# Can We Predict the Election Outcome from Sampled Votes?*

Evi Micha
Department of Computer Science
University of Toronto
emicha@cs.toronto.edu

Nisarg Shah
Department of Computer Science
University of Toronto
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 [9], 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 [1, 20], distance rationalizability [19, 12], noisy voting [26, 7], and implicit utilitarian voting [21, 5]. 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 [23, 22, 24, 13, 18, 17, 10].

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

---

*A preliminary version appeared in the *Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020*.

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.

## 1.1   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.

## 1.2   Related Work

Most closely related to ours is the work of Dey and Bhattacharyya [10]. 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

---

[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.

[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.

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 [10] 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. [4]. 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 [11, 3, 15, 2], uncertainty about which voters or candidates would participate in the election (even if all preferences are known upfront) [25], or distributional uncertainty about *each* voter's preferences [14].

## 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 $\sigma$. 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 $t^{\text{th}}$ 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 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 $\sigma^*$ 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 $\sigma^*$. 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 $\mathrm{acc}(f_1, f_2, \vec{\sigma}_n) = \mathrm{Pr}_{\vec{\pi}_k \sim U_k(\vec{\sigma}_n)}[f_2(\vec{\pi}_k) \subseteq f_1(\vec{\sigma}_n)]$.

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.

## 3.1 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 $\sigma_1$, $\sigma_2$ and $\sigma_3$ 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 $\sigma_1$, $\sigma_2$, and $\sigma_3$ 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

$$x_1 \cdot s_{m-1} + x_2 \cdot s_2 + x_3 \cdot s_2 > x_1 \cdot s_2 + x_2 \cdot s_{m-1} + x_3 \cdot s_{m-1}$$
$$\Rightarrow (x_2 + x_3) \cdot (s_2 - s_{m-1}) > x_1 \cdot (s_2 - s_{m-1}).$$

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

4

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 $p, q \in \{2, \ldots, m-2\}$ 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. $\qquad\square$

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

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

We again denote with $\sigma_1$, $\sigma_2$, $\sigma_3$ and $\sigma_4$ the rankings in rows 1, 2, 3 and 4, respectively. In ranking $\sigma_3$, $a_2$ is at position $t$ and $a_{m-1}$ is at position $t+1$. In the ranking $\sigma_4$, $a_m$ 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 $\sigma_1$, $\sigma_2$, $\sigma_3$, and $\sigma_4$ 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

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

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

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

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. $\qquad\square$

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.

**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 $\mathrm{acc}(f_1, f_2) = 0$.*

*Proof.* Fix positional scoring rules $f_1 \in \mathcal{F}_2$ and $f_2 \in \mathcal{F}$. Let $\vec{r}$ and $\vec{s}$ denote their scoring vectors, respectively. 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 $\vec{\sigma}_n$. Alternatives not shown appear in an arbitrary order in the middle.

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

We denote with $\sigma_1$, $\sigma_2$ and $\sigma_3$ the ranking of the first, second and third rows, respectively. Because $f_1 \in \mathcal{F}_2$, we have $r_1 > r_2 = r_{m-1} > r_m$, and hence for every $\sigma^* \in f_1(\vec{\sigma}_n)$, it holds $a_1 \succ_{\sigma^*} a_m$ 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 $\sigma_1$, $\sigma_2$, and $\sigma_3$ 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

$$x_1 \cdot s_1 + x_2 \cdot s_2 + x_3 \cdot s_1 > x_1 \cdot s_2 + x_2 \cdot s_1 + x_3 \cdot s_2.$$
$$\Rightarrow (x_1 + x_3) \cdot (s_1 - s_2) > x_2 \cdot (s_1 - s_2).$$

Given that $s_1 \geq s_2$, 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

$$x_1 \cdot s_m + x_2 \cdot s_{m-1} + x_3 \cdot s_{m-1} > x_1 \cdot s_{m-1} + x_2 \cdot s_m + x_3 \cdot s_m$$
$$\Rightarrow (x_2 + x_3) \cdot (s_{m-1} - s_m) > x_1 \cdot (s_{m-1} - s_m).$$

Given $s_{m-1} \geq s_m$, 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 $k$. Consider the following profile $\vec{\sigma}_n$.

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

We again denote with $\sigma_1$, $\sigma_2$, $\sigma_3$ and $\sigma_4$, the rankings in rows 1, 2, 3 and 4, respectively. Fix an arbitrary $t$ such that $2 < t < m - 2$. In ranking $\sigma_3$, alternative $a_2$ is at some position $t$ and alternative $a_{m-1}$ is at position $t + 1$. In ranking $\sigma_4$, alternative $a_m$ is at position $t$ and alternative $a_1$ 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. Recall that for the scoring vector $\vec{r}$ of $f_1 \in \mathcal{F}_2$, we have $r_1 > r_2$ and $r_{m-1} > r_m$, while $r_t = r_{t+1}, \forall t \in [2, m-2]$. 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. Here, we have to take four cases depending on whether each of $\sigma_3$ and $\sigma_4$ appears in $\vec{\pi}_k$.

Again, let $x_1$, $x_2$, $x_3$, and $x_4$ denote the number of times $\sigma_1$, $\sigma_2$, $\sigma_3$, and $\sigma_4$ 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

$$x_1 \cdot s_1 + x_2 \cdot s_2 + x_3 \cdot s_1 + x_4 \cdot s_{t+1} > x_1 \cdot s_2 + x_2 \cdot s_1 + x_3 \cdot s_2 + x_4 \cdot s_t$$
$$\Rightarrow (x_1 + x_3) \cdot (s_1 - s_2) > x_2 \cdot (s_1 - s_2) + x_4 \cdot (s_t - s_{t+1}). \tag{3}$$

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

$$x_1 \cdot s_m + x_2 \cdot s_{m-1} + x_3 \cdot s_{t+1} + x_4 \cdot s_{m-1}$$
$$> x_1 \cdot s_{m-1} + x_2 \cdot s_m + x_3 \cdot s_t + x_4 \cdot s_m$$
$$\Rightarrow (x_2 + x_4) \cdot (s_{m-1} - s_m) > x_1 \cdot (s_{m-1} - s_m) + x_3 \cdot (s_t - s_{t+1}). \tag{4}$$

We now distinguish between four cases.

**Case 1:** $x_3 = x_4 = 0.$ Since $s_1 > s_2$, from Equation (3), we obtain $x_1 > x_2$. Moreover, as $s_{m-1} \geq s_m$, from Equation (4), we obtain $x_2 > x_1$, which is a contradiction.

**Case 2:** $x_3 = 1$ **and** $x_4 = 0.$ Since $s_1 > s_2$, from Equation (3), we obtain $x_1 + 1 > x_2$. Moreover, as $s_{m-1} \geq s_m$ and $s_t \geq s_{t+1}$, from Equation (4), 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 $s_1 > s_2$ and $s_t \geq s_{t+1}$, from Equation (3) we obtain that $x_1 + 1 > x_2$. Similarly, since $s_{m-1} \geq s_m$ and $s_t \geq s_{t+1}$, from Equation (4) 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.

□

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

The remaining family $\mathcal{F}_3$ contains exactly two voting rules: plurality (denoted $f_{\text{plu}}$) and veto (denoted $f_{\text{veto}}$). 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 $\text{acc}(f_{plu}, f_{plu}) > 0$ and $\text{acc}(f_{veto}, f_{veto}) > 0$.*

*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, $f_{\text{plu}}(\vec{\pi}_k) \subseteq f_{\text{plu}}(\vec{\sigma}_n)$. $\qquad\square$

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?*

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 $\text{acc}(f_{plu}, f_2) \leq \text{acc}(f_{plu}, f_{plu})$ and $\text{acc}(f_{veto}, f_2) \leq \text{acc}(f_{veto}, f_{veto})$.*

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

Specifically, we show that for every profile $\vec{\sigma}_n$, there exists a profile $\vec{\tau}_n$ such that $\text{acc}(f_{\text{plu}}, f_{\text{plu}}, \vec{\sigma}_n) \geq \text{acc}(f_{\text{plu}}, f_2, \vec{\tau}_n)$. 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 $\tau_i$, the alternative ranked first in $\sigma_i$ is also ranked first, and the remaining alternatives are in the opposite order of how they appear in $\sigma^*$. Because we do not change the alternatives in the first position, we have that $\text{acc}(f_{\text{plu}}, f_{\text{plu}}, \vec{\sigma}_n) = \text{acc}(f_{\text{plu}}, f_{\text{plu}}, \vec{\tau}_n)$. We now show that $\text{acc}(f_{\text{plu}}, f_{\text{plu}}, \vec{\tau}_n) \geq \text{acc}(f_{\text{plu}}, f_2, \vec{\tau}_n)$. 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 $y_i \leq y_j$. 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 $\sigma^*$). This, together with $y_j \geq y_i$, 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 $f_2(\vec{\pi}_k) \not\subseteq f_{\text{plu}}(\vec{\tau}_n)$. $\qquad\square$

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 $k \geq m(m-1)/2$, we show that it is still exponentially small in $m$ when $k$ is small compared to $n$.

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

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

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 $n \geq (m-1)(m-2)/2$. 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 $k \geq (m-1)(m-2)/2$, this happens with probability at most $\binom{n-t}{k-t}/\binom{n}{k} \leq (k/n)^t \leq c^t$. $\qquad\square$

We conjecture that even when $k = n - o(n)$, $\mathrm{acc}(f_{\mathrm{plu}}, f_{\mathrm{plu}})$ and $\mathrm{acc}(f_{\mathrm{veto}}, f_{\mathrm{veto}})$ 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 $\mathrm{acc}(f_{\mathrm{plu}}, f_{\mathrm{plu}}) = \Theta(1/m)$.

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

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

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 $\sigma^* \in f_{\mathrm{plu}}(\vec{\sigma}_n)$, 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 $\hat{\sigma} \in f_{\mathrm{plu}}(\vec{\pi}_{n-1})$, $\vec{\pi}_{n-1}$ 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, $\mathrm{acc}(f_{\mathrm{plu}}, f_{\mathrm{plu}}) \leq \mathrm{acc}(f_{\mathrm{plu}}, f_{\mathrm{plu}}, \vec{\sigma}_n) = O(1/m)$.

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 $x_i \geq x_{i+1}$ for $i \in [m - 1]$. Let $i^*$ be the smallest index such that $x_{i^*} \geq x_{i^*+1} + 2$ (if $x_i \leq x_{i+1} + 1$ 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, $f_{\mathrm{plu}}(\vec{\pi}_{n-1}) \subseteq f_{\mathrm{plu}}(\vec{\sigma}_n)$. Hence, $\mathrm{acc}(f_{\mathrm{plu}}, f_{\mathrm{plu}}) = \Omega(n/m)/n = \Omega(1/m)$. $\square$

# 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 $p \geq 1/2$ 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 ($f_1 = f_{\mathrm{plu}}$). 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 $f_2 = f_{\mathrm{plu}}$ (Theorem 3). However, with the knowledge of the prior, the optimal rule $f_2$ which maximizes $\mathrm{acc}^{\mathcal{D}}(f_1, f_2)$ 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 $\Pr[f_{\mathrm{plu}}(\vec{\sigma}_n) = b \succ a | \vec{\pi}_k] > \Pr[f_{\mathrm{plu}}(\vec{\sigma}_n) = a \succ b | \vec{\pi}_k]$, but returns $a \succ b$ otherwise. It is easy to show that $\Pr[f_{\mathrm{plu}}(\vec{\sigma}_n) = a \succ b | \vec{\pi}_k]$ is monotonically decreasing in $k$ and in $\varphi$. Hence, there exists a unique $\varphi_k^*$ such that the optimal rule returns

$a \succ b$ when $\varphi < \varphi_k^*$ and returns $b \succ a$ when $\varphi > \varphi_k^*$. Further, $\varphi_k^*$ is monotonically decreasing in $k$. The next result sheds more light on the relation between $\varphi_k^*$ 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 $\varphi_k^*$ 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 $k = (n-1)/4$, $\varphi_k^* \in [1/4, 2/3]$.*

*Proof.* We begin by examining the probability of each possible underlying plurality outcome given the samples. Recall that under the prior, the probability of generating $a \succ b$ is $p \in [1/2, 1]$, and we have $\varphi = (1-p)/p$. Given a sample $\vec{\pi}_k$ which consists of $k$ copies of $b \succ a$, we get

$$
\begin{aligned}
&\Pr[f_{\mathrm{plu}}(\vec{\sigma}_n) = a \succ b | \vec{\pi}_k] \\
&= \frac{\Pr[f_{\mathrm{plu}}(\vec{\sigma}_n) = a \succ b \wedge \vec{\pi}_k]}{\Pr[\vec{\pi}_k]} \\
&= \frac{\sum_{\vec{\sigma}_n} \Pr[\vec{\sigma}_n] \cdot \Pr[f_{\mathrm{plu}}(\vec{\sigma}_n) = a \succ b \wedge \vec{\pi}_k | \vec{\sigma}_n]}{\Pr[\vec{\pi}_k]} \\
&= \frac{\sum_{\vec{\sigma}_n} \Pr[\vec{\sigma}_n] \cdot \Pr[f_{\mathrm{plu}}(\vec{\sigma}_n)) = a \succ b | \vec{\sigma}_n] \cdot \Pr[\vec{\pi}_k | \vec{\sigma}_n]}{\Pr[\vec{\pi}_k]} \\
&= \frac{\sum_{\vec{\sigma}_n : f_{\mathrm{plu}}(\vec{\sigma}_n) = a \succ b} \Pr[\vec{\sigma}_n] \cdot \Pr[\vec{\pi}_k | \vec{\sigma}_n]}{\Pr[\vec{\pi}_k]} \\
&= \frac{\sum_{i=k}^{\frac{n-1}{2}} \binom{n}{i} p^{n-i}(1-p)^i \frac{\binom{i}{k}}{\binom{n}{k}}}{\Pr[\vec{\pi}_k]} \\
&= \frac{\sum_{i=k}^{\frac{n-1}{2}} \binom{i}{k}\binom{n}{i}\varphi^i p^n}{\binom{n}{k}\Pr[\vec{\pi}_k]}.
\end{aligned}
\tag{5}
$$

Similarly, we conclude that:

$$
\Pr[f_{\mathrm{plu}}(\vec{\sigma}_n) = b \succ a | \vec{\pi}_k] = = \frac{\sum_{i=\frac{n+1}{2}}^{n} \binom{i}{k}\binom{n}{i}\varphi^i p^n}{\binom{n}{k}\Pr[\vec{\pi}_k]}.
\tag{6}
$$

Recall that $\varphi_k^*$ is the unique value which satisfies $\Pr[f_{\mathrm{plu}}(\vec{\sigma}_n) = a \succ b | \vec{\pi}_k] = \Pr[f_{\mathrm{plu}}(\vec{\sigma}_n) = b \succ a | \vec{\pi}_k]$, i.e.,

$$
\sum_{i=k}^{\frac{n-1}{2}} \binom{i}{k}\binom{n}{i}(\varphi_k^*)^i = \sum_{i=\frac{n+1}{2}}^{n} \binom{i}{k}\binom{n}{i}(\varphi_k^*)^i.
$$

**Case 1:** $k = 1$. We want to show that $\varphi_k^* \geq 1 - \frac{4 \ln n}{n+1}$. Hence, it is sufficient to show that for all $\varphi < 1 - \frac{4 \ln n}{n+1}$,

$$
\sum_{i=k}^{\frac{n-1}{2}} \binom{i}{k}\binom{n}{i}\varphi^i > \sum_{i=\frac{n+1}{2}}^{n} \binom{i}{k}\binom{n}{i}\varphi^i.
$$

For $k = 1$, this reduces to

$$\sum_{i=1}^{\frac{n-1}{2}} \binom{n}{i} \cdot i \cdot \varphi^i > \sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} \cdot i \cdot \varphi^i$$

$$\Leftrightarrow \sum_{i=1}^{\frac{n-1}{2}} \binom{n}{i} \cdot \left(i\varphi^i - (n-i)\varphi^{n-i}\right) > n\varphi^n.$$

A sufficient condition for this to hold is that

$$i\varphi^i - (n-i)\varphi^{n-i} > 0, \forall i \in [(n-1)/2]$$

and

$$\frac{n-1}{2}\varphi^{\frac{n-1}{2}} - \frac{n+1}{2}\varphi^{\frac{n+1}{2}} \geq n\varphi^n.$$

For the former condition, it is sufficient to have

$$\frac{n-1}{2}\varphi^{\frac{n-1}{2}} - \frac{n+1}{2}\varphi^{\frac{n+1}{2}} > 0$$

$$\Leftrightarrow \varphi < \frac{(n-1)/2}{(n+1)/2} = 1 - \frac{2}{n+1}.$$

This is true because $\varphi < 1 - 4\ln n/(n+1)$. We can simplify the latter condition to

$$\frac{n-1}{2} \geq \frac{n+1}{2}\varphi + n\varphi^{\frac{n+1}{2}}.$$

For $\varphi \leq 1 - \frac{4\ln n}{n+1}$, we have

$$\frac{n+1}{2}\varphi + n\varphi^{\frac{n+1}{2}} \leq \frac{n+1-4\ln n}{2} + ne^{-\frac{4\ln n}{n+1} \cdot \frac{n+1}{2}}$$

$$= \frac{n+1-4\ln n}{2} + 1 \leq \frac{n-1}{2},$$

where the first transition holds because $1 - x \leq e^{-x}$ for all $x$, and the last transition holds when $\ln n \geq 1$, which is true for $n \geq 3$.

**Case 2:** $k = (n-1)/2$. In this case, we want to show $\varphi_k^* \leq 2/(n+1)$. It is sufficient to show that for all $\varphi > 2/(n+1)$,

$$\sum_{i=k}^{\frac{n-1}{2}} \binom{i}{k}\binom{n}{i}\varphi^i < \sum_{i=\frac{n+1}{2}}^{n} \binom{i}{k}\binom{n}{i}\varphi^i.$$

For $k = (n-1)/2$, this reduces to

$$\binom{n}{\frac{n-1}{2}}\varphi^{\frac{n-1}{2}} < \sum_{i=\frac{n+1}{2}}^{n} \binom{i}{\frac{n-1}{2}}\binom{n}{i}\varphi^i.$$

It suffices to show that

$$\binom{n}{\frac{n-1}{2}}\varphi^{\frac{n-1}{2}} < \binom{\frac{n+1}{2}}{\frac{n-1}{2}}\binom{n}{\frac{n+1}{2}}\varphi^{\frac{n+1}{2}}.$$

Upon simplification, it is easy to see that this is true when $\varphi > \frac{2}{n+1}$.

11

**Case 3:** $k = (n-1)/4$. In this case, we want to show that $\varphi_k^* \in [1/4, 2/3]$. It is sufficient to show that

$$\sum_{i=\frac{n-1}{4}}^{\frac{n-1}{2}} \binom{i}{\frac{n-1}{4}} \binom{n}{i} \varphi^i > \sum_{i=\frac{n+1}{2}}^{n} \binom{i}{\frac{n-1}{4}} \binom{n}{i} \varphi^i, \ \forall \varphi < 1/4, \tag{7}$$

and

$$\sum_{i=\frac{n-1}{4}}^{\frac{n-1}{2}} \binom{i}{\frac{n-1}{4}} \binom{n}{i} \varphi^i < \sum_{i=\frac{n+1}{2}}^{n} \binom{i}{\frac{n-1}{4}} \binom{n}{i} \varphi^i, \ \forall \varphi > 2/3. \tag{8}$$

We show each inequality separately.

**Case 3a:** $\varphi < 1/4$. For Equation (7), it suffices to show that for all $\varphi < 1/4$ and $i \in \{0, 1, \ldots, \frac{n-1}{4}\}$, we have

$$\binom{\frac{n-1}{2} - i}{\frac{n-1}{4}} \binom{n}{\frac{n-1}{2} - i} \varphi^{\frac{n-1}{2} - i} >$$

$$\binom{\frac{n+1}{2} + 2i}{\frac{n-1}{4}} \binom{n}{\frac{n+1}{2} + 2i} \varphi^{\frac{n+1}{2} + 2i}$$

$$+ \binom{\frac{n+1}{2} + 2i + 1}{\frac{n-1}{4}} \binom{n}{\frac{n+1}{2} + 2i + 1} \varphi^{\frac{n+1}{2} + 2i + 1}$$

$$\Leftarrow 1 > \frac{(\frac{n-1}{2} + i + 1) \cdots (\frac{n-1}{2} - 2i + 1)}{(\frac{n-1}{4} + 2i + 1) \cdots (\frac{n-1}{4} - i + 1)} \left(\frac{1}{4}\right)^{3i+1}$$

$$+ \frac{(\frac{n-1}{2} + i + 1) \cdots (\frac{n-1}{2} - 2i)}{(\frac{n-1}{4} + 2i + 2) \cdots (\frac{n-1}{4} - i + 1)} \left(\frac{1}{4}\right)^{3i+2}$$

We show that both terms of the right hand side are less than $1/2$ and the inequality follows. First, for $i \in \{0, \ldots, \frac{n-1}{4}\}$, we have

$$\left(\frac{n-1}{4} + 2i + 1\right) \cdot \frac{1}{2} > \left(\frac{n-1}{2} + i + 1\right) \cdot \frac{1}{4},$$

and for $i \in \{0, \ldots, \frac{n-1}{4}\}$ and $j \in [3i]$, we have

$$\frac{n-1}{4} + 2i + 1 - j > \left(\frac{n-1}{2} + i + 1 - j\right) \cdot \frac{1}{4}.$$

This shows that the first term is less than $1/2$. To show that the second term is less than $1/2$, it is sufficient to notice that for $i \in \{0, \ldots, \frac{n-1}{4}\}$,

$$\left(\frac{n-1}{4} + 2i + 2\right) \cdot \frac{1}{2} > \left(\frac{n-1}{2} + i + 1\right) \cdot \frac{1}{4},$$

and for $i \in \{0, \ldots, \frac{n-1}{4}\}$ and $j \in [3i + 1]$,

$$\frac{n-1}{4} + 2i + 2 - j > \left(\frac{n-1}{2} + i + 1 - j\right) \cdot \frac{1}{4}.$$

**Case 3b:** $\varphi > 2/3$. We now show Equation (8) for $\varphi = 2/3$. Then, the equation clearly holds for $\varphi > 2/3$.

12

First, note that for $i \in \left\{0, \ldots, \frac{n-1}{4} - 1\right\}$, it holds that

$$\binom{\frac{n+1}{2} + i}{\frac{n-1}{4}} \binom{n}{\frac{n+1}{2} + i} \left(\frac{2}{3}\right)^{\frac{n+1}{2} + i} \geq$$

$$\binom{\frac{n-1}{2} - i}{\frac{n-1}{4}} \binom{n}{\frac{n-1}{2} - i} \left(\frac{2}{3}\right)^{\frac{n-1}{2} - i}$$

$$\Leftrightarrow \left(\frac{2}{3}\right)^{2i+1} \geq \frac{\left(\frac{n-1}{4} + i + 1\right) \ldots \left(\frac{n-1}{4} - i + 1\right)}{\left(\frac{n-1}{2} + i + 1\right) \ldots \left(\frac{n-1}{2} - i + 1\right)}.$$

This is because for each $j \in \{0, \ldots, 2i\}$, we have

$$\frac{2}{3} \cdot \left(\frac{n-1}{2} + i + 1 - j\right) \geq \frac{n-1}{4} + i + 1 - j.$$

Next, we argue that

$$\binom{\frac{3n+1}{4}}{\frac{n-1}{4}} \binom{n}{\frac{3n+1}{4}} \left(\frac{2}{3}\right)^{\frac{3n+1}{4}}$$

$$+ \binom{\frac{3n+1}{4} + 1}{\frac{n-1}{4}} \binom{n}{\frac{3n+1}{4} + 1} \left(\frac{2}{3}\right)^{\frac{3n+1}{4} + 1} > \binom{n}{\frac{n-1}{4}} \left(\frac{2}{3}\right)^{\frac{n-1}{4}}$$

$$\Leftrightarrow \frac{\left(\frac{3n+1}{4}\right) \ldots \left(\frac{n+1}{2} + 1\right)}{\left(\frac{n-1}{4}\right)!} \left(\frac{2}{3}\right)^{\frac{n+1}{2}}$$

$$+ \frac{\left(\frac{3n+1}{4}\right) \ldots \left(\frac{n+1}{2} + 2\right)}{\left(\frac{n-1}{4} - 1\right)!} \left(\frac{2}{3}\right)^{\frac{n+1}{2} + 1} > 1.$$

To prove this, we show that the first term on the LHS is greater than $7/8$, and the second term on the LHS is greater than $1/8$. First, we notice that

$$\left(\frac{2}{3}\right)^3 \cdot \frac{3n+1}{4} > \frac{7}{8} \cdot \left(\frac{n-1}{4}\right).$$

Next, for each $j \in [\frac{n-1}{4} - 1]$,

$$\left(\frac{2}{3}\right)^2 \cdot \left(\frac{3n+1}{4} - j\right) > \frac{n-1}{4} - j.$$

This is sufficient to show that the first term is greater than $7/8$. For the second term, note that

$$\left(\frac{2}{3}\right)^6 \cdot \frac{3n+1}{4} > \frac{1}{8} \cdot \left(\frac{n-1}{4} - 1\right),$$

and for each $j \in [\frac{n-1}{4} - 2]$,

$$\left(\frac{2}{3}\right)^2 \left(\frac{3n+1}{4} - j\right) > \frac{n-1}{4} - 1 - j.$$

Thus, the second term is greater than $1/8$. This concludes the proof. $\qquad\square$

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 $\varphi_k^*$ that is not arbitrarily close to either endpoint when $n$ is large.

| | Borda | Bucklin | Copeland | Harmonic | Maximin | Plurality | STV | Veto | 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 | 8.32 | 5.73 | 3.83 | 6.49 | 4.93 | 3.12 | 4.03 | 3.32 |
| Bucklin | 1.27 | 1.39 | 0.38 | 1.37 | 0.53 | 0.46 | 0.65 | 0.54 | 5.67 | 8.16 | 2.67 | 3.54 | 3.55 | 1.59 | 2.24 | 2.54 |
| Copeland | 4.18 | 3.81 | 2.46 | 4.48 | 2.64 | 2.61 | 3.01 | 2.55 | 15.67 | 11.75 | 12.10 | 12.99 | 10.40 | 7.81 | 10.44 | 8.16 |
| Harmonic | 1.22 | 1.04 | 0.36 | 1.67 | 0.52 | 0.67 | 0.79 | 0.44 | 5.90 | 3.02 | 2.72 | 8.43 | 3.80 | 5.70 | 5.46 | 1.60 |
| Maximin | 1.59 | 1.47 | 0.51 | 1.82 | 0.77 | 0.73 | 0.90 | 0.71 | 6.77 | 5.05 | 3.03 | 5.73 | 6.59 | 3.13 | 4.17 | 3.12 |
| Plurality | 1.30 | 1.05 | 0.43 | 1.88 | 0.61 | 0.89 | 0.98 | 0.51 | 4.08 | 1.97 | 2.08 | 7.99 | 2.93 | 8.12 | 5.76 | 1.20 |
| STV | 1.36 | 1.16 | 0.49 | 1.86 | 0.62 | 0.84 | 1.19 | 0.54 | 5.01 | 2.76 | 3.13 | 7.21 | 3.77 | 5.46 | 10.66 | 1.54 |
| Veto | 1.33 | 1.23 | 0.45 | 1.35 | 0.59 | 0.52 | 0.66 | 0.89 | 4.48 | 3.63 | 2.19 | 2.56 | 3.01 | 1.31 | 1.78 | 8.34 |
| Borda | 99.84 | 99.80 | 99.95 | 98.69 | 62.54 | 40.35 | 40.65 | 40.43 | 100 | 100 | 100 | 100 | 99.87 | 97.35 | 97.35 | 97.60 |
| Bucklin | 99.84 | 99.80 | 99.95 | 98.69 | 62.54 | 40.35 | 40.65 | 40.43 | 100 | 100 | 100 | 100 | 99.87 | 97.35 | 97.35 | 97.60 |
| Copeland | 99.84 | 99.80 | 99.95 | 98.69 | 62.54 | 40.35 | 40.65 | 40.43 | 100 | 100 | 100 | 100 | 99.87 | 97.35 | 99.35 | 97.60 |
| Harmonic | 99.84 | 99.80 | 99.95 | 98.69 | 62.54 | 40.35 | 40.65 | 40.43 | 100 | 100 | 100 | 100 | 99.87 | 97.35 | 97.35 | 97.60 |
| Maximin | 99.84 | 99.80 | 99.95 | 98.69 | 62.54 | 40.35 | 40.65 | 40.43 | 100 | 100 | 100 | 100 | 99.87 | 97.35 | 97.35 | 97.60 |
| Plurality | 99.64 | 99.60 | 99.75 | 98.52 | 62.50 | 40.42 | 40.72 | 40.34 | 99.9 | 99.9 | 99.9 | 99.9 | 99.77 | 97.66 | 97.66 | 97.50 |
| STV | 99.64 | 99.60 | 99.75 | 98.52 | 62.50 | 40.42 | 40.72 | 40.34 | 99.9 | 99.9 | 99.9 | 99.9 | 99.77 | 97.66 | 97.66 | 97.50 |
| Veto | 99.84 | 99.80 | 99.95 | 98.69 | 62.54 | 40.35 | 40.65 | 40.55 | 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).
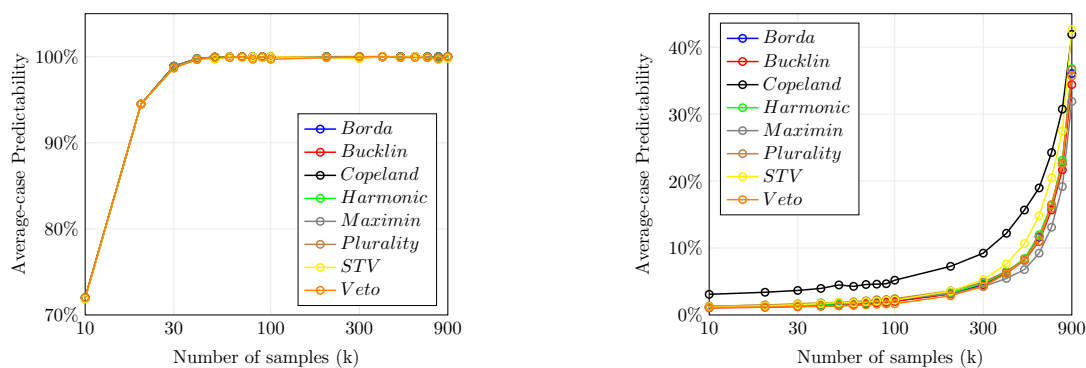


Figure 2: Average-case predictability of different voting rules as a function of the number of samples $k$, under the Mallows model with $\varphi = 1/3$ (left) and under the uniform distribution (right).

# 5 Experiments

In this section, we conduct experiments to measure the predictability of popular voting rules in the average case.[3] 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).[4] 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 [16] and to rank students in MOOCs [8]. But sometimes we may be interested in predicting only the top alternative. Could this lead to more optimistic results? The answer is yes and no. In the appendix, we show that the worst-case accuracy of predicting the Borda count winner using Borda count itself is still zero. But the case of the $r$-approval rule *is* more optimistic.

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 [21, 5], where the goal is to find an alternative with small distortion, how small can we make the expected distortion given only sampled votes?

---

[3]Refer to the book by Brandt et al. [6] for definitions.

[4]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).

# References

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

[2] Haris Aziz, Markus Brill, Felix Fischer, Paul Harrenstein, Jérôme Lang, and Hans Georg Seedig. Possible and necessary winners of partial tournaments. *Journal of Artificial Intelligence Research*, 54:493–534, 2015.

[3] Dorothea Baumeister, Piotr Faliszewski, Jérôme Lang, and Jörg Rothe. Campaigns for lazy voters: truncated ballots. In *Proceedings of the 11th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 577–584, 2012.

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

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

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

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

[8] Ioannis Caragiannis, Xenophon Chatzigeorgiou, George A Krimpas, and Alexandros A Voudouris. Optimizing positional scoring rules for rank aggregation. *Artificial Intelligence*, 267:58–77, 2019.

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

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

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

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

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

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

[15] Jérôme Lang, Maria Silvia Pini, Francesca Rossi, Domenico Salvagnin, Kristen Brent Venable, and Toby Walsh. Winner determination in voting trees with incomplete preferences and weighted votes. *Autonomous Agents and Multi-Agent Systems*, 25(1):130–157, 2012.

[16] Jonathan Levin and Barry Nalebuff. An introduction to vote-counting schemes. *Journal of Economic Perspectives*, 9(1):3–26, 1995.

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

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

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

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

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

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

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

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

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

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

# Appendix

## A    Winner Prediction

**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.*

*Proof.* We consider cases of even and odd $k$ and for each case we construct a profile on which Borda count cannot predict its own winner.

**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 bottom.

| $\frac{n-1}{2}$ votes | $a_1 \succ a_3 \succ a_2 \succ a_4 \succ \ldots$ |
|---|---|
| $\frac{n-1}{2}$ votes | $a_4 \succ a_2 \succ a_3 \succ a_1 \succ \ldots$ |
| 1 vote | $a_3 \succ a_1 \succ a_4 \succ a_2 \succ \ldots$ |

We denote with $\sigma_1$, $\sigma_2$ and $\sigma_3$ the ranking of the first, second and third rows, respectively. Under Borda rule, clearly winner is $a_3$.

We show that for every sample $\vec{\pi}_k \subset \vec{\sigma}_n$ of size $k$, $f_{\mathrm{Borda}}(\vec{\pi}_k) \neq \{a_3\}$. Let $x_1$, $x_2$, and $x_3$ denote the number of times $\sigma_1$, $\sigma_2$, and $\sigma_3$ appear in $\vec{\pi}_k$, respectively. Note that $x_3 \in \{0, 1\}$.

In order for $a_3$ to be the unique winner under $f_{\text{Borda}}$ on $\vec{\pi}_k$, it must defeat $a_1$. Hence, we need

$$x_1 \cdot (m-2) + x_2 \cdot (m-3) + x_3 \cdot (m-1)$$
$$> x_1 \cdot (m-1) + x_2 \cdot (m-4) + x_3 \cdot (m-2)$$
$$\Rightarrow x_2 + x_3 > x_1.$$

Similarly, $a_3$ must also defeat $a_4$. Hence, we also need

$$x_1 \cdot (m-2) + x_2 \cdot (m-3) + x_3 \cdot (m-1)$$
$$> x_1 \cdot (m-4) + x_2 \cdot (m-1) + x_3 \cdot (m-3)$$
$$\Rightarrow x_1 + x_3 > x_2.$$

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 $k$. Consider the following profile $\vec{\sigma}_n$.

| $\frac{n-1}{2}$ votes | $a_1 \succ a_3 \succ a_2 \succ a_4 \succ \ldots$ |
|---|---|
| $\frac{n-1}{2}$ votes | $a_4 \succ a_2 \succ a_3 \succ a_1 \succ \ldots$ |
| 1 vote | $a_3 \succ a_2 \succ a_1 \succ a_4 \succ \ldots$ |
| 1 vote | $a_4 \succ a_2 \succ a_1 \succ a_3 \succ \ldots$ |

We again denote with $\sigma_1$, $\sigma_2$, $\sigma_3$ and $\sigma_4$ the rankings in rows 1, 2, 3 and 4, respectively. When Borda count is applied on $\vec{\sigma}_n$, clearly $a_2$ is the unique winner.

We now argue that for every sample $\vec{\pi}_k \subset \vec{\sigma}_n$ of size $k$, $f_{\text{Borda}}(\vec{\pi}_k) \neq \{a_2\}$. Again, let $x_1$, $x_2$, $x_3$, and $x_4$ denote the number of times $\sigma_1$, $\sigma_2$, $\sigma_3$, and $\sigma_4$ appear in $\vec{\pi}_k$, respectively. Note that $x_3, x_4 \in \{0, 1\}$. In order for $a_2$ to be the unique winner under $f_{\text{Borda}}$ on $\vec{\pi}_k$, it must defeat $a_3$. Hence, we need

$$x_1 \cdot (m-3) + x_2 \cdot (m-2) + x_3 \cdot (m-2) + x_4 \cdot (m-2)$$
$$> x_1 \cdot (m-2) + x_2 \cdot (m-3) + x_3 \cdot (m-1) + x_4 \cdot (m-4)$$
$$\Rightarrow x_2 + 2 \cdot x_4 > x_1 + x_3. \tag{9}$$

Similarly, $a_2$ must also defeat $a_4$. Hence, we also need

$$x_1 \cdot (m-3) + x_2 \cdot (m-2) + x_3 \cdot (m-2) + x_4 \cdot (m-2)$$
$$> x_1 \cdot (m-4) + x_2 \cdot (m-1) + x_3 \cdot (m-4) + x_4 \cdot (m-1)$$
$$\Rightarrow x_1 + 2 \cdot x_3 > x_2 + x_4. \tag{10}$$

**Case 1:** $x_3 = x_4 = 0$. From Equation (9), we obtain $x_2 > x_1$, while from Equation (10), we obtain $x_1 > x_2$, which is clearly a contradiction.

**Case 2:** $x_3 = 1$ and $x_4 = 0$. From Equation (9), we obtain $x_2 > x_1 + 1$, while from Equation (10), we obtain $x_1 + 2 > x_2$, 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 $x_4 = 1$. From Equation (9) we obtain that $x_2 + 2 > x_1 + 1$, while from Equation (10) we obtain $x_1 + 2 > x_2 + 1$. This means that $x_1 = x_2$, which implies $|\vec{\pi}_k| = x_1 + x_2 + 2$ is even, which in turn contradicts the fact that $k$ is odd. □

**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.*

*Proof.* Fix $r \in [m-1]$ and $k \geq r$. Consider any profile $\vec{\sigma}_n$. Let $W = f_{r\text{-app}}(\vec{\sigma}_n)$ denote the set of winners under $f_{r\text{-app}}$ on $\vec{\sigma}_n$, and let $\bar{W} = A \setminus W$. Let $x_i$ denote the $r$-approval score of $a_i$ in $\vec{\sigma}_n$ for each $i \in [m]$.

We now show that there exists a sample $\vec{\pi}_k \in \mathcal{S}_k(\vec{\sigma}_n)$ such that $f_{r\text{-app}}(\vec{\pi}_k) \subseteq f_{r\text{-app}}(\vec{\sigma}_n)$. Without loss of generality, assume that $a_1 \in W$. Hence, we have $x_1 > x_j$ for all $a_j \in \bar{W}$.

Let $\vec{\rho} \subseteq \vec{\sigma}_n$ denote the subset of votes in which $a_1$ is approved (i.e. it appears in the top $r$ positions). If $k \geq |\vec{\rho}|$, then we can simply choose an arbitrary $\vec{\pi}_k \in \mathcal{S}_k(\vec{\sigma}_n)$ such that $\vec{\rho} \subseteq \vec{\pi}_k$. In this case, the score of $a_1$ would be exactly equal to $x_1 = |\vec{\rho}|$, while the score of any $a_j \in \bar{W}$ will be at most $x_j < x_1$. This implies $f_{r\text{-app}}(\vec{\pi}_k) \subseteq W$, as desired.

Now, suppose $r \leq k \leq |\vec{\rho}|$. Note that the previous argument establishes that for $k = |\vec{\rho}|$, we can use $\vec{\pi}_k = \vec{\rho}$, and have $f_{r\text{-app}}(\vec{\pi}_k) \subseteq W$. For $k < |\vec{\rho}|$, we show that one can start from $\vec{\rho}$, and iteratively remove one vote at a time — until $k$ votes remain — such that the set of winners is always a subset of $W$.

More specifically, we show that given $k \geq r$, and any $\vec{\pi}_{k+1} \subseteq \vec{\rho}$ with $|\vec{\pi}_{k+1}| = k+1$ and $f_{r\text{-app}}(\vec{\pi}_{k+1}) \subseteq W$, there exists $\vec{\pi}_k \subset \vec{\pi}_{k+1}$ with $|\vec{\pi}_k| = k$ and $f_{r\text{-app}}(\vec{\pi}_k) \subseteq W$.

Suppose this is false for some $\vec{\pi}_{k+1}$. That is, removing any single ranking from $\vec{\pi}_{k+1}$ results in at least one alternative from $\bar{W}$ becoming a winner. For each $\sigma \in \vec{\pi}_{k+1}$, let us denote $\vec{\pi}_k^{\sigma} = \vec{\pi}_{k+1} \setminus \{\sigma\}$ and $S(\sigma) = f_{r\text{-app}}(\vec{\pi}_k^{\sigma}) \cap \bar{W}$. Then, our assumption implies that $S(\sigma) \neq \emptyset$ for each $\sigma \in \vec{\pi}_{k+1}$.

We further show that $S(\sigma) \cap S(\sigma') = \emptyset$ for distinct $\sigma, \sigma' \in \vec{\pi}_{k+1}$.[5] Note that for each $a_j \in S(\sigma)$, $a_j$ is a winner under $\vec{\pi}_k^{\sigma}$ but not under $\vec{\pi}_{k+1}$. Also, note that the $r$-approval score of $a_1$ under $\vec{\pi}_k^{\sigma}$ and $\vec{\pi}_{k+1}$ is $k$ and $k+1$, respectively, since $\vec{\pi}_{k+1} \subseteq \vec{\rho}$. Hence, it must be the case that $a_j$ is approved in every ranking in $\vec{\pi}_k^{\sigma}$ but not in $\sigma$. Hence, $a_j \notin S(\sigma')$ for every other ranking $\sigma'$ in $\vec{\pi}_{k+1}$.

Finally, let us partition $\bar{W}$ into two sets $S_{\text{win}}$ and $S_{\text{lose}}$ such that $S_{\text{win}} = \cup_{\sigma \in \vec{\pi}_{k+1}} S(\sigma)$ and $S_{\text{lose}} = \bar{W} \setminus S_{\text{win}}$. Let $t = \min_{\sigma \in \vec{\pi}_{k+1}} |S(\sigma)|$. Then, $|S_{\text{win}}| \geq t \cdot (k+1)$. Further, consider a ranking $\sigma \in \vec{\pi}_{k+1}$ under which exactly $t$ of the unapproved alternatives are in $S_{\text{win}}$. In this ranking, the remaining $m - r - t$ unapproved alternatives must be in $S_{\text{lose}}$. Hence, $|S_{\text{lose}}| \geq m - r - t$. Thus, we have $\bar{W} \geq t \cdot (k+1) + m - r - t$. However, since $a_1 \in W$, we also have $\bar{W} \leq m - 1$. Combining the two inequalities, we get

$$t \cdot (k+1) + m - r - t \leq m - 1 \Rightarrow t \cdot k \leq r - 1$$
$$\Rightarrow k \leq r - 1,$$

where the last inequality holds because $t \geq 1$ (since $S(\sigma) \neq \emptyset$ for each $\sigma \in \vec{\pi}_{k+1}$). However, this is a contradiction since we assumed $k \geq r$. $\square$

It is easy to notice that the lower bound of $r$ is tight. For example, consider the following preference profile with $m$ alternatives and $r$ votes.

| Approved | Not Approved |
|---|---|
| $a_1 \succ a_2 \succ a_3 \succ \ldots \succ a_r$ | $a_{r+1} \succ a_{r+2} \succ \ldots \succ a_m$ |
| $a_1 \succ a_{r+1} \succ a_2 \succ \ldots \succ a_{r-1}$ | $a_r \succ a_{r+2} \succ \ldots \succ a_m$ |
| $\vdots$ | $\vdots$ |
| $a_1 \succ a_3 \succ a_4 \succ \ldots \succ a_{r+1}$ | $a_2 \succ a_{r+2} \succ \ldots \succ a_m$ |

Basically, alternative $a_1$ is approved in each vote, and in each of the $r$ votes, a distinct subset of $r-1$ alternatives from the set $\{a_2, \ldots, a_{r+1}\}$ of $r$ alternatives is approved. It is easy to check that $a_1$ is the unique $r$-approval winner on this profile. However, there is no subset of $r-1$ (or fewer) votes under which $a_1$ is the unique $r$-approval winner.

---

[5] Here, even two copies of the same ranking are referred to as distinct.