierarchical and deep generative models for complex physical and social processes, and approximate Bayesian computation. His email address is adway.cse@gmail.com. His website is https://sites.google.com/site/adwayresearch/.