Statistical Foundations of Virtual Democracy

Anson Kahng [1] **Min Kyung Lee** [1] **Ritesh Noothigattu** [1] **Ariel D. Procaccia** [1] **Alexandros Psomas** [1]

Abstract

Virtual democracy is an approach to automating decisions, by learning models of the preferences of individual people, and, at runtime, aggregating the *predicted* preferences of those people on the dilemma at hand. One of the key questions is which aggregation method — or *voting rule* — to use; we offer a novel statistical viewpoint that provides guidance. Specifically, we seek voting rules that are *robust* to prediction errors, in that their output on people's true preferences is likely to coincide with their output on noisy estimates thereof. We prove that the classic Borda count rule is robust in this sense, whereas any voting rule belonging to the wide family of pairwise-majority consistent rules is not. Our empirical results further support, and more precisely measure, the robustness of Borda count.

1. Introduction

One of the most basic ideas underlying democracy is that complicated decisions can be made by asking a group of people to vote on the alternatives at hand. As a decision-making framework, this paradigm is versatile, because people can express a sensible opinion about a wide range of issues. One of its seemingly inherent shortcomings, though, is that voters must take the time to cast a vote — hopefully an informed one — every time a new dilemma arises.

But what if we could *predict* the preferences of voters — instead of explicitly asking them each time — and then aggregate those predicted preferences to arrive at a decision? This is exactly the idea behind the work of Noothigattu et al. (2018), who are motivated by the challenge of automating *ethical* decisions. Specifically, their approach consists of three[1] steps: first, collect preferences from voters on exam-

ple dilemmas; second, learn models of their preferences, which generalize to any (previously unseen) dilemma; and third, at runtime, use those models to predict the voters' preferences on the current dilemma, and aggregate the predicted preferences to reach a decision. The idea is that we would ideally like to consult the voters on each decision, but in order to automate those decisions we instead use the models that we have learned as a proxy for the flesh and blood voters. In other words, the models serve as virtual voters, which is why we refer to this paradigm as *virtual democracy*.

Since 2017, we have been building on this approach in a collaboration with a Pittsburgh-based non-profit, 412 Food Rescue, that provides on-demand food donation distribution services. The goal is to design and deploy an algorithm that would automatically make the decisions they most frequently face: given an incoming food donation, which recipient organization (such as a housing authority or food pantry) should receive it? The voters in our implementation are stakeholders: donors, recipients, volunteers (who pick up the food from the donor and deliver it to the recipient), and employees. We have collected roughly 100 pairwise comparisons from each voter, where in each comparison, the voter is provided information about the type of donation, as well as seven relevant features of the two alternatives that are being compared, e.g., the distance between donor and recipient, and when the recipient last received a donation. Using this data, we have learned a model of the preferences of each voter, which allows us to predict the voter's preference *ranking* over hundreds of recipients. And given a predicted ranking for each voter, we map them into a ranking over the alternatives by applying a *voting rule*.

While this implementation sounds simple enough, the choice of voting rule can have a major impact on the efficacy of the system. In fact, the question of which voting rule to employ is one of the central questions in computational social choice (Brandt et al., 2016), and in social choice theory more broadly. A long tradition of impossibility results establishes that there are no perfect voting rules (Arrow, 1951), so the answer, such as it is, is often context-dependent.

The central premise of this paper is that, in the context of virtual democracy, certain statistical considerations should guide the choice of voting rule. Indeed, the voting rule

---

[1]Technically four, see Section 1.2.

inherently operates on noisy predictions of the voters' true preferences, yet one might hope that it would still output the same ranking as it would in the 'real' election based on the voters' true preferences (after all, this is the ideal that virtual democracy is trying to approximate). Our research question, therefore, is

> ... which voting rules have the property that their output on the true preferences is likely to coincide with their output on noisy estimates thereof?

1.1. Our Approach and Results

Our technical approach relies on the observation that the classic Mallows (1957) model is an unusually good fit with our problem. Typically the Mallows model describes situations where there is a true ranking of the alternatives σ^*. The probability that voter i would be associated with a given ranking σ_i decreases exponentially with the number of pairs of alternatives on which σ_i and σ^* disagree (formally known as the *Kendall tau* distance). The model is parameterized by a parameter $\phi \in (0, 1]$, which is directly related to the probability that σ_i agrees with σ^* on any particular pair of alternatives. This model is very well studied (see Section 1.2), but, even in situations where there is a ground-truth ranking, the Mallows model may not be an accurate representation of reality (Mao et al., 2013). This observation has motivated a body of work on generalized (Caragiannis et al., 2016; 2014) and adversarial (Procaccia et al., 2016; Benade et al., 2017) noise models.

In our setting each voter has a (possibly different) true ranking σ_i^*, and the voter's predicted ranking σ_i is drawn from a Mallows distribution around σ_i^*. Crucially, since the learning algorithm is, in fact, trying to predict pairwise comparisons (which make up the training set), the accuracy of the predictor can be directly mapped to the Mallows parameter ϕ. In other words, instead of making the classic assumption that voters may fail to identify the ordering of some pairs of alternatives with some probability, we are essentially observing that the machine learning algorithm fails to accurately predict some of the pairwise comparisons, and mapping that to a separate Mallows model for each voter. To drive the point home, although the Mallows model is widely believed to be a tenuous fit with previously studied applications (as discussed earlier), it is intuitively the correct way of reasoning about the errors that arise when machine learning algorithms predict rankings based on pairwise comparisons. This insight is a key part of our conceptual contribution.

Our main positive result (Theorem 1) is that the classic Borda count rule is *robust* to random noise, that is, it satisfies the property stated earlier, in a precise sense. Specifically, we establish an upper bound on the probability that two alternatives are ranked differently when Borda count is applied to the true preferences and to their noisy estimates.

The bound depends on the parameters of the model, as well as on the difference between the scores of the two alternatives in the true profile. On a high level, the theorem implies that if one alternative is stronger than another by a moderate margin under the true profile, Borda count is highly unlikely to swap the two when given noisy preferences.

By contrast, we show that voting rules belonging to the wide family of *pairwise-majority consistent* rules are *not* robust (Theorem 2). We do this by constructing an instance where there are significant margins between alternatives, yet any voting rule belonging to this family is likely to flip a pair of alternatives.

Finally, we provide empirical results that further strengthen our case for the robustness of Borda count. Specifically, these results suggest that the probability of making a mistake on a pair of alternatives decreases very quickly with their average Borda score difference, independently of the distribution used to generate the underlying true preferences.

1.2. Related Work

A number of recent papers have explored the idea of automating ethical decisions via machine learning and social choice (Conitzer et al., 2017; Freedman et al., 2018; Noothigattu et al., 2018). As mentioned above, our work builds on the framework proposed by Noothigattu et al. (2018). However, it is important to clarify why the questions we explore here do not arise in their work. Since they deal with 1.3 million voters, and split-second decisions (what should a self-driving car do in an emergency?), they cannot afford to consult the individual voter models at runtime. Hence, they have added an additional *summarization* step, whereby the individual voter models are summarized as a single, concise model of societal preferences (with possibly significant loss to accuracy). The structure of the summary model is such that, for any given set of alternatives, almost all reasonable voting rules agree on the outcome (this is their main theoretical result), hence the choice of voting rule is a nonissue under that particular implementation. By contrast, our work is motivated by the food bank application of the virtual democracy framework, where the number of voters is small and speed is not of the essence, hence we predict the preferences of individual voters at runtime.

It is worth mentioning that another prominent approach to the allocation of food donations is based on (online) fair division (Aleksandrov et al., 2015). That said, it is important to emphasize that we study a general question about the foundations of the virtual democracy paradigm, that is, our work is not technically tied to any particular application.

Furthermore, the Mallows model underlies a large body of work in computational social choice (Conitzer & Sandholm, 2005; Conitzer et al., 2009; Elkind et al., 2010; Elkind &

Shah, 2014; Xia et al., 2010; Xia & Conitzer, 2011; Lu & Boutilier, 2011; Procaccia et al., 2012; Jiang et al., 2014; Azari Soufiani et al., 2012; 2013; 2014; Mao et al., 2013; Caragiannis et al., 2014; 2016; Xia, 2016). Our model is loosely related to that of Jiang et al. (2014), where individual rankings are derived from a single ground truth ranking via a Mallows model, and then a second Mallows model is applied to obtain a noisy version of each voter's ranking. Our technical question is completely different from theirs.

Finally, there is a large body of work in social choice on finding aggregation rules that satisfy axiomatic properties that formally capture notions of fairness or efficiency (Arrow, 1951; Tennenholtz & Zohar, 2016). However, many common axiomatic properties in social choice do not apply to standard applications of virtual democracy, including the autonomous vehicle domain of Noothigattu et al. (2018) and our setting of food rescue, although they may be relevant in other differently-constrained domains.

2. Preliminaries

We deal with a set of alternatives A such that $|A| = m$. Preferences over A are represented via a ranking $\sigma \in \mathcal{L}$, where $\mathcal{L} = \mathcal{L}(A)$ is the set of rankings (or permutations) over A. We denote by $\sigma(j)$ the alternative ranked in position j in σ, where position 1 is the highest, and m the lowest. We denote by $\sigma^{-1}(x)$ the position in which $x \in A$ is ranked. We use $x \succ_\sigma y$ to denote that x is preferred to y according to σ, i.e., that $\sigma^{-1}(x) < \sigma^{-1}(y)$.

The setting also includes a set of voters $N = \{1, \ldots, n\}$. Each voter $i \in N$ is associated with a ranking $\sigma_i \in \mathcal{L}$. The preferences of N are represented as a *preference profile* $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n) \in \mathcal{L}^n$.

Given a preference profile $\boldsymbol{\sigma} \in \mathcal{L}^n$, we say that $x \in A$ *beats* $y \in A$ *in a pairwise comparison* if a majority of voters prefer x to y, that is,

$$|\{i \in N : x \succ_{\sigma_i} y\}| > n/2.$$

The profile $\boldsymbol{\sigma}$ induces a *weighted pairwise majority graph* $\Gamma(\boldsymbol{\sigma})$, where we have a vertex for each alternative in A. For each $x \in A$ and $y \in A \setminus \{x\}$, there is an edge from x to y if x beats y in a pairwise comparison; the weight on this edge is

$$w_{(x,y)}(\boldsymbol{\sigma}) \triangleq |\{i \in N : x \succ_{\sigma_i} y\}| - |\{i \in N : y \succ_{\sigma_i} x\}|.$$

2.1. Voting Rules

A *voting rule* (formally known as a *social welfare function*) is a function $f : \mathcal{L}^n \to \mathcal{L}$, which receives a preference profile as input, and returns a 'consensus' ranking of the alternatives. We are especially interested in two families of voting rules.

- *Positional scoring rules*. Each such rule is defined by a score vector $(\alpha_1, \ldots, \alpha_m)$. Given a preference profile $\boldsymbol{\sigma}$, the score of alternative x is

$$\sum_{i=1}^n \alpha_{\sigma_i^{-1}(x)}.$$

In words, each voter who ranks x in position p gives α_p points to x. The positional scoring rule returns a ranking of the alternatives by non-increasing score, with ties broken arbitrarily.

Our main positive result pertains to the classic *Borda count* voting rule, which is the positional scoring rule defined by the score vector $(m-1, m-2, \ldots, 0)$. Denote the Borda count score of $x \in A$ in $\boldsymbol{\sigma} \in \mathcal{L}^n$ by

$$B(x, \boldsymbol{\sigma}) \triangleq \sum_{i=1}^n \left(m - \sigma_i^{-1}(x) \right).$$

- *Pairwise-majority consistent (PMC) rules* (Caragiannis et al., 2016): These rules satisfy a fairly weak requirement that extends the classic notion of Condorcet consistent social *choice* functions: Given a profile $\boldsymbol{\sigma}$, if the pairwise majority graph $\Gamma(\boldsymbol{\sigma}) = (A, E)$ is such that for all $x \in A, y \in A \setminus \{x\}$, either $(x, y) \in E$ or $(y, x) \in E$ (i.e., it is a *tournament*), and, moreover, Γ is acyclic, then $f(\boldsymbol{\sigma}) = \tau$ for the unique ranking τ induced by $\Gamma(\boldsymbol{\sigma})$. Caragiannis et al. (2016) give many examples of prominent voting rules that are PMC, including the Kemeny rule, the Slater rule, the ranked pairs method, Copeland's method, and Schulze's method.

2.2. The Mallows Model

Let the *Kendall tau* distance between two rankings $\sigma, \sigma' \in \mathcal{L}$ be

$$d_{\mathrm{KT}}(\sigma, \sigma') \triangleq |\{(x, y) \in A^2 : x \succ_\sigma y \land y \succ_{\sigma'} x\}|.$$

In words, it is the number of pairs of alternatives on which σ and σ' disagree. For example, if $\sigma = (a, b, c, d)$, and $\sigma' = (a, c, d, b)$, then $d_{\mathrm{KT}}(\sigma, \sigma') = 2$.

In the Mallows (1957) model, there is a *ground truth ranking* $\sigma^\star$, which induces a probability distribution over perceived rankings. Specifically, the probability of a ranking σ, given the ground truth ranking $\sigma^\star$, is given by

$$\Pr[\sigma \mid \sigma^\star] \triangleq \frac{\phi^{d_{\mathrm{KT}}(\sigma, \sigma^\star)}}{Z},$$

where $\phi \in (0, 1]$ is a parameter, and

$$Z \triangleq \sum_{\sigma' \in \mathcal{L}} \phi^{d_{\mathrm{KT}}(\sigma', \sigma^\star)}$$

is a normalization constant. Note that for $\phi = 1$ this is a uniform distribution, whereas the probability of $\sigma^\star$ goes to 1 as ϕ goes to 0. In the rest of the paper we assume that $\phi < 1$ for ease of exposition.

The *repeated insertion model* (Doignon et al., 2004) provides a convenient alternative way of reasoning about the Mallows model. In the former model, alternatives are sequentially inserted into a partial ranking, until all alternatives have been ranked. Specifically, after alternatives $\sigma^\star(1), \dots, \sigma^\star(\ell - 1)$ have been inserted, the alternative $\sigma^\star(\ell)$ is inserted into the first position with probability P_1^ℓ, into the second with probability P_2^ℓ, and so on until P_ℓ^ℓ. The following lemma connects the parameters of the random insertion model with the parameter ϕ of the Mallows model.

Lemma 1 (Doignon et al. 2004). *The Mallows model with parameter $\phi \in (0, 1)$ induces the same distribution over rankings as the random insertion model with parameters*

$$P_i^\ell = \phi^{\ell - i} \cdot \frac{1 - \phi}{1 - \phi^\ell}.$$

We also require a lemma that gives bounds on the probability that the position of an element x in a ranking sampled from the Mallows model with parameter ϕ is far from its position in the true ranking.

Lemma 2 (Braverman & Mossel 2009). *Let σ be sampled from a Mallows model with parameter ϕ and true ranking $\sigma^\star$. Then for all alternatives $x \in A$ and all $s \geq 0$,*

$$\Pr[\sigma^{-1}(x) \leq (\sigma^\star)^{-1}(x) - s] \leq \frac{\phi^s}{1 - \phi}$$

$$\Pr[\sigma^{-1}(x) \geq (\sigma^\star)^{-1}(x) + s] \leq \frac{\phi^s}{1 - \phi}.$$

3. From Predictions to Mallows

In the virtual democracy framework, we are faced at runtime with a dilemma that induces a set of alternatives A. For example, when a food bank receives a donation, the set of alternatives is the current set of recipient organizations, each associated with information specific to the current donation, such as the distance between the donor and the recipient. Each voter $i \in N$ has a ranking $\sigma_i^\star \in \mathcal{L}$ over the given set of alternatives; together these rankings comprise the true preference profile $\boldsymbol{\sigma}^\star$.

One of the novel components of this paper is the assumption that, for each voter $i \in N$, we obtain a *predicted* ranking σ_i drawn from a Mallows distribution with parameter ϕ and true ranking $\sigma_i^\star$. We emphasize that, in contrast to almost all work on the Mallows Model, in our setting each voter has her own true ranking.

Why is the Mallows Model a good choice here? Recall that we are building preference models using pairwise compar-

isons as training data. When validating a model, we therefore test its accuracy on pairwise comparisons. And the Mallows model itself, because it is defined via the Kendall tau distance, is essentially determined by pairwise comparisons. In fact, the Mallows model (with parameter ϕ and true ranking $\sigma^\star$) is equivalent to the following generative process: for each pair of alternatives x and y such that $x \succ_{\sigma^\star} y$, x is preferred to y with probability $1/(1 + \phi)$, and y is preferred to x with probability $\phi/(1 + \phi)$; if this preference relation corresponds to a ranking (i.e., it is transitive), return that ranking, otherwise restart.

In more detail, let β be the average probability that we predict a pairwise comparison correctly; in our food bank implementation, $\beta \approx 0.9$. Based on the preceding discussion, one might be tempted to set $\beta = 1/(1 + \phi)$, i.e., set β to be the probability of getting the relative ordering of two *adjacent* alternatives correctly. While this is not unreasonable (and would have been very convenient for us), for $\beta \approx 0.9$ it would lead to extremely high probability of correctly ranking alternatives that are, say, 30 positions apart in the ground truth ranking. In order to moderate this effect, we define another parameter $\kappa \in \{2, \dots, m\}$, and assume that our observed pairwise comparisons are between $\sigma_i^\star(1)$ (the top-ranked alternative in the true ranking of i) and $\sigma_i^\star(\kappa)$ (the alternative ranked in position κ). Formally, the parameters β and κ are such that, for the ranking σ_i sampled from a Mallows Model with ϕ and $\sigma_i^\star$,

$$\Pr[\sigma_i^\star(1) \succ_{\sigma_i} \sigma_i^\star(\kappa)] = \beta. \tag{1}$$

It is worth noting that the implicit assumption that we are observing comparisons between $\sigma_i^\star(1)$ and $\sigma_i^\star(\kappa)$ specifically is not meant to be realistic. Rather, the idea is that there is *some* appropriate value of κ such that the observed accuracy β can be related to the underlying Mallows model through Equation (1), and, if we can establish results that are *general* with respect to the choice of κ, they would carry over to the real world.

Moving from conceptual issues to novel technical results, we start with the following lemma, which expresses the probability on the right hand side of Equation (1) in terms of the Mallows parameter ϕ.

Lemma 3. *Let σ_i be sampled from a Mallows Model with parameter ϕ and true ranking $\sigma_i^\star$. Then*

$$\Pr[\sigma_i^\star(1) \succ_{\sigma_i} \sigma_i^\star(\kappa)] = \frac{\kappa}{1 - \phi^\kappa} - \frac{\kappa - 1}{1 - \phi^{\kappa - 1}}.$$

Equation (1) and Lemma 3 imply that

$$\beta = \frac{\kappa}{1 - \phi^\kappa} - \frac{\kappa - 1}{1 - \phi^{\kappa - 1}},$$

but for subsequent results we need to express ϕ in terms of β and κ, and it is unclear whether this can be done in closed

form. Nevertheless, we are are able to derive a bound that suffices for our purposes.

Lemma 4. *For β and κ defined as in Equation (1), it holds that*

$$\phi \leq \left(\frac{1-\beta}{\beta}\right)^{\frac{1}{2\kappa-1}}.$$

We relegate the proofs of both lemmas to the full version of the paper. Note that Lemma 3 can be proved via a theorem of Désir et al. (2018). Their theorem gives a closed form for the probability that an alternative x is ranked first out of a subset of alternatives S. This closed form is complex, and requires quite a bit of additional notation, so we instead derive the probability we are interested in, i.e., the probability that $\sigma_i^\star(\kappa)$ is ranked above $\sigma_i^\star(1)$, from scratch.

4. Robustness of Borda Count

In this section, we rigorously establish the robustness of Borda count to prediction error by showing that it satisfies a formal version of the desired property stated in Section 1. We do this by building on the machinery developed in Section 3, as well as additional lemmas that we will state and prove momentarily.

As we have already discussed, we do not have access to the Mallows parameter ϕ. Instead, we can measure β, the probability that we correctly predict a pairwise comparison of alternatives that are κ positions apart. On a very high level, the theorem bounds the probability that the noisy Borda ranking (based on the sampled profile) would disagree with the true Borda ranking (based on the true profile) on a given pair of alternatives.

Theorem 1. *For any $\beta > 1/2$ and $\epsilon > 0$ there exists a universal constant $T = T(\beta, \epsilon)$ such that for all $n, m, \kappa \in \mathbb{N}$ such that $n, m \geq 2$, for all $s \geq T\kappa \log \kappa$, for all $\boldsymbol{\sigma}^\star \in \mathcal{L}^n$, and for all $x, x' \in A$ such that $\frac{1}{n}B(x, \boldsymbol{\sigma}^\star) \geq \frac{1}{n}B(x', \boldsymbol{\sigma}^\star) + 2s$, it holds that*

$$\Pr\left[\frac{1}{n}B(x, \boldsymbol{\sigma}) > \frac{1}{n}B(x', \boldsymbol{\sigma})\right] \geq 1 - \epsilon^n,$$

where the probability is taken over the sampling of $\boldsymbol{\sigma}$.

Let us discuss the statement of the theorem. First, note that the probability of mistake, ϵ^n, converges to 0 exponentially fast as n grows, so the theorem immediately implies a "with high probability" statement. Moreover, one can easily derive such a statement with respect to all pairs of alternatives (whose Borda scores are sufficiently separated) *simultaneously*, using a direct application of the union bound. Second, it is intuitive that the separation in Borda scores has to depend on κ, but it is encouraging (and, to us, surprising) that this dependence is almost linear. In particular, *even if κ is almost linear in m, i.e., $\kappa \in o(m/\log m)$, the theorem*

implies that our noisy Borda ranking is highly unlikely to make mistakes on pairs of alternatives whose average score difference is linear in m.

Turning to the proof, we start by bounding the probability that the Borda count score $B(x, \boldsymbol{\sigma})$ of an alternative $x \in A$ in the observed profile $\boldsymbol{\sigma}$ is far from the Borda count score $B(x, \boldsymbol{\sigma}^\star)$ in the true profile $\boldsymbol{\sigma}^\star$. The proof of the following lemma adapts that of a lemma of (Braverman & Mossel, 2009), which deals with average rank (instead of average Borda count score), but in the case of a single true ranking, i.e., $\sigma_i^\star = \sigma_j^\star$, for all i, j.

Lemma 5. *For all alternatives $x \in A$, and all $s \geq 0$*

$$\Pr\left[\frac{1}{n}B(x, \boldsymbol{\sigma}) \leq \frac{1}{n}B(x, \boldsymbol{\sigma}^\star) - s\right]$$
$$\leq \left(\frac{2e(n+ns-1)}{n-1} \cdot \frac{\phi^s}{1-\phi}\right)^n,$$

$$\Pr\left[\frac{1}{n}B(x, \boldsymbol{\sigma}) \geq \frac{1}{n}B(x, \boldsymbol{\sigma}^\star) + s\right]$$
$$\leq \left(\frac{2e(n+ns-1)}{n-1} \cdot \frac{\phi^s}{1-\phi}\right)^n.$$

Proof. We prove the first inequality; the proof of the second is analogous. Given a subset of voters $S \subseteq N$ and a nonnegative vector $\mathbf{b} = (b_i)_{i \in S} \in \mathbb{N}^{|S|}$, let $\mathcal{E}_{S,\mathbf{b}}$ be the event that

$$B(x, \sigma_i) \leq B(x, \sigma_i^\star) - b_i$$

for all voters $i \in S$, where we abuse notation by using

$$B(x, \sigma_i) \triangleq m - \sigma_i^{-1}(x)$$

to denote the Borda count score of alternative x in the *ranking* σ_i. Lemma 2 implies that for all $s \geq 0$,

$$\Pr[B(x, \sigma_i) \leq B(x, \sigma_i^\star) - s] \leq \frac{\phi^s}{1-\phi}. \qquad (2)$$

Therefore,

$$\Pr[\mathcal{E}_{S,\mathbf{b}}] = \prod_{i \in S} \Pr[B(x, \sigma_i) \leq B(x, \sigma_i^\star) - b_i]$$
$$\leq \prod_{i \in S} \frac{\phi^{b_i}}{1-\phi} = \frac{\phi^{\sum_{i \in S} b_i}}{(1-\phi)^{|S|}},$$

where the inequality follows from Equation (2).

Let $\mathcal{E}$ be the event that $\frac{1}{n}B(x, \boldsymbol{\sigma}) \leq \frac{1}{n}B(x, \boldsymbol{\sigma}^\star) - s$. Notice that

$$\mathcal{E} \subset \bigcup_{S \subseteq N, \mathbf{b} \in \mathbb{N}^{|S|}: \sum_{i \in S} b_i = ns} \mathcal{E}_{S,\mathbf{b}},$$

as there must exist a subset of voters who contribute sufficiently to the difference in Borda scores. Moreover, for

a fixed S, the number of vectors $\mathbf{b} \in \mathbb{N}^{|S|}$ such that $\sum_{i \in S} b_i = ns$ is exactly $\binom{|S|+ns-1}{|S|-1}$. Therefore,

$$
\begin{aligned}
\Pr[\mathcal{E}] &\leq \sum_{S \subseteq N} \left| \left\{ \mathbf{b} \in \mathbb{N}^{|S|} : \sum_{i=1}^{n} b_i = ns \right\} \right| \cdot \frac{\phi^{ns}}{(1-\phi)^{|S|}} \\
&\leq 2^n \cdot \binom{n+ns-1}{n-1} \cdot \frac{\phi^{ns}}{(1-\phi)^n} \\
&\leq 2^n \cdot \left(\frac{e(n+ns-1)}{n-1} \right)^{n-1} \cdot \left(\frac{\phi^s}{1-\phi} \right)^n \\
&\leq \left(\frac{2e(n+ns-1)}{n-1} \cdot \frac{\phi^s}{1-\phi} \right)^n,
\end{aligned}
$$

where we used the fact that $\binom{n}{t} \leq \left(\frac{en}{t} \right)^t$. $\qquad\square$

Using Lemma 5 we can bound, given the Mallows parameter ϕ, the probability that two alternatives, whose Borda count scores in the true profile $\boldsymbol{\sigma}^\star$ are sufficiently far apart, are ranked by the Borda count voting rule in the correct order (in the sampled profile $\boldsymbol{\sigma}$).

Lemma 6. *Let $x, x' \in A$ such that $\frac{1}{n} B(x, \boldsymbol{\sigma}^\star) \geq \frac{1}{n} B(x', \boldsymbol{\sigma}^\star) + 2s$. Then*

$$
\begin{aligned}
\Pr &\left[\frac{1}{n} B(x, \boldsymbol{\sigma}) > \frac{1}{n} B(x', \boldsymbol{\sigma}) \right] \\
&\geq 1 - 2 \left(\frac{2e(n+ns-1)}{n-1} \cdot \frac{\phi^s}{1-\phi} \right)^n.
\end{aligned}
$$

Proof. Let $\mathcal{E}_1$ be the event that

$$
\frac{1}{n} B(x, \boldsymbol{\sigma}) \leq \frac{1}{n} B(x, \boldsymbol{\sigma}^\star) - s,
$$

and $\mathcal{E}_2$ be the event that

$$
\frac{1}{n} B(x', \boldsymbol{\sigma}) \geq \frac{1}{n} B(x', \boldsymbol{\sigma}^\star) + s.
$$

By Lemma 5 and a union bound we have that

$$
\Pr\left[\mathcal{E}_1 \cup \mathcal{E}_2 \right] \leq 2 \left(\frac{2e(n+ns-1)}{n-1} \cdot \frac{\phi^s}{1-\phi} \right)^n.
$$

Next, notice that every time the Borda count scores of x and x' in the sampled preference profile are in the wrong order (or tied), then at least one of $\mathcal{E}_1, \mathcal{E}_2$ occurred, i.e.,

$$
\Pr\left[\frac{1}{n} B(x, \boldsymbol{\sigma}) \leq \frac{1}{n} B(x', \boldsymbol{\sigma}) \right] \leq \Pr[\mathcal{E}_1 \cup \mathcal{E}_2].
$$

The lemma directly follows. $\qquad\square$

Recall that Lemma 4 gives an upper bound on ϕ as a function of β and κ. Combining with Lemma 6, we can bound the probability of getting the correct ranking as a function of β and κ, and prove our main result.

Proof of Theorem 1. By Lemma 6,

$$
\begin{aligned}
\Pr &\left[\frac{1}{n} B(x, \boldsymbol{\sigma}) > \frac{1}{n} B(x', \boldsymbol{\sigma}) \right] \\
&\geq 1 - 2 \left(\frac{2e(n+ns-1)}{n-1} \cdot \frac{\phi^s}{1-\phi} \right)^n \\
&\geq 1 - 2 \left(\frac{4en}{n-1} \cdot \frac{s\phi^s}{1-\phi} \right)^n \\
&\geq 1 - 2 \left(8e \cdot \frac{s\phi^s}{1-\phi} \right)^n.
\end{aligned}
$$

It suffices to give a bound on s such that

$$
\frac{s\phi^s}{1-\phi} \leq \frac{\epsilon}{16e}. \tag{3}
$$

By Lemma 4,

$$
\phi \leq \left(\frac{1-\beta}{\beta} \right)^{\frac{1}{2\kappa-1}}.
$$

Since $\beta > 1/2$, there is a universal constant $c > 1$ such that $\frac{1-\beta}{\beta} = \frac{1}{c}$. Therefore,

$$
\begin{aligned}
\frac{s\phi^s}{1-\phi} &\leq s \cdot \frac{\left(\frac{1-\beta}{\beta} \right)^{\frac{s}{2\kappa-1}}}{1 - \left(\frac{1-\beta}{\beta} \right)^{\frac{1}{2\kappa-1}}} = s \cdot \frac{c^{-\frac{s}{2\kappa-1}}}{1 - c^{-\frac{1}{2\kappa-1}}} \\
&= \frac{s}{c^{\frac{s}{2\kappa-1}} - c^{\frac{s-1}{2\kappa-1}}} = \frac{s}{c^{\frac{s-1}{2\kappa-1}} \left(c^{\frac{1}{2\kappa-1}} - 1 \right)} \\
&\leq \frac{s}{c^{\frac{s-1}{2\kappa-1}} \cdot \frac{c^{\frac{1}{2\kappa-1}}(c-1)}{c(2\kappa-1)}} \leq \frac{c}{c-1} \cdot \frac{s(2\kappa-1)}{c^{\frac{s}{2\kappa-1}}},
\end{aligned}
$$

where for the penultimate inequality we use the inequality

$$
rz(z^{1/r} - 1) > z^{1/r}(z - 1),
$$

which holds for all $z, r \geq 1$,[2] with $z = c$ and $r = 2\kappa - 1$. It is now easy to verify that there is a universal constant $T > 0$ such that if $s \geq T\kappa \log \kappa$ then Equation (3) holds. $\qquad\square$

---

[2]To see this, let

$$
\begin{aligned}
f(z, r) &\triangleq \frac{rz(z^{1/r} - 1) - z^{1/r}(z-1)}{z} \\
&= (r-1)z^{1/r} + z^{1/r-1} - r.
\end{aligned}
$$

Taking the partial derivative with respect to z, we have

$$
\frac{\partial}{\partial z} f(z, r) = \frac{(r-1)(z-1)z^{1/r-2}}{r},
$$

which is clearly non-negative for $z, r \geq 1$. Also, $f(1, r) = 0$. So, we have shown that $f(z, r) \geq 0$ for all $z, r \geq 1$, which implies the claim.

It is important to note that it should be possible to extend Theorem 1 to other positional scoring rules defined by a score vector $(\alpha_1, \ldots, \alpha_m)$ where $\alpha_j > \alpha_{j+1}$ for all $j = 1, \ldots, m-1$. However, Borda count is especially practical and easy to explain (see Section 7 for more on this), which is why we focus on it for our positive result.

5. Non-Robustness of PMC Rules

Theorem 1 shows that Borda count is robust against noisy perturbations of the preference profile. It is natural to ask whether 'many' voting rules satisfy a similar property. In this section we answer this question in the negative, by proving that any voting rule that belongs to the important family of PMC rules is *not* robust in a similar sense.

Specifically, recall that under a PMC rule, when the weighted pairwise majority graph is acyclic, the output ranking is the topological ordering of the pairwise majority graph. We show that there exist profiles in which the pairwise majority graph is acyclic and all edge weights are large, but, with high probability, the noisy profile also has an acyclic pairwise majority graph which induces a *different* ranking. This means that *any* PMC rule would return different rankings when applied to the true profile and the noisy profile.

Theorem 2. *For all $\delta > 0$, $\phi \in (0, 1)$, and $m \in \mathbb{N}$ such that $m \geq 3$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, there exists a profile $\boldsymbol{\sigma}^\star \in \mathcal{L}^n$ such that $\Gamma(\boldsymbol{\sigma}^\star)$ is acyclic and all edges have weight $\Omega(n)$, but with probability at least $1 - \delta$ $\Gamma(\boldsymbol{\sigma})$ is acyclic and there is a pair of alternatives on which the unique rankings induced by $\Gamma(\boldsymbol{\sigma}^\star)$ and $\Gamma(\boldsymbol{\sigma})$ disagree, where the probability is taken over the sampling of $\boldsymbol{\sigma}$.*

It is instructive to contrast our positive result, Theorem 1, with this negative result. On a very high level, the former result asserts that "if Borda count says that the gaps between alternatives are significant, then the alternatives will not flip under Borda count," whereas the latter says "even if a PMC rule says that the gaps between alternatives are very significant, some alternatives are likely to flip under that rule." On a technical level, a subtle difference is that Theorem 1 is stated for β and κ, whereas Theorem 2 is stated directly for ϕ. This actually *strengthens* the negative result, because a constant β and $\kappa \in \omega(1)$ lead to $\phi = 1 - o(1)$, i.e., very noisy distributions — and still the positive result of Theorem 1 holds. By contrast, the negative result of Theorem 2 is true *even* when ϕ is constant, i.e., for settings that are not nearly as noisy. That said, the two results are not directly comparable, as Borda count and PMC rules deal with very different notions of score or weight. Nevertheless, the take-home message is that the notion of score that defines Borda count is inherently more robust to random perturbations of the preference profile.

The proof of Theorem 2 is rather technical, and appears in the full version of the paper. In a nutshell, we construct a preference profile $\boldsymbol{\sigma}^\star$ with αn voters whose preferences are $x^\star \succ x_1 \succ \cdots$, and $(1 - \alpha)n$ voters whose preferences are $x_1 \succ \cdots \succ x^\star$, for $\alpha > 1/2$. This profile induces a ranking where $x^\star$ is first and x_1 is second. However, it can be seen that, in the sampled profile $\boldsymbol{\sigma}$, many voters from the first group would flip $x^\star$ and x_1, leading to a majority who prefer x_1 to $x^\star$. Furthermore, we prove the nontrivial claim that $\Gamma(\boldsymbol{\sigma})$ is likely to be acyclic ('nontrivial' because it is unclear there would not be a cycle involving $x^\star$), which completes the argument.

6. Empirical Results

In Section 4 we have established that Borda count is robust to prediction error. However, our positive theoretical result, Theorem 1, only provides asymptotic guarantees. In this section, we evaluate the performance of Borda count on profiles of size that is more representative of real-world instances. For our evaluation metric, we consider the probability of the rule flipping alternatives when aggregating noisy rankings against their difference in Borda score in the underlying true profile.

All of our code is open-source and can be found at https://github.com/akahng/VirtualDemocracy-ICML2019.

6.1. Methodology

Given n voters, m alternatives, a Mallows parameter $\phi \in (0, 1)$, and a probability $p \in [0, 1]$, we generate a true profile $\boldsymbol{\sigma}^\star = (\sigma_1^\star, \ldots, \sigma_n^\star)$ from a mixture of Mallows models. Specifically, each ranking is drawn with probability p from a Mallows model with base ranking $x_1 \succ x_2 \succ \cdots \succ x_m$ and parameter ϕ, and with probability $1 - p$ from a Mallows model with base ranking $x_m \succ x_{m-1} \succ \cdots \succ x_1$ and parameter ϕ.

We then repeatedly generate noisy profiles $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ where each σ_i is generated by a Mallows model centered at $\sigma_i^\star$ with parameter ϕ. For every pair of alternatives (x_i, x_j) such that $B(x_i, \boldsymbol{\sigma}^\star) > B(x_j, \boldsymbol{\sigma}^\star)$ — that is, x_i beat x_j when Borda count was applied to the true profile — we calculate the percentage of noisy profiles that flipped the order of x_i and x_j, i.e., those where $B(x_j, \boldsymbol{\sigma}) > B(x_i, \boldsymbol{\sigma})$. Based on the true difference in Borda scores $B(x_i, \boldsymbol{\sigma}^\star) - B(x_j, \boldsymbol{\sigma}^\star)$, we place this data point in the appropriate bucket, where the width of each bucket corresponds to an average Borda score difference of 1. This way we can relate the Borda score difference to the probability of making a pairwise prediction error. Note that starting from a mixture of 'opposite' ranking models allows us to vary the distribution over score differences in $\boldsymbol{\sigma}^\star$ by varying p.
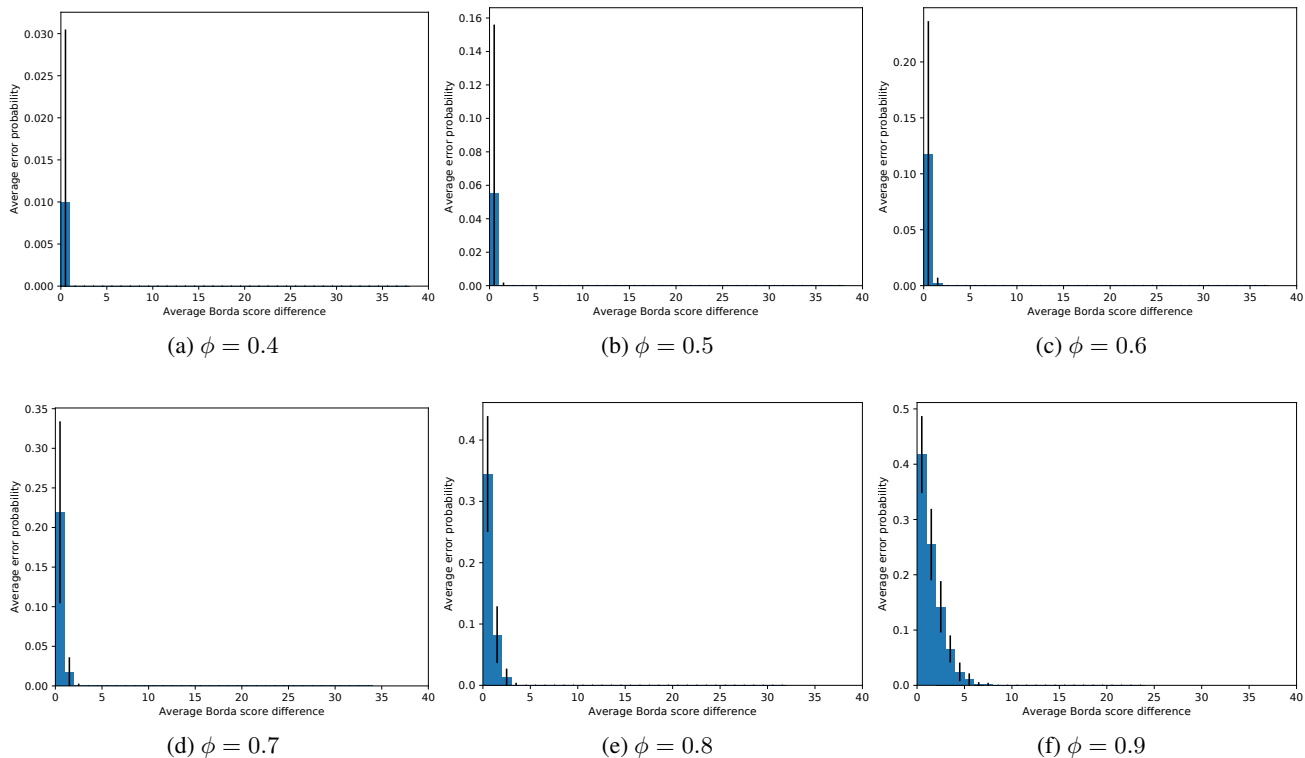
Figure 1. $p = 1$ mixture of Mallows, $n = 100$ voters, $m = 40$ alternatives

6.2. Results

Throughout our experiments, we let $n = 100$, $m = 40$, $\phi \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, and $p \in \{1, 0.7, 0.5\}$. Our results for $p = 1$, shown in Figure 1, plot the average probability of flipping the order of alternatives as a function of the difference in average Borda scores of the alternatives, where comparisons are bucketed by the difference in average Borda score. For $\phi \in \{0.1, 0.2, 0.3\}$, the observed probability of flipping any two alternatives, regardless of average Borda score difference, is 0; i.e., there are no mistakes.

At a high level, error rate decreases with true average Borda score distance in all experiments. Note that the maximum observed error rate increases with the Mallows parameter ϕ, which is intuitive because higher values of ϕ imply noisier (more uniformly random) rankings, so the probability of swapping alternatives should increase. However, for all values of ϕ and under all methods of generating profiles, the probability of making errors quickly decreases with average Borda score difference in the true profile.

Similar plots for $p = 0.7$ and $p = 0.5$ are included in the full version of the paper; these plots support the observation that the probability of making a mistake depends on the average Borda score difference, and not on the particular methods used to sample the underlying true profile.

7. Discussion

Our theoretical and empirical results identify Borda count as an especially attractive voting rule for virtual democracy, from a statistical viewpoint. However, Borda count is also compelling in terms of usability and explainability.

In more detail, in our implemented donor-recipient matching system, clicking on a recommended alternative displays an explanation for why it was ranked highly by Borda count, which consists of two components. First, we show the alternative's average position in the predicted preferences of each of the four stakeholder groups. Note that this information determines the Borda score of the alternative, given the weight of each stakeholder group.[3] Second — this is the more novel component — we show specific features in which the recommended alternative stands out. This is interesting because classic social choice theory does not have features for alternatives, and we are able to give this type of explanation precisely because our alternatives are represented as vectors of features (which is crucial for the application of learning-to-rank algorithms).

Based on the results presented in this paper, as well as these additional insights, we use Borda count in our implemented virtual-democracy-based system.

---

[3]These weights were decided by the stakeholders themselves.

References

Aleksandrov, M., Aziz, H., Gaspers, S., and Walsh, T. Online fair division: Analysing a food bank problem. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2540–2546, 2015.

Arrow, K. *Social Choice and Individual Values*. Wiley, 1951.

Azari Soufiani, H., Parkes, D. C., and Xia, L. Random utility theory for social choice. In *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS)*, pp. 126–134, 2012.

Azari Soufiani, H., Parkes, D. C., and Xia, L. Preference elicitation for general random utility models. In *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 596–605, 2013.

Azari Soufiani, H., Parkes, D. C., and Xia, L. Computing parametric ranking models via rank-breaking. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pp. 360–368, 2014.

Benade, G., Kahng, A., and Procaccia, A. D. Making right decisions based on wrong opinions. In *Proceedings of the 18th ACM Conference on Economics and Computation (EC)*, pp. 267–284, 2017.

Brandt, F., Conitzer, V., Endriss, U., Lang, J., and Procaccia, A. D. (eds.). *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

Braverman, M. and Mossel, E. Sorting from noisy information. arXiv preprint arXiv:0910.1191, 2009.

Caragiannis, I., Procaccia, A. D., and Shah, N. Modal ranking: A uniquely robust voting rule. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 616–622, 2014.

Caragiannis, I., Procaccia, A. D., and Shah, N. When do noisy votes reveal the truth? *ACM Transactions on Economics and Computation*, 4(3): article 15, 2016.

Conitzer, V. and Sandholm, T. Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 145–152, 2005.

Conitzer, V., Rognlie, M., and Xia, L. Preference functions that score rankings and maximum likelihood estimation. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 109–115, 2009.

Conitzer, V., Sinnott-Armstrong, W., Schaich Borg, J., Deng, Y., and Kramer, M. Moral decision making frameworks for artificial intelligence. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pp. 4831–4835, 2017.

Désir, A., Goyal, V., Jagabathula, S., and Segev, D. Mallows-smoothed distribution over rankings approach for modeling choice. SSRN preprint, 2018.

Doignon, J.-P., Pekeč, A., and Regenwetter, M. The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika*, 69(1):33–54, 2004.

Elkind, E. and Shah, N. Electing the most probable without eliminating the irrational: Voting over intransitive domains. In *Proceedings of the 30th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 182–191, 2014.

Elkind, E., Faliszewski, P., and Slinko, A. Good rationalizations of voting rules. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 774–779, 2010.

Freedman, R., Schaich Borg, J., Sinnott-Armstrong, W., Dickerson, J. P., and Conitzer, V. Adapting a kidney exchange algorithm to align with human values. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1636–1643, 2018.

Jiang, A. X., Marcolino, L. S., Procaccia, A. D., Sandholm, T., Shah, N., and Tambe, M. Diverse randomized agents vote to win. In *Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS)*, pp. 2573–2581, 2014.

Lu, T. and Boutilier, C. Robust approximation and incremental elicitation in voting protocols. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 287–293, 2011.

Mallows, C. L. Non-null ranking models. *Biometrika*, 44: 114–130, 1957.

Mao, A., Procaccia, A. D., and Chen, Y. Better human computation through principled voting. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1142–1148, 2013.

Noothigattu, R., Gaikwad, S. S., Awad, E., Dsouza, S., Rahwan, I., Ravikumar, P., and Procaccia, A. D. A voting-based system for ethical decision making. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1587–1594, 2018.

Procaccia, A. D., Reddi, S. J., and Shah, N. A maximum likelihood approach for selecting sets of alternatives. In *Proceedings of the 28th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 695–704, 2012.

Procaccia, A. D., Shah, N., and Zick, Y. Voting rules as error-correcting codes. *Artificial Intelligence*, 231:1–16, 2016.

Tennenholtz, M. and Zohar, A. The axiomatic approach and the Internet. In Brandt, F., Conitzer, V., Endress, U., Lang, J., and Procaccia, A. D. (eds.), *Handbook of Computational Social Choice*, chapter 18. Cambridge University Press, 2016.

Xia, L. Bayesian estimators as voting rules. In *Proceedings of the 32nd Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 2016.

Xia, L. and Conitzer, V. A maximum likelihood approach towards aggregating partial orders. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 446–451, 2011.

Xia, L., Conitzer, V., and Lang, J. Aggregating preferences in multi-issue domains by using maximum likelihood estimators. In *Proceedings of the 9th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 399–408, 2010.