The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)

Can We Predict the Election Outcome from Sampled Votes?

Evi Micha, Nisarg Shah
University of Toronto
{emicha, nisarg}@cs.toronto.edu

Abstract

In the standard model of voting, it is assumed that a voting rule observes the ranked preferences of each individual over a set of alternatives and makes a collective decision. In practice, however, not every individual votes. Is it possible to make a good collective decision for a group given the preferences of only a few of its members?

We propose a framework in which we are given the ranked preferences of k out of n individuals sampled from a distribution, and the goal is to predict what a given voting rule would output if applied on the underlying preferences of all n individuals. We focus on the family of positional scoring rules, derive a strong negative result when the underlying preferences can be arbitrary, and discover interesting phenomena when they are generated from a known distribution.

1 Introduction

The aim of voting is to make a good collective decision for a group of individuals based on the preferences of its members over a set of alternatives. In the vast literature published on voting since the work of Condorcet [1785], numerous voting rules have been proposed which intuitively provide different notions of what makes an outcome best for the group. Additionally, the literature offers several frameworks that define which collective decisions are good, and allow evaluating as well as systematically designing voting rules; examples include the axiomatic approach (Arrow 1951; Moulin 1988), distance rationalizability (Meskanen and Nurmi 2008; Elkind, Faliszewski, and Slinko 2015), noisy voting (Young 1988; Caragiannis, Procaccia, and Shah 2016), and implicit utilitarian voting (Procaccia and Rosenschein 2006; Boutilier et al. 2015). However, most of this literature assumes that the voting rule in question is able to observe the preferences of every individual in the group.

In most real-world applications, only a fraction of the members actually participate in voting. The goal of the decision-making system is then to *predict* what the right collective decision is for the whole group (including individuals whose preferences are not observed) given the preferences of some of its members. In this work, we assume that there are n individuals (a.k.a. voters), and we are given a voting rule f_1 which can make the desired collective decision given the preferences of all n voters. Instead, we observe the preferences of only k out of n voters, and our goal is to design a voting rule f_2 which, when applied on the k observed preferences, predicts the outcome of f_1 on all n preferences.

This framework has two immediate motivations. As described so far, we could think of f_1 as the idealized voting rule we would like to implement if every individual votes, and f_2 as the voting rule we should really implement if only a fraction of individuals are expected to vote. Alternatively, we can imagine a setting where voting rule f_1 is implemented for an upcoming election, and we would like to conduct a poll to observe a subset of preferences and use f_2 to predict the outcome of the upcoming election. Predicting outcomes of political elections using surveys has been extensively studied (Tumasjan et al. 2010; Rothschild and Wolfers 2011; Walther 2015; Graefe 2014; Lewis-Beck and Tien 1999; Lewis-Beck 2005; Dey and Bhattacharyya 2015).

One question still lingers: *Which k out of n voters would participate?* If they are adversarially chosen, it is not difficult to see that predicting the election outcome is impossible. For instance, if 49 out of 100 voters report that they prefer alternative a over alternative b, it is impossible to know if this is in fact the majority opinion. In the worst case, each of the remaining 51 voters might prefer b over a. However, we can argue that this is an unlikely scenario.

In this work, we assume that the k voters are sampled from a distribution.[1] This raises a number of questions: Which voting rule f_2, given the k observed votes, would best predict the outcome of f_1 on all n votes? When is it optimal to simply apply the same voting rule on the observed votes (i.e. $f_2 = f_1$)? Which voting rules f_1 can be predicted well? If we have stochastic information about the unobserved votes, how do we incorporate it into the prediction rule f_2? Such questions are the focus of this work.

[1] Our main results hold for all possible distributions, but we provide additional results for the special case where the k voters are sampled uniformly at random from all n voters.

Our Results

We study a setting with m alternatives and n voters who have ranked preferences over the alternatives. A voting rule takes as input a set of rankings and returns a societal ranking. In case of ties, it may return a set of societal rankings.

We use $\vec{\sigma}_n$ to denote the profile of all n ranked votes, and $\vec{\pi}_k$ to denote the sample of k votes. For simplicity, assume for now that the k votes are selected uniformly at random. We focus on the family of positional scoring rules, which includes popular voting rules such as plurality, Borda count, harmonic rule, k-approval, and veto. Given positional scoring rules f_1 and f_2, we are interested in the probability that $f_2(\vec{\pi}_k) \subseteq f_1(\vec{\sigma}_n)$, i.e., that f_2 predicts the outcome of f_1 on $\vec{\sigma}_n$ by producing a refinement of this outcome.

This probability depends on the underlying profile $\vec{\sigma}_n$. In Section 3, we consider the prediction accuracy in the worst case over $\vec{\sigma}_n$. We show that for a positional scoring rule f_1 other than plurality and veto, no positional scoring rule f_2 can predict its outcome with a positive probability in the worst case, when n and k have different parity (Theorems 1 and 2). This holds for any distribution from which the k observed votes are sampled. When f_1 is plurality or veto, and the distribution of samples is uniform, we show that f_1 is the optimal predictor of itself among all positional scoring rules (Theorem 3), but its prediction accuracy is still small when the number of alternatives is large (Theorems 4 and 5). In summary, it is impossible to predict the outcome of any positional scoring rule with a reasonable accuracy when no additional information is known about $\vec{\sigma}_n$.[2]

In Section 4, we consider the expected prediction accuracy when $\vec{\sigma}_n$ is drawn from a known prior. Using the simplest case of two alternatives, where our goal is to predict the majority rule f_1, we show that the knowledge of prior can have little to significant effect on the optimal prediction rule f_2, depending on how large k is compared to n and how concentrated the prior is (Theorem 6).

Our experiments in Section 5 show that when $\vec{\sigma}_n$ is drawn from a concentrated prior (the Mallows model with $\varphi = 1/3$), most voting rules can be predicted with at least 98% accuracy given only 3% of the votes. However, when $\vec{\sigma}_n$ is drawn from the uniform prior, most voting rules cannot be predicted with accuracy more than 4% given only 3% of the votes, although the accuracy increases with more observed votes. We also curiously discover that, in certain settings, the harmonic rule predicts other voting rules better than they predict themselves.

Related Work

Most closely related to ours is the work of Dey and Bhattacharyya [2015]. They consider voting rules which output a single alternative instead of a ranking, and study the problem of predicting the output of a given voting rule on an unknown election by sampling votes. Their work differs from ours in two key aspects. First, they sample with replacement and allow the prediction rule to determine how many votes to sample. In contrast, our sampling is without replacement (which becomes dramatically different when k is comparable to n) and the sampled votes are given. But more importantly, they assume that the underlying election has a margin of victory that is at least a constant fraction of n, that is, the underlying election is such that changing a constant fraction of the votes cannot change the outcome of the voting rule. We do not make this assumption. In fact, our negative results are derived precisely by considering elections that are borderline. In that sense, our results complement the results of Dey and Bhattacharyya [2015] by showing that their positive results are replaced by strong negative results when their margin of victory assumption is dropped.

Our results also have a surprising connection to the work of Borodin et al. [2019]. They consider an implicit utilitarian voting framework, in which voters and alternatives are embedded in an underlying metric space, each voter ranks the alternatives, and the social cost of an alternative is measured by its total distance from the voters. One of their results (informally) shows that given a voting rule and an *arbitrary* set of k votes, where $k = \Theta(n)$, it is possible to produce an alternative that is almost as good as the alternative that would be produced by the voting rule with all n votes. That is, in their framework, it is possible to do almost as well as the idealized voting rule even if the sampled votes are adversarially. This is fundamentally impossible in our setting.

A bit further afield, there is also work on predicting election outcomes under different types of uncertainty such as partial preferences (Doucette, Larson, and Cohen 2014; Baumeister et al. 2012; Lang et al. 2012; Aziz et al. 2015), uncertainty about which voters or candidates would participate in the election (even if all preferences are known upfront) (Wojtas and Faliszewski 2012), or distributional uncertainty about *each* voter's preferences (Hazon et al. 2012).

2 Preliminaries

For $k \in \mathbb{N}$, let $[k] = \{1, \ldots, k\}$. We consider a set $A = \{a_1, \ldots, a_m\}$ of m *alternatives* and a set $N = \{1, \ldots, n\}$ of n *voters*. We denote by $\mathcal{L}(A)$ the set of all rankings over A. We use $a \succ_\sigma b$ to denote that alternative a is preferred to alternative b under ranking σ. Each voter i has a preference ranking (*vote*), denoted $\sigma_i \in \mathcal{L}(A)$. The *(preference) profile* $\vec{\sigma}_n = (\sigma_1, \ldots, \sigma_n)$ is the collection of all n votes.

A voting rule (technically, a social welfare function) is a function $f : \mathcal{L}(A)^n \to 2^{\mathcal{L}(A)}$, which takes as input a profile and outputs a set of tied rankings. In this work, we focus on the family of *positional scoring rules*, denoted $\mathcal{F}$. A positional scoring rule $f_{\vec{s}}$ is characterized by a scoring vector $\vec{s} = (s_1, \ldots, s_m) \in \mathbb{R}^m$, where $s_t \geq s_{t+1}$ for each $t \in [m-1]$ and $s_1 > s_m$. Given a profile $\vec{\sigma}_n$, $f_{\vec{s}}$ assigns s_t points to the t^{th} alternative in voter i's vote, for each $i \in N$ and $t \in [m]$. Let $\mathrm{sc}_{\vec{s}}(a, \vec{\sigma}_n) = \sum_{i=1}^{n} s_{\sigma_i(a)}$ denote the total score of $a \in A$, where $\sigma_i(a)$ is the rank of a in voter i's vote. Then, $f_{\vec{s}}$ returns the set of rankings where the alternatives are sorted in a non-ascending order of their scores.

We partition $\mathcal{F}$ into three subfamilies, $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$. Family $\mathcal{F}_1$ consists of all rules $f_{\vec{s}}$ for which $s_2 > s_{m-1}$. This

[2] This is partly due to the fact that we want to predict the entire ranking of alternatives returned by the rule. In Section 6, we consider the weaker requirement of predicting only the winning alternative, and present a mix of positive and negative results.

includes the well known Borda rule ($\vec{s} = (m, m-1, \ldots, 1)$) and harmonic rule ($\vec{s} = (1, 1/2, \ldots, 1/m)$). The remaining rules $f_{\vec{s}}$ satisfy $s_2 = s_3 = \ldots = s_{m-1}$. Among these, family $\mathcal{F}_2$ consists of rules for which $s_1 > s_2 = \ldots = s_{m-1} > s_m$, while family $\mathcal{F}_3$ contains the two remaining rules: $s_1 > s_2 = \ldots = s_m$ is equivalent to plurality, and $s_1 = \ldots = s_{m-1} > s_m$ is equivalent to veto.

The *Mallows model* is a distribution over $\mathcal{L}(A)$, parametrized by a *central ranking* $\sigma^* \in \mathcal{L}(A)$ and a *noise parameter* $\varphi \in [0, 1]$. To obtain a sample ranking from this distribution, one generates an independent comparison between each pair of alternatives which matches with σ^* with probability p (where $p \geq 1/2$ and $\varphi = (1-p)/p$), and restarts if the comparisons violate transitivity. When $\varphi = 0$ (i.e. $p = 1$), the distribution puts all the probability mass on σ^*. When $\varphi = 1$ (i.e. $p = 1/2$), we obtain the uniform distribution, also known as impartial culture.

3 Worst-Case Predictability

Given a positional scoring rule $f_1 \in \mathcal{F}$, our goal in this paper is to study how accurately one can predict its outcome on a profile $\vec{\sigma}_n$ given a sample of k votes from the profile, where $k \leq n$. Specifically, let $\mathcal{S}_k(\vec{\sigma}_n)$ denote the set of all subsets of $\vec{\sigma}_n$ of size k, and let $U_k(\vec{\sigma}_n)$ be the uniform distribution over $\mathcal{S}_k(\vec{\sigma}_n)$. Define the accuracy of predicting f_1 using a positional scoring rule $f_2 \in \mathcal{F}$ on profile $\vec{\sigma}_n$ as $\mathrm{acc}(f_1, f_2, \vec{\sigma}_n) = \Pr_{\vec{\pi}_k \sim U_k(\vec{\sigma}_n)}[f_2(\vec{\pi}_k) \subseteq f_1(\vec{\sigma}_n)]$.

Note that $f_2(\vec{\pi}_k) \subseteq f_1(\vec{\sigma}_n)$ allows f_2 to break some of the ties produced by f_1 on $\vec{\sigma}_n$. This makes our negative results stronger than if we had required $f_2(\vec{\pi}_k) = f_1(\vec{\sigma}_n)$. Similarly, although we defined accuracy for $\vec{\pi}_k$ sampled from the uniform distribution $U_k(\vec{\sigma}_n)$, our main negative results (Theorems 1 and 2) hold for all distributions since they establish zero accuracy.

We then define the *worst-case accuracy* of predicting f_1 using f_2 as $\mathrm{acc}(f_1, f_2) = \min_{\vec{\sigma}_n} \mathrm{acc}(f_1, f_2, \vec{\sigma}_n)$. Taking this one step further, we define the *worst-case predictability* of f_1 as $\mathrm{acc}(f_1) = \sup_{f_2 \in \mathcal{F}} \mathrm{acc}(f_1, f_2)$, which is the worst-case accuracy of predicting f_1 using the best positional scoring rule f_2. Note that these quantities depend on n, m, and k, which are fixed in our framework. Motivated by political applications, we are interested in cases where n is large but k and m are relatively smaller.

Predicting a Rule from $\mathcal{F}_1$ or $\mathcal{F}_2$

We begin by establishing a strong negative result: every positional scoring rule in $\mathcal{F}_1$ and $\mathcal{F}_2$ has zero worst-case predictability. That is, such a rule cannot be predicted by any positional scoring rule with positive worst-case accuracy. The strength of the result lies in two observations. First, as we argued above, zero predictability implies that the outcome of the rule cannot be predicted given *any* subset of k votes; thus, the negative result holds for any distribution from which the k observed votes are drawn. Second, while the impossibility of prediction may be intuitive for small values of k, the result holds even when $k = n-1$, i.e., when all but one of the votes are observed.

Theorem 1. *Let $n \geq 2$, $m \geq 7$, and $k \in [n-1]$ such that n and k have different parity. Then, for any positional scoring rules $f_1 \in \mathcal{F}_1$ and $f_2 \in \mathcal{F}$, we have $\mathrm{acc}(f_1, f_2) = 0$.*

Proof. Fix positional scoring rules $f_1 \in \mathcal{F}_1$ and $f_2 \in \mathcal{F}$. Let $\vec{r}$ and $\vec{s}$ denote their scoring vectors, respectively. Because $f_1 \in \mathcal{F}_1$, we have $r_2 > r_{m-1}$. We consider cases of even and odd k, and for each case, construct a profile on which f_2 predicts f_1 with zero accuracy.

Odd n, even k: We start with the case where n is odd and k is even. Consider the following profile $\vec{\sigma}_n$. Each row represents a ranking where alternatives are listed from left to right in the most preferred to least preferred order. The first two rankings appear $(n-1)/2$ times each, and the third ranking appears once. Alternatives not shown appear in an arbitrary order in the middle.

$\frac{n-1}{2}$ votes	$a_1 \succ a_2 \succ a_3 \succ \ldots \succ a_{m-2} \succ a_{m-1} \succ a_m$
$\frac{n-1}{2}$ votes	$a_m \succ a_{m-1} \succ a_{m-2} \succ \ldots \succ a_3 \succ a_2 \succ a_1$
1 vote	$a_1 \succ a_{m-1} \succ a_3 \succ \ldots \succ a_{m-2} \succ a_2 \succ a_m$

We denote with σ_1, σ_2 and σ_3 the ranking of the first, second and third rows, respectively. Because $f_1 \in \mathcal{F}_1$, it holds that for every $\sigma^* \in f_1(\vec{\sigma}_n)$, $a_1 \succ_{\sigma^*} a_m$ and $a_{m-1} \succ_{\sigma^*} a_2$. We show that for every sample $\vec{\pi}_k \in \mathcal{S}_k(\vec{\sigma}_n)$, $f_2(\vec{\pi}_k) \not\subseteq f_1(\vec{\sigma}_n)$, i.e., for some $\hat{\sigma} \in f_2(\vec{\pi}_k)$ at least one of $a_1 \succ_{\hat{\sigma}} a_m$ and $a_{m-1} \succ_{\hat{\sigma}} a_2$ fails to hold. Suppose for contradiction that there exists $\vec{\pi}_k$ for which this does not happen.

Let x_1, x_2, and x_3 denote the number of times σ_1, σ_2, and σ_3 appear in $\vec{\pi}_k$, respectively. Note that $x_3 \in \{0, 1\}$. To have $a_1 \succ_{\hat{\sigma}} a_m$ for every $\hat{\sigma} \in f_2(\vec{\pi}_k)$, we need

$$x_1 \cdot s_1 + x_2 \cdot s_m + x_3 \cdot s_1 > x_1 \cdot s_m + x_2 \cdot s_1 + x_3 \cdot s_m$$
$$\Rightarrow (x_1 + x_3) \cdot (s_1 - s_m) > x_2 \cdot (s_1 - s_m).$$

Given that $s_1 > s_m$, this implies $x_1 + x_3 > x_2$.

On the other hand, to have $a_{m-1} \succ_{\hat{\sigma}} a_2$ for every $\hat{\sigma} \in f_2(\vec{\pi}_k)$, we need

$$\begin{aligned} & x_1 \cdot s_{m-1} + x_2 \cdot s_2 + x_3 \cdot s_2 \\ & \quad > x_1 \cdot s_2 + x_2 \cdot s_{m-1} + x_3 \cdot s_{m-1} \\ & \Rightarrow (x_2 + x_3) \cdot (s_2 - s_{m-1}) > x_1 \cdot (s_2 - s_{m-1}). \end{aligned}$$

Given $s_2 \geq s_{m-1}$, this implies $x_2 + x_3 > x_1$.

Given $x_1 + x_3 > x_2$ and $x_2 + x_3 > x_1$, we can derive $x_3 > 0$, i.e., $x_3 = 1$. But then, we have $x_1 + 1 > x_2$ and $x_2 + 1 > x_1$, which implies $x_1 = x_2$. In this case, $|\vec{\pi}_k| = x_1 + x_2 + 1$ is odd, which contradicts the fact that k is even.

Even n, odd k: We now consider the case of even n and odd k. We begin by establishing the following property of f_1. Recall that for the scoring vector $\vec{r}$ of $f_1 \in \mathcal{F}_1$, we have $r_2 > r_{m-1}$.

Lemma 1. *There exists a $t \in \{3, \ldots, m-3\}$ such that $r_2 - r_{m-1} > r_t - r_{t+1}$.*

Proof. First, suppose there exists a $p \in \{2, \ldots, m-2\}$ such that $r_2 = r_p > r_{p+1} = r_{m-1}$ (i.e., in going from

r_2 to r_{m-1}, the score drops only once). If $p \geq 4$, then we set $t = 3$. In this case, we have $r_t - r_{t+1} = 0 < r_2 - r_{m-1}$, as desired. If $p \leq 3$, then we set $t = 4$. Because $m \geq 7$, we have $t \leq m - 3$. Also, we again have $r_t - r_{t+1} = 0 < r_2 - r_{m-1}$.

Next, suppose there exist distinct $p, q \in \{2, \dots, m-2\}$ such that $r_p > r_{p+1}$ and $r_q > r_{q+1}$ (i.e., in going from r_2 to r_{m-1}, the score drops at least twice). Then, we can simply set $t = 3$. This ensures that $r_t - r_{t+1} < r_2 - r_{m-1}$ (if it were equal, then the score would drop only once in going from r_2 to r_{m-1}). This completes the proof. □

Let us fix $t \in \{3, \dots, m-3\}$ for which Lemma 1 holds. Consider the following profile $\vec{\sigma}_n$.

$\frac{n-2}{2}$ votes	$a_1 \succ a_2 \succ a_3 \succ a_4 \succ \dots \succ a_{m-2} \succ a_{m-1} \succ a_m$
$\frac{n-2}{2}$ votes	$a_m \succ a_{m-1} \succ a_{m-2} \succ \dots \succ a_4 \succ a_3 \succ a_2 \succ a_1$
1 vote	$a_1 \succ a_3 \succ \dots \succ a_2 \succ a_{m-1} \succ \dots \succ a_{m-2} \succ a_m$
1 vote	$a_3 \succ a_{m-1} \succ \dots \succ a_m \succ a_1 \succ \dots \succ a_2 \succ a_{m-2}$

We again denote with σ_1, σ_2, σ_3 and σ_4 the rankings in rows 1, 2, 3 and 4, respectively. In ranking σ_3, a_2 is at position t and a_{m-1} is at position $t+1$. In the ranking σ_4, a_m is at position t and a_1 is at position $t+1$. In each ranking, alternatives not shown appear in the unfilled positions arbitrarily.

First, we argue about the outcome of f_1 on this profile. From the Lemma 1, it is obvious that $r_1 - r_m \geq r_2 - r_{m-1} > r_t - r_{t+1}$. Using this, it is easy to see that for every $\sigma^* \in f_1(\vec{\sigma}_n)$, $a_1 \succ_{\sigma^*} a_m$ and $a_{m-1} \succ_{\sigma^*} a_2$. We now argue that for every sample $\vec{\pi}_k \in \mathcal{S}_k(\vec{\sigma}_n)$, there exists $\hat{\sigma} \in f_2(\vec{\pi}_k)$ which violates at least one of $a_1 \succ_{\hat{\sigma}} a_m$ and $a_{m-1} \succ_{\hat{\sigma}} a_2$. Suppose for contradiction that there exists a sample $\vec{\pi}_k$ for which this does not happen.

Again, let x_1, x_2, x_3, and x_4 denote the number of times σ_1, σ_2, σ_3, and σ_4 appear in $\vec{\pi}_k$, respectively. Note that $x_3, x_4 \in \{0, 1\}$. To have $a_1 \succ_{\hat{\sigma}} a_m$ for every $\hat{\sigma} \in f_2(\vec{\pi}_k)$, we need

$$\begin{aligned} & x_1 \cdot s_1 + x_2 \cdot s_m + x_3 \cdot s_1 + x_4 \cdot s_{t+1} \\ & \quad > x_1 \cdot s_m + x_2 \cdot s_1 + x_3 \cdot s_m + x_4 \cdot s_t \\ & \Rightarrow (x_1 + x_3) \cdot (s_1 - s_m) \\ & \quad > x_2 \cdot (s_1 - s_m) + x_4 \cdot (s_t - s_{t+1}). \end{aligned} \tag{1}$$

On the other hand, to have $a_{m-1} \succ_{\hat{\sigma}} a_2$ for every $\hat{\sigma} \in f_2(\vec{\pi}_k)$, we need

$$\begin{aligned} & x_1 \cdot s_{m-1} + x_2 \cdot s_2 + x_3 \cdot s_{t+1} + x_4 \cdot s_2 \\ & \quad > x_1 \cdot s_2 + x_2 \cdot s_{m-1} + x_3 \cdot s_t + x_4 \cdot s_{m-1} \\ & \Rightarrow (x_2 + x_4) \cdot (s_2 - s_{m-1}) \\ & \quad > x_1 \cdot (s_2 - s_{m-1}) + x_3 \cdot (s_t - s_{t+1}). \end{aligned} \tag{2}$$

We now distinguish between four cases.

Case 1: $x_3 = x_4 = 0$**.** Since $s_1 > s_m$, from Equation (1), we obtain $x_1 > x_2$. Moreover, as $s_2 \geq s_{m-1}$, from Equation (2), we obtain $x_2 > x_1$, which is a contradiction.

Case 2: $x_3 = 1$ **and** $x_4 = 0$**.** Since $s_1 > s_m$, from Equation (1), we obtain $x_1 + 1 > x_2$. Moreover, as $s_2 \geq s_{m-1}$ and $s_t \geq s_{t+1}$, from Equation (2), we obtain $x_2 > x_1$, which is a contradiction.

Case 3: $x_3 = 0$ **and** $x_4 = 1$**.** This case leads to a contradiction in a manner similar to Case 2, so we omit the details.

Case 4: $x_3 = 1$ **and** $x_4 = 1$**.** Since $s_1 > s_m$ and $s_t \geq s_{t+1}$, from Equation (1) we obtain that $x_1 + 1 > x_2$. Similarly, since $s_2 \geq s_{m-1}$ and $s_t \geq s_{t+1}$, from Equation (2) we obtain $x_2 + 1 > x_1$. This implies $x_1 = x_2$, which implies $|\vec{\pi}_k| = x_1 + x_2 + 2$ is even, which in turn contradicts the fact that k is odd. □

Unfortunately, the proof of Theorem 1 does not directly work when $f_1 \in \mathcal{F}_2$. However, a similar proof with somewhat more intricate profiles works, yielding the following result. Its proof is given in the full version.[3]

Theorem 2. *Let $n \geq 4$, $m \geq 5$, and $k \in [n-1]$ such that n and k have different parity. Then, for any positional scoring rules $f_1 \in \mathcal{F}_2$ and $f_2 \in \mathcal{F}$, we have* $\mathrm{acc}(f_1, f_2) = 0$.

Predicting a Rule from $\mathcal{F}_3$

The remaining family $\mathcal{F}_3$ contains exactly two voting rules: plurality (denoted f_{plu}) and veto (denoted f_{veto}). These two rules are special within the family of positional scoring rules. While Theorems 1 and 2 establish that every other positional scoring rule has zero worst-case predictability, we will show that this is not the case with plurality or veto.

Proposition 1. *Let $k \geq m(m-1)/2$. Then, we have* $\mathrm{acc}(f_{plu}, f_{plu}) > 0$ *and* $\mathrm{acc}(f_{veto}, f_{veto}) > 0$.

Proof. We provide a proof for plurality. The proof for veto is similar. Consider any profile $\vec{\sigma}_n$. Let x_i denote the number of times alternative a_i is ranked first. Without loss of generality, assume $x_i \geq x_{i+1}$ for all $i \in [m-1]$. Now, we construct a sample $\vec{\pi}_k \in \mathcal{S}_k(\vec{\sigma}_n)$ such that $f_{\text{plu}}(\vec{\pi}_k) \subseteq f_{\text{plu}}(\vec{\sigma}_n)$.

Let $y_i = |\{j \in [i, m-1] : x_j > x_{j+1}\}|$ for each $i \in [m-1]$, and $y_m = 0$. We begin by choosing y_i arbitrary rankings from $\vec{\sigma}_n$ which rank a_i first, for each i, and adding them to the sample. It is easy to see that $y_i = y_{i+1}$ if and only if $x_i = x_{i+1}$ and $y_i > y_{i+1}$ if and only if $x_i > x_{i+1}$, for all $i \in [m-1]$. Further, $\sum_{i=1}^{m} y_i \leq m(m-1)/2 \leq k$. If we ran plurality on the sample constructed so far, the set of rankings it returns would be precisely $f_{\text{plu}}(\vec{\sigma}_n)$. However, this sample may contain fewer than k votes.

We complete the sample $\vec{\pi}_k$ by adding any remaining votes which rank a_1 first, then adding any remaining votes which rank a_2 first, etc, until the sample size becomes k. Let z_i denote the final number of votes in the sample which rank a_i first. Then, for all $i, j \in [m]$, $x_i > x_j$ implies $y_i > y_j$, which implies $z_i > z_j$. Thus, $f_{\text{plu}}(\vec{\pi}_k) \subseteq f_{\text{plu}}(\vec{\sigma}_n)$. □

Proposition 1 raises two important questions: a) *How well can plurality or veto predict itself?*; and b) *Can some positional scoring rule predict plurality (resp. veto) better than plurality (resp. veto) itself?*

[3]The full version is available at: http://www.cs.toronto.edu/~nisarg/papers/samples.pdf

We begin by answering the latter question negatively. We show that among all positional scoring rules, the best predictor of plurality (resp. veto) is plurality (resp. veto) itself.

Theorem 3. *For every positional scoring rule $f_2 \in \mathcal{F}$, we have that* $\mathrm{acc}(f_{plu}, f_2) \le \mathrm{acc}(f_{plu}, f_{plu})$ *and* $\mathrm{acc}(f_{veto}, f_2) \le \mathrm{acc}(f_{veto}, f_{veto})$.

Proof. Fix $f_2 \in \mathcal{F}$. We show $\mathrm{acc}(f_{\text{plu}}, f_2) \le \mathrm{acc}(f_{\text{plu}}, f_{\text{plu}})$. The proof for $\mathrm{acc}(f_{\text{veto}}, f_2) \le \mathrm{acc}(f_{\text{veto}}, f_{\text{veto}})$ is similar.

Specifically, we show that for every profile $\vec{\sigma}_n$, there exists a profile $\vec{\tau}_n$ such that $\mathrm{acc}(f_{\text{plu}}, f_{\text{plu}}, \vec{\sigma}_n) \ge \mathrm{acc}(f_{\text{plu}}, f_2, \vec{\tau}_n)$. This implies the desired result.

Consider any profile $\vec{\sigma}_n$. Fix $\sigma^* \in f_{\text{plu}}(\vec{\sigma}_n)$. We construct the profile $\vec{\tau}_n$ as follows. In each ranking τ_i, the alternative ranked first in σ_i is also ranked first, and the remaining alternatives are in the opposite order of how they appear in σ^*. Because we do not change the alternatives in the first position, we have that $\mathrm{acc}(f_{\text{plu}}, f_{\text{plu}}, \vec{\sigma}_n) = \mathrm{acc}(f_{\text{plu}}, f_{\text{plu}}, \vec{\tau}_n)$. We now show that $\mathrm{acc}(f_{\text{plu}}, f_{\text{plu}}, \vec{\tau}_n) \ge \mathrm{acc}(f_{\text{plu}}, f_2, \vec{\tau}_n)$. More specifically, we show that for every sample $\vec{\pi}_k \in \mathcal{S}_k(\vec{\tau}_n)$, $f_{\text{plu}}(\vec{\pi}_k) \not\subseteq f_{\text{plu}}(\vec{\tau}_n)$ implies $f_2(\vec{\pi}_k) \not\subseteq f_{\text{plu}}(\vec{\tau}_n)$.

Consider any sample $\vec{\pi}_k \in \mathcal{S}_k(\vec{\tau}_n)$. Let x_i and y_i denote the number of times a_i is ranked first in $\vec{\tau}_n$ and $\vec{\pi}_k$, respectively. Suppose $f_{\text{plu}}(\vec{\pi}_k) \not\subseteq f_{\text{plu}}(\vec{\tau}_n)$. Then, there exist alternatives $a_i, a_j \in A$ such that $x_i > x_j$ but $y_i \le y_j$. Note that $x_i > x_j$ implies that $a_i \succ_{\sigma^*} a_j$. Hence, in every ranking in $\vec{\tau}_n$ (and therefore in $\vec{\pi}_k$) where a_i or a_j is not ranked first, a_j must appear before a_i (since we order them in the opposite order of σ^*). This, together with $y_j \ge y_i$, implies that under $\vec{\pi}_k$, f_2 assigns at least as much score to a_j as to a_i. Hence, there exists $\hat{\sigma} \in f_2(\vec{\pi}_k)$ for which $a_j \succ_{\hat{\sigma}} a_i$, and thus $\hat{\sigma} \notin f_{\text{plu}}(\vec{\tau}_n)$. Hence, we conclude $f_2(\vec{\pi}_k) \not\subseteq f_{\text{plu}}(\vec{\tau}_n)$. □

From Theorem 3, an upper bound on the worst-case accuracy of predicting plurality using plurality gives us an upper bound on the worst-case predictability of plurality. The same holds for veto. While Proposition 1 shows that this quantity is non-zero for $k \ge m(m-1)/2$, we show that it is still exponentially small in m when k is small compared to n.

Theorem 4. *For $n \ge (m-1)(m-2)/2$ and $k \le cn$, where $c < 1$ is a constant, we have* $\mathrm{acc}(f_{plu}) = \mathrm{acc}(f_{plu}, f_{plu}) \le c^{\Omega(m^2)}$ *and* $\mathrm{acc}(f_{veto}) = \mathrm{acc}(f_{veto}, f_{veto}) \le c^{\Omega(m^2)}$.

Proof. We provide a proof for plurality. The proof for veto is similar. Note that $\mathrm{acc}(f_{\text{plu}}) = \mathrm{acc}(f_{\text{plu}}, f_{\text{plu}})$ follows from Theorem 3. We now show that there exists a profile $\vec{\sigma}_n$ for which $\mathrm{acc}(f_{\text{plu}}, f_{\text{plu}}) \le \mathrm{acc}(f_{\text{plu}}, f_{\text{plu}}, \vec{\sigma}_n) \le c^m$.

Consider the profile $\vec{\sigma}_n$ in which alternative a_i appears first in exactly $i-1$ rankings, for each $i \in [m-1]$. In every other ranking, alternative a_m appears first. This is feasible because $n \ge (m-1)(m-2)/2$. Note that for $f_{\text{plu}}(\vec{\pi}_k) \subseteq f_{\text{plu}}(\vec{\sigma}_n)$, sample $\vec{\pi}_k$ must at least contain all of $t = (m-1)(m-2)/2$ rankings in which an alternative from $\{a_1, \ldots, a_{m-1}\}$ is ranked first. For $k < (m-1)(m-2)/2$, this happens with zero probability. For $k \ge (m-1)(m-2)/2$, this happens with probability at most $\binom{n-t}{k-t}/\binom{n}{k} \le (k/n)^t \le c^t$. □

We conjecture that even when $k = n - o(n)$, $\mathrm{acc}(f_{\text{plu}}, f_{\text{plu}})$ and $\mathrm{acc}(f_{\text{veto}}, f_{\text{veto}})$ are still $O(1/m)$. For $k = n-1$, it is easy to see that they are in fact $\Theta(1/m)$.

Theorem 5. *For $k = n-1$,* $\mathrm{acc}(f_{plu}) = \mathrm{acc}(f_{plu}, f_{plu}) = \Theta(1/m)$ *and* $\mathrm{acc}(f_{veto}) = \mathrm{acc}(f_{veto}, f_{veto}) = \Theta(1/m)$.

Proof. Once again, we provide a proof for plurality. The proof for veto is similar. Assume $n \gg m$. Given Theorem 3, we simply need to show that $\mathrm{acc}(f_{\text{plu}}, f_{\text{plu}}) = \Theta(1/m)$.

First, we show the upper bound. For all $i \in [m]$, define

$$x_i = \begin{cases} \lceil n/m \rceil + 1 & \text{if } i = 1 \\ \lceil n/m \rceil & \text{if } 2 \le i \le n \bmod m \\ \lfloor n/m \rfloor & \text{if } n \bmod m \le i \le m-1 \\ \lfloor n/m \rfloor - 1 & \text{if } i = m \end{cases}$$

Note that this satisfies $x_m < x_i$ for all $i \in [m-1]$ and $\sum_{i=1}^m x_i = n$. Now, consider a profile $\vec{\sigma}_n$ in which alternative a_i is in the top position in x_i rankings, for each $i \in [m]$. For every $\sigma^* \in f_{\text{plu}}(\vec{\sigma}_n)$, we have $a_i \succ_{\sigma^*} a_m$ for all $i \in [m-1]$. Consider a sample $\vec{\pi}_{n-1} \in \mathcal{S}_{n-1}(\vec{\sigma}_n)$. To have $a_i \succ_{\hat{\sigma}} a_m$ for all $i \in [m-1]$ and $\hat{\sigma} \in f_{\text{plu}}(\vec{\pi}_{n-1})$, $\vec{\pi}_{n-1}$ must contain all rankings of $\vec{\sigma}_n$ except a ranking in which a_m appears first. This happens with probability $(\lfloor n/m \rfloor - 1)/n = O(1/m)$. Hence, $\mathrm{acc}(f_{\text{plu}}, f_{\text{plu}}) \le \mathrm{acc}(f_{\text{plu}}, f_{\text{plu}}, \vec{\sigma}_n) = O(1/m)$.

Next, we show the lower bound. Consider any profile $\vec{\sigma}_n$. For $i \in [m]$, let x_i denote the number of times alternative a_i appears first. Without loss of generality, assume $x_i \ge x_{i+1}$ for $i \in [m-1]$. Let i^* be the smallest index such that $x_{i^*} \ge x_{i^*+1} + 2$ (if $x_i \le x_{i+1} + 1$ for all $i \in [m-1]$, then let $i^* = m$). It is easy to see that $x_{i^*} = \Omega(n/m)$, and for any $\vec{\pi}_{n-1} \in \mathcal{S}_{n-1}(\vec{\sigma}_n)$ which is obtained by removing one of the rankings in which a_{i^*} appears first, $f_{\text{plu}}(\vec{\pi}_{n-1}) \subseteq f_{\text{plu}}(\vec{\sigma}_n)$. Hence, $\mathrm{acc}(f_{\text{plu}}, f_{\text{plu}}) = \Omega(n/m)/n = \Omega(1/m)$. □

4 Average-Case Predictability

In the previous section, we considered the accuracy of predicting the outcome of a voting rule f_1 using a voting rule f_2 in the *worst case* over the underlying profile $\vec{\sigma}_n$, and defined $\mathrm{acc}(f_1, f_2) = \min_{\vec{\sigma}_n} \mathrm{acc}(f_1, f_2, \sigma_n)$.

In this section, we take a less pessimistic viewpoint, assume that the profile $\vec{\sigma}_n$ consists of n rankings drawn iid from a known prior $\mathcal{D}$, and define $\mathrm{acc}^{\mathcal{D}}(f_1, f_2) = \mathbb{E}_{\vec{\sigma}_n \sim \mathcal{D}^n}[\mathrm{acc}(f_1, f_2, \vec{\sigma}_n)]$, where $\vec{\sigma}_n \sim \mathcal{D}^n$ denotes that $\vec{\sigma}_n$ is drawn from the product distribution $\mathcal{D}^n$.

We show that this leads to interesting phenomena even in the simplest setting with two alternatives. Let $A = \{a, b\}$. Without loss of generality, suppose $\mathcal{D}$ generates $a \succ b$ with probability $p \ge 1/2$ and $b \succ a$ with probability $1-p$. This coincides with the Mallows model with central ranking $\sigma^* = a \succ b$ and noise parameter $\varphi = (1-p)/p \in [0,1]$.

For two alternatives, all reasonable voting rules (including all positional scoring rules) coincide with plurality, which is simply the majority rule. Hence, we fix the target voting rule as plurality ($f_1 = f_{\text{plu}}$). Our goal is to predict which of $a \succ b$ and $b \succ a$ appears more frequently in the underlying profile $\vec{\sigma}_n$. Without any distributional information, we cannot outperform running plurality on the sample $\vec{\pi}_k$, i.e., using $f_2 = f_{\text{plu}}$ (Theorem 3). However, with the knowledge of the prior, the optimal rule f_2 which maximizes $\text{acc}^{\mathcal{D}}(f_1, f_2)$ computes the posterior distribution of $\vec{\sigma}_n$ given both sample $\vec{\pi}_k$ and prior $\mathcal{D}$, and returns the more likely outcome of plurality on $\vec{\sigma}_n$ drawn from the posterior.

When the sample contains at least as many $a \succ b$ as $b \succ a$, the optimal rule would also return $a \succ b$. However, when the sample contains more $b \succ a$ than $a \succ b$, there is tension between the sample and the prior, and the output of the optimal rule is less clear.

Consider the extreme case in which the sample $\vec{\pi}_k$ consists of k copies of $b \succ a$. If $k \geq n/2$, the optimal rule safely returns $b \succ a$. When $k < n/2$, the optimal rule returns $b \succ a$ if $\Pr[f_{\text{plu}}(\vec{\sigma}_n) = b \succ a | \vec{\pi}_k] > \Pr[f_{\text{plu}}(\vec{\sigma}_n) = a \succ b | \vec{\pi}_k]$, but returns $a \succ b$ otherwise. It is easy to show that $\Pr[f_{\text{plu}}(\vec{\sigma}_n) = a \succ b | \vec{\pi}_k]$ is monotonically decreasing in k and in φ. Hence, there exists a unique φ_k^* such that the optimal rule returns $a \succ b$ when $\varphi < \varphi_k^*$ and returns $b \succ a$ when $\varphi > \varphi_k^*$. Further, φ_k^* is monotonically decreasing in k. The next result sheds more light on the relation between φ_k^* and k. Its proof is provided in the full version.

Theorem 6. *Let $n \geq 5$ and $n-1$ be divisible by 4. Given a sample $\vec{\pi}_k$ which consists of k copies of $b \succ a$, let φ_k^* be such that the optimal predictor returns $a \succ b$ if $\varphi < \varphi_k^*$ and $b \succ a$ if $\varphi > \varphi_k^*$. Then the following hold.*

1. *For $k = 1$, $\varphi_k^* \geq 1 - \frac{4 \ln n}{n+1}$.*
2. *For $k = (n-1)/2$, $\varphi_k^* \leq \frac{2}{n+1}$.*
3. *For $k = (n-1)/4$, $\varphi_k^* \in [1/4, 2/3]$.*

Let us consider the implications of Theorem 6 as $n \to \infty$. The first part implies that if we observe only a single $b \succ a$ sample, we should predict $a \succ b$ for any $\varphi < 1$. This makes sense because the $n-1$ unobserved votes vastly overshadow the single observed vote, and the prior places at least somewhat more probability on $a \succ b$ than on $b \succ a$.

The second part implies that if we observe $(n-1)/2$ votes (just a little less than a majority), we should predict $b \succ a$ for any $\varphi > 0$. This again makes sense because the probability that there is at least one $b \succ a$ in the remaining $(n+1)/2$ votes — sufficient to make $b \succ a$ the plurality outcome on the original profile — approaches 1.

The final part shows that the transition between $\varphi_k^* \approx 0$ and $\varphi_k^* \approx 1$ is not sudden; for $k = (n-1)/4$, the transition happens at φ_k^* that is not arbitrarily close to either endpoint when n is large.

5 Experiments

In this section, we conduct experiments to measure the predictability of popular voting rules in the average case.[4] We consider profiles $\vec{\sigma}_n$ with $n = 1,000$ voters and $m = 5$ alternatives. We use two distributions to draw i.i.d. rankings in $\vec{\sigma}_n$: the Mallows model with $\varphi = 1/3$ (in short, "Mallows distribution") and the uniform distribution. The former is more concentrated than the latter. We average our results across 10^6 draws of profile $\vec{\sigma}_n$.

Figure 1 shows the average predictability of different voting rules f_1 (rows) using different voting rules f_2 (columns), under the uniform distribution (tables on the top) and under the Mallows distribution (tables on the bottom), with $k = 50$ (tables on the left) and with $k = 500$ (tables on the right).[5] The entries in the table indicate the percentage of instances on which prediction was successful. Generally, we observe that prediction accuracy increases as the prior becomes more concentrated and as the number of samples k increases, as expected. We also note a few peculiarities. Under the uniform distribution with $k = 50$, the harmonic rule is the best predictor of every voting rule (except Bucklin), although the prediction accuracy is small. As k increases to 500, however, each voting rule (except Copeland and maximin) becomes the best predictor of itself. Under the Mallows distribution, it is evident that Borda, Bucklin, Copeland, and the harmonic rule predict other voting rules well — often because they return fewer ties — while maximin, plurality, STV, and veto perform worse.

Figure 2 shows the average-case predictability of different voting rules f_1 (using the best voting rule f_2 from the same list) as a function of the number of samples, under the Mallows distribution (left) and under the uniform distribution (right). Once again, more concentrated prior and more samples allow greater predictability. The effect of the prior is significant: under the Mallows distribution, observing just 3% of the votes allows predicting every voting rule with at least 98% accuracy, while the same number of samples under the uniform distribution does not allow predicting any voting rule with more than 4% accuracy.

6 Discussion

Predicting election outcomes using limited information is a broad research agenda, and while our work makes progress towards painting the full picture, there are a number of areas yet unexplored. The most immediate direction is to fill the gaps in our results, e.g., analyzing the accuracy of plurality and veto predicting themselves (Theorems 4 and 5) for all values of k, and extending the average-case analysis to heterogeneous samples, all values of k, and more than two alternatives. The next step would be to study other voting rules (e.g. Copeland's or Kemeny's method) and other models of sampling votes (e.g. when each voter i independently participates in the poll with probability p_i).

While Theorems 1 and 2 paint an extremely pessimistic picture of predictability of positional scoring rules, this could be because we want to predict the entire ranking of alternatives returned by the rule. This is indeed what is required in several real-world applications, e.g., Borda count is used to rank college football teams in the Associated Press poll (Levin and Nalebuff 1995) and to rank students in MOOCs (Caragiannis et al. 2019). But sometimes we may be interested in predicting only the top alternative. Could this lead to more optimistic results? The answer is yes and

[4]Refer to the book by Brandt et al. [2016] for definitions.

[5]For the Bucklin rule, we define the Bucklin score of an alternative as the smallest t such that a majority of voters rank the alternative in the first t positions. Alternatives are first compared by their Bucklin score (lower is better), and alternatives with the same Bucklin score t are compared by the number of voters who rank them in the first t positions (higher is better).

	Borda	Bucklin	Copeland	Harmonic	Maximin	Plurality	STV	Veto
Borda	1.40	1.26	0.43	1.65	0.59	0.61	0.79	0.59
Bucklin	1.27	1.39	0.38	1.37	0.53	0.46	0.65	0.54
Copeland	4.18	3.81	2.46	4.48	2.64	2.61	3.01	2.55
Harmonic	1.22	1.04	0.36	1.67	0.52	0.67	0.79	0.44
Maximin	1.59	1.47	0.51	1.82	0.77	0.73	0.90	0.71
Plurality	1.30	1.05	0.43	1.88	0.61	0.89	0.98	0.51
STV	1.36	1.16	0.49	1.86	0.62	0.84	1.19	0.54
Veto	1.33	1.23	0.45	1.35	0.59	0.52	0.66	0.89

	Borda	Bucklin	Copeland	Harmonic	Maximin	Plurality	STV	Veto
Borda	8.32	5.73	3.83	6.49	4.93	3.12	4.03	3.32
Bucklin	5.67	8.16	2.67	3.54	3.55	1.59	2.24	2.54
Copeland	15.67	11.75	12.10	12.99	10.40	7.81	10.44	8.16
Harmonic	5.90	3.02	2.72	8.43	3.80	5.70	5.46	1.60
Maximin	6.77	5.05	3.03	5.73	6.59	3.13	4.17	3.12
Plurality	4.08	1.97	2.08	7.99	2.93	8.12	5.76	1.20
STV	5.01	2.76	3.13	7.21	3.77	5.46	10.66	1.54
Veto	4.48	3.63	2.19	2.56	3.01	1.31	1.78	8.34

	Borda	Bucklin	Copeland	Harmonic	Maximin	Plurality	STV	Veto
Borda	99.84	99.80	99.95	98.69	62.54	40.35	40.65	40.43
Bucklin	99.84	99.80	99.95	98.69	62.54	40.35	40.65	40.43
Copeland	99.84	99.80	99.95	98.69	62.54	40.35	40.65	40.43
Harmonic	99.84	99.80	99.95	98.69	62.54	40.35	40.65	40.43
Maximin	99.84	99.80	99.95	98.69	62.54	40.35	40.65	40.43
Plurality	99.64	99.60	99.75	98.52	62.50	40.42	40.72	40.34
STV	99.64	99.60	99.75	98.52	62.50	40.42	40.72	40.34
Veto	99.84	99.80	99.95	98.69	62.54	40.35	40.65	40.55

	Borda	Bucklin	Copeland	Harmonic	Maximin	Plurality	STV	Veto
Borda	100	100	100	100	99.87	97.35	97.35	97.60
Bucklin	100	100	100	100	99.87	97.35	97.35	97.60
Copeland	100	100	100	100	99.87	97.35	99.35	97.60
Harmonic	100	100	100	100	99.87	97.35	97.35	97.60
Maximin	100	100	100	100	99.87	97.35	97.35	97.60
Plurality	99.9	99.9	99.9	99.9	99.77	97.66	97.66	97.50
STV	99.9	99.9	99.9	99.9	99.77	97.66	97.66	97.50
Veto	99.9	99.9	99.9	99.9	99.77	97.25	97.25	97.75

Figure 1: Average-case predictability of different voting rules f_1 (rows) using different voting rules f_2 (columns) under the uniform distribution (top) and the Mallows model with $\varphi = 1/3$ (bottom) with $k = 50$ (left) and $k = 500$ (right).

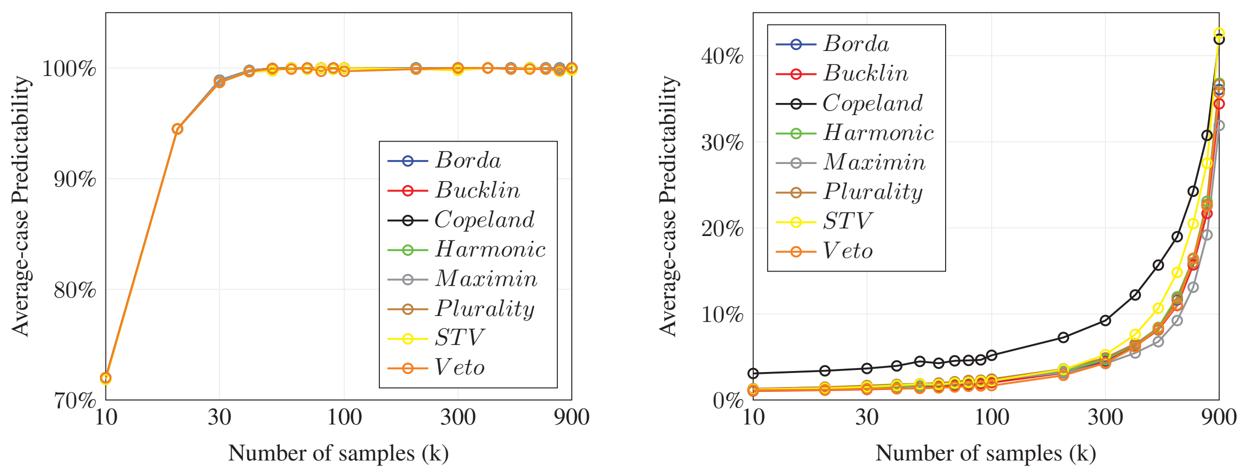

Figure 2: Average-case predictability of different voting rules as a function of the number of samples k, under the Mallows model with $\varphi = 1/3$ (left) and under the uniform distribution (right).

no. We show that the worst-case accuracy of predicting the Borda count winner using Borda count itself is still zero.

Theorem 7. *Let $n \geq 2$, $m \geq 4$, and $k \in [n-1]$ such that n and k have different parity. Then,* $\mathrm{acc}(f_{Borda}, f_{Borda}) = 0$, *where f_{Borda} denotes Borda count.*

But the case of the r-approval rule *is* more optimistic.

Theorem 8. *For $r \in [m-1]$ and $k \geq r$, we have* $\mathrm{acc}(f_{r\text{-}app}, f_{r\text{-}app}) > 0$, *where $f_{r\text{-}app}$ denotes the r-approval voting rule.*

The lower bound of r is tight. The proofs of these results appear in the full version.

Finally, we can also consider the use of limited information to make good collective decisions in other frameworks of voting. For example, in the implicit utilitarian voting framework (Procaccia and Rosenschein 2006; Boutilier et al. 2015), where the goal is to find an alternative with small distortion, how small can we make the expected distortion given only sampled votes?

References

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

Aziz, H.; Brill, M.; Fischer, F.; Harrenstein, P.; Lang, J.; and Seedig, H. G. 2015. Possible and necessary winners of partial tournaments. *Journal of Artificial Intelligence Research* 54:493–534.

Baumeister, D.; Faliszewski, P.; Lang, J.; and Rothe, J. 2012. Campaigns for lazy voters: truncated ballots. In *Proceedings of the 11th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 577–584.

Borodin, A.; Lev, O.; Shah, N.; and Strangway, T. 2019. Primarily about primaries. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, 1804–1811.

Boutilier, C.; Caragiannis, I.; Haber, S.; Lu, T.; Procaccia, A. D.; and Sheffet, O. 2015. Optimal social choice functions: A utilitarian view. *Artificial Intelligence* 227:190–213.

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

Caragiannis, I.; Chatzigeorgiou, X.; Krimpas, G. A.; and Voudouris, A. A. 2019. Optimizing positional scoring rules for rank aggregation. *Artificial Intelligence* 267:58–77.

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

Condorcet, M. d. 1785. Essai sur l'application de l'analyse à la probabilité de décisions rendues à la pluralité de voix. Imprimerie Royal. Facsimile published in 1972 by Chelsea Publishing Company, New York.

Dey, P., and Bhattacharyya, A. 2015. Sample complexity for winner prediction in elections. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 1421–1430.

Doucette, J. A.; Larson, K.; and Cohen, R. 2014. Approximate winner selection in social choice with partial preferences. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Elkind, E.; Faliszewski, P.; and Slinko, A. 2015. Distance rationalization of voting rules. *Social Choice and Welfare* 45(2):345–377.

Graefe, A. 2014. Accuracy of vote expectation surveys in forecasting elections. *Public Opinion Quarterly* 78(S1):204–232.

Hazon, N.; Aumann, Y.; Kraus, S.; and Wooldridge, M. 2012. On the evaluation of election outcomes under uncertainty. *Artificial Intelligence* 189:1–18.

Lang, J.; Pini, M. S.; Rossi, F.; Salvagnin, D.; Venable, K. B.; and Walsh, T. 2012. Winner determination in voting trees with incomplete preferences and weighted votes. *Autonomous Agents and Multi-Agent Systems* 25(1):130–157.

Levin, J., and Nalebuff, B. 1995. An introduction to vote-counting schemes. *Journal of Economic Perspectives* 9(1):3–26.

Lewis-Beck, M. S., and Tien, C. 1999. Voters as forecasters: a micromodel of election prediction. *International Journal of Forecasting* 15(2):175–184.

Lewis-Beck, M. S. 2005. Election forecasting: principles and practice. *The British Journal of Politics and International Relations* 7(2):145–164.

Meskanen, T., and Nurmi, H. 2008. Closeness counts in social choice. In Braham, M., and Steffen, F., eds., *Power, Freedom, and Voting*. Springer-Verlag.

Moulin, H. 1988. *Axioms of Cooperative Decision Making*, volume 15 of *Econometric Society monographs*. Cambridge University Press.

Procaccia, A. D., and Rosenschein, J. S. 2006. The distortion of cardinal preferences in voting. In *Proceedings of the 10th International Workshop on Cooperative Information Agents (CIA)*, 317–331.

Rothschild, D., and Wolfers, J. 2011. Forecasting elections: Voter intentions versus expectations. *SSRN Electronic Journal*.

Tumasjan, A.; Sprenger, T. O.; Sandner, P. G.; and Welpe, I. M. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the 4th AAAI conference on Weblogs and Social Media (ICWSM)*.

Walther, D. 2015. Picking the winner(s): Forecasting elections in multiparty systems. *Electoral Studies* 40:1–13.

Wojtas, K., and Faliszewski, P. 2012. Possible winners in noisy elections. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI)*, 1499–1505.

Young, H. P. 1988. Condorcet's theory of voting. *The American Political Science Review* 82(4):1231–1244.