X

When Do Noisy Votes Reveal the Truth?

IOANNIS CARAGIANNIS, University of Patras & CTI
ARIEL D. PROCACCIA, Carnegie Mellon University
NISARG SHAH, Carnegie Mellon University

A well-studied approach to the design of voting rules views them as maximum likelihood estimators; given votes that are seen as noisy estimates of a true ranking of the alternatives, the rule must reconstruct the most likely true ranking. We argue that this is too stringent a requirement, and instead ask: *How many* votes does a voting rule need to reconstruct the true ranking? We define the family of *pairwise-majority consistent* rules, and show that for all rules in this family the number of samples required from the Mallows noise model is logarithmic in the number of alternatives, and that no rule can do asymptotically better (while some rules like plurality do much worse). Taking a more normative point of view, we consider voting rules that surely return the true ranking as the number of samples tends to infinity (we call this property *accuracy in the limit*); this allows us to move to a higher level of abstraction. We study families of noise models that are parametrized by distance functions, and find voting rules that are accurate in the limit for all noise models in such general families. We characterize the distance functions that induce noise models for which pairwise-majority consistent rules are accurate in the limit, and provide a similar result for another novel family of *position-dominance consistent* rules. These characterizations capture three well-known distance functions.

1. INTRODUCTION

Social choice theory studies the aggregation of individual preferences towards a collective choice. In one of the most common models, both the individual preferences and the collective decision are represented as rankings of the alternatives. A *voting rule*[1] takes the individual rankings as input and outputs a social ranking.

One can imagine many different voting rules; which are better than others? The popular *axiomatic* approach suggests that the best voting rules are the ones that satisfy intuitive social choice axioms. For example, if we replicate the votes, the outcome should not change; or, if each and every voter prefers one alternative to another, the social ranking should follow suit. It is well-known though that natural combinations of axioms are impossible to achieve [Arrow 1951], hence the axiomatic approach cannot give a crisp answer to the above question.

A different — in a sense competing — approach views voting rules as *estimators*. From this viewpoint, some alternatives are objectively better than others, i.e., the votes are simply noisy estimates of an underlying ground truth. One voting rule is therefore better than another if it is more likely to output the true underlying ranking; the best voting rule is a *maximum likelihood estimator (MLE)* of the true ranking. This approach dates back to Marquis de Condorcet, who also proposed a compellingly simple noise model: each voter ranks each pair of alternatives correctly with probability $p > 1/2$ and incorrectly with probability $1 - p$, and the mistakes are i.i.d.[2] Today this noise model is typically named after Mallows [1957]. Probability theory was still in its infancy in the 18th Century (in fact Condorcet was one of its pioneers), so the maximum likelihood estimator in the Mallows model — the *Kemeny* rule — had to wait another two centuries to receive due recognition [Young 1988]. More recently, the

[1] More formally known in this context as a *social welfare function*.
[2] Intuitively, if a ranking is not obtained because of cycle formation, the process is restarted.

Authors' addresses: I. Caragiannis, Computer Technology Institute and Dept. of Computer Engineering and Informatics, University of Patras, caragian@ceid.upatras.gr; A. D. Procaccia and N. Shah, Computer Science Department, Carnegie Mellon University, {arielpro,nkshah}@cs.cmu.edu.

MLE approach has received some attention in computer science [Conitzer and Sandholm 2005; Conitzer et al. 2009; Elkind et al. 2010b; Procaccia et al. 2012; Mao et al. 2013], in part because its main prerequisite (underlying true ranking) is naturally satisfied by some of the crowdsourcing and human computation domains, where voting is in fact commonly used [Procaccia et al. 2012; Mao et al. 2013].

As compelling as the MLE approach is, there are many different considerations in choosing a voting rule, and insisting that the voting rule be an MLE is a tall order (there is only one MLE per noise model); this is reflected in existing negative results [Conitzer and Sandholm 2005; Elkind et al. 2010b]. We relax this requirement by asking: *How many* votes do prominent voting rules need to recover the true ranking with high probability? In crowdsourcing tasks, for example, the required number of votes directly translates to the amount of time and money one must spend to obtain accurate results. Taking one step further and adopting a more normative viewpoint, we ask: Which voting rules are guaranteed to return the correct ranking given an *infinite* number of samples from Mallows' model? Finally, at the highest level of abstraction we consider general classes of noise models, and seek similar guarantees with respect to any noise model in one of these classes.

1.1. Our contribution

In Section 3 we focus on the Mallows model. We define the class of *pairwise-majority consistent (PM-c)* rules. Intuitively, if there is a ranking σ of the alternatives such that for every pair of alternatives a majority of voters agree with σ on their comparison then a PM-c rule must return σ. The Kemeny rule is a PM-c rule, and so are several other prominent voting rules. Our main result for this section is that to output the true ranking with probability $1-\epsilon$ any PM-c rule requires only a logarithmic number of samples in $1/\epsilon$ and m, where m is the number of alternatives. We also establish a matching lower bound that holds for any voting rule. Among other results, we show that a similar bound does not hold for the plurality rule — the most ubiquitous among voting rules — and indeed it requires an exponential number of samples.

Section 4 is an interlude of sorts. Instead of quantifying the required number of samples, we consider a relaxed guarantee that we call *accuracy in the limit*: a voting rule should return the correct ranking given an infinite number of samples. We view this as a *normative* property, and in this sense we are connecting the axiomatic approach with the estimation approach. In the Mallows model accuracy in the limit is easy to satisfy. Clearly, it is satisfied by all PM-c rules in light of the abovementioned result, but we also show that it is satisfied by all rules that belong to another novel class — *position-dominance consistent (PD-c)* rules. Roughly speaking, PD-c rules focus on the exact positions in which alternatives appear in the individual rankings, rather than pairwise comparisons, and are disjoint from PM-c rules. We show that all PD-c rules are also accurate in the limit under the Mallows model. While we view accuracy in the limit as a normative constraint, asking for a voting rule to be accurate in the limit only for the Mallows model is perhaps asking too little. In the Mallows model the probability of a ranking decreases, but in a specific way (exponentially), as its Kendall-Tau (KT) distance from the true ranking increases; this distance function measures the number of disagreements on pairs of alternatives. We want the voting rules to be accurate in the limit with respect to any noise model that is similarly monotonic with respect to the KT distance, and show that this is indeed the case with respect to all PM-c and PD-c rules.

At the highest level of abstraction, we wish to extend our results to noise models that are derived from a variety of distance functions. We define the family of *majority-concentric (MC)* distances and prove the following characterization result: All PM-c rules are accurate in the limit with respect to any noise model that is monotonic with respect to a distance function d if and only if d is MC. Similarly, we define the family of *position-concentric (PC)* distances and prove an analogous results for PD-c rules and PC distances. To verify that these results are indeed very general, we prove that three popular distance functions are both MC and PC.

1.2. Related work

The theme of quantifying the number of samples that are required to uncover the truth plays a central role in a recent paper by Chierichetti and Kleinberg [2012]. They study a setting with a single correct alternative and noisy signals about its identity. Focusing on a single voting rule — the plurality rule — they give an upper bound on the number of votes that are required to pinpoint the correct winner. They also prove a lower bound that applies to any voting rule and suggests that plurality is not far from optimal. Interestingly, under the Mallows model we show that plurality is far worse than all PM-c rules, but note that we consider rules that output a ranking while Chierichetti and Kleinberg [2012] study rules that output a single winner.

Our initial results regarding the Kemeny rule are related to the work of Braverman and Mossel [2008]. Given samples from the Mallows model, they aim to compute the Kemeny ranking; this problem is known to be NP-hard. They focus on circumventing the complexity barrier by giving an efficient algorithm that computes the Kemeny ranking with arbitrarily high probability. In contrast, we ask: How many samples do PM-c rules (including Kemeny) need to reconstruct the *true* ranking?

There is a significant body of literature on MLEs and parameter estimation for noise models over rankings that generalize Mallows' model [Fligner and Verducci 1986; Critchlow et al. 1991; Meilă et al. 2012; Lebanon and Lafferty 2002; Lu and Boutilier 2011]. In particular, the classic paper by Fligner and Verducci [1986] analyzes extensions of the Mallows model with distance functions from two families: those that are based on discordant pairs (including the KT distance) and those that are based on cyclic structure. Critchlow et al. [1991] introduce four categories of noise models; they also define desirable axiomatic properties that noise models should satisfy, and determine which properties are satisfied by the different categories.

Somewhat further afield, a recent line of work in computational social choice studies the *distance rationalizability* of voting rules [Meskanen and Nurmi 2008; Elkind et al. 2009, 2010a,b; Boutilier and Procaccia 2012]. Voting rules are said to be distance rationalizable if they always select an alternative or a ranking that is "closest" to being a consensus winner, under some notion of distance and some notion of consensus. Among these papers, the one by Elkind et al. [2010b] is the most closely related to our work; they observe that the Kemeny rule is both an MLE and distance rationalizable, and ask whether at least one of several other common rules has the same property (the answer is "no").

2. PRELIMINARIES

We consider a set A of m alternatives. Let $\mathcal{L}(A)$ be the set of votes (which we may think of as rankings or permutations), where each vote is a bijection $\sigma : A \to \{1, 2, \ldots, m\}$. Hence, $\sigma(a)$ is the position of alternative a in σ. In particular, $\sigma(a) < \sigma(b)$ denotes that a is preferred to b under σ; we alternatively denote this by $a \succ_\sigma b$. A *vote profile* (or simply *profile*) $\pi \in \mathcal{L}(A)^n$ consists of a set of n votes for some $n \in \mathbb{N}$.

2.1. Voting rules

A *deterministic voting rule* is a function $r : \cup_{n \geq 1} \mathcal{L}(A)^n \to \mathcal{L}(A)$ which operates on a vote profile and outputs a ranking. First, note that we define the voting rule to output a ranking over alternatives rather than a single alternative; such functions are also known as *social welfare functions* in the literature. Second, in contrast to the traditional notation, we define a voting rule to operate on any number of votes, which is required to analyze its asymptotic properties as the number of votes grows. We consider *randomized voting rules* which are denoted by $r : \cup_{n\geq 1}\mathcal{L}(A)^n \to D(\mathcal{L}(A))$ where $D(\cdot)$ denotes the set of all distributions over an outcome space. We use $\Pr[r(\pi) = \sigma]$ to denote the probability of rule r returning ranking σ given profile π. The following voting rules (or families of voting rules) play a key role in the paper.

(Positional) Scoring Rules. A scoring rule is given by a scoring vector $\alpha = (\alpha_1, \ldots, \alpha_m)$ where $\alpha_i \geq \alpha_{i+1}$ for all $i \in \{1, \ldots, m\}$ and $\alpha_1 > \alpha_m$. Under this rule for each vote σ and $i \in \{1, \ldots, m\}$, α_i points are awarded to the alternative $\sigma^{-1}(i)$, that is, α_1 points to the first alternative, α_2 points to the second alternative, and so on. The alternative with the most points overall is selected as the winner. We naturally extend this to output the ranking where alternatives are sorted in the descending order of their total points. Our results on positional scoring rules hold irrespective of the tie-breaking rule used. Special scoring rules include *plurality* with $\alpha = (1, 0, 0, \ldots, 0)$, *Borda count* with $\alpha = (m, m-1, \ldots, 1)$, the *veto* rule with $\alpha = (1, 1, \ldots, 1, 0)$, and the *harmonic* rule [Boutilier et al. 2012] with $\alpha = (1, 1/2, \ldots, 1/m)$.

The Kemeny Rule. Given a profile $\pi = (\sigma_1, \ldots, \sigma_n) \in \mathcal{L}(A)^n$, the Kemeny rule selects a ranking $\sigma \in \mathcal{L}(A)$ that minimizes $\sum_{i=1}^{n} d_{KT}(\sigma, \sigma_i)$, where d_{KT} is the *Kendall tau (KT)* distance defined as

$$d_{KT}(\sigma_1, \sigma_2) = |\{(a, b) | ((a \succ_{\sigma_1} b) \wedge (b \succ_{\sigma_2} a)) \vee ((b \succ_{\sigma_1} a) \wedge (a \succ_{\sigma_2} b))\}|.$$

In words, the KT distance between two rankings is their number of disagreements over pairs of alternatives, and informally it is equal to the minimum number of adjacent swaps required to convert one ranking into the other. We give special attention to the *Kemeny rule with uniform tie-breaking* — the randomized version of the Kemeny rule where ties are broken uniformly, i.e., each ranking in $\arg\min_{\sigma\in\mathcal{L}(A)} \sum_{i=1}^{n} d_{KT}(\sigma, \sigma_i)$ is returned with equal probability.

2.2. Noise models and distances

We assume that there exists a true hidden order $\sigma^* \in \mathcal{L}(A)$ over the alternatives. We denote the alternative at position i in σ^* by a_i, i.e., $\sigma^*(a_i) = i$.

Our noise models are parametrized by distance functions over rankings. A function $d : \mathcal{L}(A) \times \mathcal{L}(A) \to \mathbb{R}_{\geq 0}$ is called a distance function if for every $\sigma, \sigma', \tau \in \mathcal{L}(A)$ it satisfies: (1) $d(\sigma, \sigma') \geq 0$, (2) $d(\sigma, \sigma') = 0$ if and only if $\sigma = \sigma'$, (3) $d(\sigma, \sigma') = d(\sigma', \sigma)$, and (4) $d(\sigma, \sigma') \leq d(\sigma, \tau) + d(\tau, \sigma')$. We assume that our distance functions are right-invariant: the distance between any two rankings does not change if the alternatives are relabeled, which is a standard assumption. A right-invariant distance function is fully specified by the distances of all rankings from any single base ranking.

We consider three popular distance functions in this paper: the Kendall tau (KT) distance (which we have defined above), the *footrule* distance, and the *maximum displacement* distance. We investigate the KT distance in detail in Section 3. Definitions of the other distance functions are given in Appendix E.

A noise model defines the probability of observing a ranking given an underlying true ranking, i.e., $\Pr[\sigma|\sigma^*]$ for all $\sigma, \sigma^* \in \mathcal{L}(A)$. In Section 3, we focus on a particular noise model, known as the *Mallows model* [Mallows 1957], which is widely used in machine learning and statistics. In this model, a ranking is generated given the true ranking σ^* as follows. When two alternatives a and b with $a \succ_{\sigma^*} b$ are compared, the outcome is consistent with the true ranking, i.e., $a \succ b$, with a fixed probability $1/2 < p < 1$. Every two alternatives are compared in this manner, and the process is restarted if the generated vote has a cycle (e.g., $a \succ b \succ c \succ a$). It is easy to check that the probability of drawing a ranking σ, given that the true order is σ^*, is proportional to

$$p^{\binom{m}{2}-d_{KT}(\sigma,\sigma^*)} \cdot (1-p)^{d_{KT}(\sigma,\sigma^*)},$$

which upon normalization gives

$$\Pr[\sigma|\sigma^*] = \frac{\varphi^{d_{KT}(\sigma,\sigma^*)}}{Z_\varphi^m},$$

where $\varphi = (1-p)/p < 1$ and Z_φ^m is the normalization factor which is independent of the true ranking σ^* (see, e.g., [Lu and Boutilier 2011]). We denote by $p_{i,j}$ the probability that the alternative at position i in the true ranking (a_i) appears in position j in a random vote , so

$$p_{i,j} = \sum_{\sigma \in \mathcal{L}(A) | \sigma(a_i) = j} \Pr[\sigma|\sigma^*].$$

Let $q_{i,k} = \sum_{j=1}^{k} p_{i,j}$. Votes are sampled independently, so the probability of observing a profile $\pi = (\sigma_1, \ldots, \sigma_n) \in \mathcal{L}(A)^n$ is $\Pr[\pi|\sigma^*] = \prod_{i=1}^{n} \Pr[\sigma_i|\sigma^*]$. We note that this model is equivalent to the *Condorcet noise model*.

3. SAMPLE COMPLEXITY IN MALLOWS' MODEL

We first consider the Mallows model and analyze the number of samples needed by different voting rules to determine the true ranking with high probability; we use this *sample complexity* as a criterion to distinguish between voting rules or families of voting rules. For any (randomized) voting rule r, integer $k \in \mathbb{N}$ and ranking $\sigma \in \mathcal{L}(A)$, let $\mathrm{Acc}^r(k,\sigma) = \sum_{\pi \in \mathcal{L}(A)^k} \Pr[\pi|\sigma] \cdot \Pr[r(\pi) = \sigma]$ denote the accuracy of rule r with k samples and true ranking σ, that is, the probability that rule r returns σ given k samples from Mallows' model with true ranking σ. We overload the notation by letting $\mathrm{Acc}^r(k) = \min_{\sigma \in \mathcal{L}(A)} \mathrm{Acc}^r(k,\sigma)$. In words, given k samples from Mallows' model, rule r returns the underlying true ranking with probability at least $\mathrm{Acc}^r(k)$ irrespective of what the true ranking is. Finally, we denote $N^r(\epsilon) = \min\{k \mid \mathrm{Acc}^r(k) \geq 1-\epsilon\}$, which is the number of samples required by rule r to return the true ranking with probability at least $1-\epsilon$. Informally, we call $N^r(\epsilon)$ the sample complexity of rule r.

We begin by showing that for any number of alternatives m and any accuracy level ϵ, the Kemeny rule (with uniform tie-breaking) requires the minimum number of samples from Mallows' model to determine the true ranking with probability at least $1-\epsilon$. It is already known that the Kemeny rule is the maximum likelihood estimator (MLE) for the true ranking given samples from Mallows' model. Formally, given a profile $\pi = (\sigma_1, \ldots, \sigma_n)$ from Mallows' model, the MLE estimator of the true ranking is

$$\arg\max_{\sigma \in \mathcal{L}(A)} \Pr[\pi|\sigma] = \arg\max_{\sigma \in \mathcal{L}(A)} \prod_{i=1}^{n} \frac{\varphi^{-d_{KT}(\sigma_i,\sigma)}}{Z_\varphi^m} = \arg\min_{\sigma \in \mathcal{L}(A)} \sum_{i=1}^{n} d_{KT}(\sigma_i, \sigma),$$

where the expression on the right hand side is a Kemeny ranking. While at first glance it may seem that this directly implies optimal sample complexity of the Kemeny rule, we give an example in Appendix A of a noise model where the MLE rule does not have optimal sample complexity. However, we show that for the Mallows model, the Kemeny rule is optimal in terms of sample complexity. The proof is given in Appendix A.

THEOREM 3.1. *The Kemeny rule with uniform tie-breaking has the optimal sample complexity in Mallows' model, that is, for any number of alternatives* m *and any* $\epsilon > 0$*,* $N^{\mathrm{KEM}}(\epsilon) \leq N^r(\epsilon)$ *for every (randomized) voting rule* r.

Now that we know that the Kemeny rule has the optimal sample complexity, a natural question is to determine how many samples it really requires. Instead of analyzing the sample complexity of the Kemeny rule particularly, we consider a family of voting rules (which includes the Kemeny rule itself) such that each rule in this family has the same asymptotic sample complexity as that of the Kemeny rule.

3.1. The family of PM-c rules

Our family of voting rules crucially relies on the standard concept of *pairwise-majority graph* (PM graph). Given a profile $\pi \in \mathcal{L}(A)^n$, the PM graph of π is the directed graph $G = (V, E)$, where the alternatives are the vertices ($V = A$) and there is an edge from alternative a to alternative a' if a is preferred to a' in a (strong) majority of the rankings of π. Formally, $(a, a') \in E$ if $|\{\sigma \in \pi | a \succ_\sigma a'\}| > |\{\sigma \in \pi | a' \succ_\sigma a\}|$. Note that there may be pairs of alternatives such that there is no edge in the PM graph in either direction (if they are tied), but there can never be an edge in both directions. A PM graph can also have directed cycles. When a PM graph is complete (i.e., there is an edge between every pair of alternatives) and acyclic, there exists a unique $\sigma \in \mathcal{L}(A)$ such that there is an edge from a to a' if and only if $a \succ_\sigma a'$. In this case, we say that the PM graph reduces to σ.

Definition 3.2 (*Pairwise-Majority Consistent Rules*). A deterministic voting rule r is called pairwise-majority consistent (PM-c) if $r(\pi) = \sigma$ whenever the PM graph of π reduces to σ. For randomized voting rules, we require that $\Pr[r(\pi) = \sigma] = 1$.

To the best of our knowledge this family of rules is novel. Note though that an acyclic and complete PM graph is similar to — and in some sense an extension of — having a Condorcet winner. A Condorcet winner is an alternative that beats every other alternative in a pairwise election. It is easy to check that if such an alternative exists, then it is unique and it is a source in the PM graph with $m - 1$ outgoing edges and no incoming edges. Thus, profiles where the PM graph reduces to a ranking necessarily have a Condorcet winner. In addition, they have a second alternative with $m - 2$ outgoing edges and only 1 incoming edge, a third alternative with $m - 3$ outgoing edges and 2 incoming edges, and so on.

THEOREM 3.3. *The Kemeny rule, the ranked pairs method, Copeland's method, and Schulze's method are PM-c.*

The definitions of these rules and the proof of the theorem appear in Appendix C. Note that all the rules in Theorem 3.3 are Condorcet consistent when they output a single alternative. If we take any Condorcet consistent method, apply it on a profile, remove the winner from every vote in the profile, apply the method again on the reduced profile, and keep repeating this process, then the extended rule that outputs the alternatives in the order of removal is always a PM-c rule. In contrast, Copeland's method in Theorem 3.3 is extended by outputting a ranking where the alternatives are sorted by their Copeland scores.

We now proceed to prove that any PM-c rule requires at most a logarithmic number of samples in m (the number of alternatives) and $1/\epsilon$ to determine the true ranking with probability at least $1 - \epsilon$. First, we introduce a property of distance functions that will be used throughout the paper. For any $\sigma \in \mathcal{L}(A)$ and $a, b \in A$, define $\sigma_{a \leftrightarrow b}$ to be the ranking obtained by swapping a and b in σ. That is, $\sigma_{a \leftrightarrow b}(c) = \sigma(c)$ for any $c \in A \setminus \{a, b\}$, $\sigma_{a \leftrightarrow b}(a) = \sigma(b)$ and $\sigma_{a \leftrightarrow b}(b) = \sigma(a)$.

Definition 3.4 (*Swap-Increasing Distance Functions*). An integer-valued distance function d is called swap-increasing if for any $\sigma^*, \sigma \in \mathcal{L}(A)$ and alternatives $a, b \in A$

such that $a \succ_{\sigma^*} b$ and $a \succ_\sigma b$, we have $d(\sigma_{a\leftrightarrow b}, \sigma^*) \geq d(\sigma, \sigma^*)+1$, and if $\sigma^*(b) = \sigma^*(a)+1$ (a and b are adjacent in σ^*) then $d(\sigma_{a\leftrightarrow b}, \sigma^*) = d(\sigma, \sigma^*)+1$.

The following lemma is a folklore result; we reconstruct its proof in Appendix E for the sake of completeness.

LEMMA 3.5. *The Kendall tau (KT) distance is swap-increasing.*

We are now ready to analyze the sample complexity of PM-c rules.

THEOREM 3.6. *For any given $\epsilon > 0$, any PM-c rule determines the true ranking with probability at least $1-\epsilon$ given $O(\log(m/\epsilon))$ samples from Mallows' model.*

PROOF. Let σ^* denote the true underlying ranking. We show that the PM graph of a profile of $O(\log(m/\epsilon))$ votes from Mallows' model reduces to σ^* with probability at least $1-\epsilon$. It follows that any PM-c rule would output σ^* with probability at least $1-\epsilon$.

Let $\pi \in \mathcal{L}(A)^n$ denote a profile of n samples from Mallows' model. For any $a, b \in A$, let n_{ab} denote the number of rankings $\sigma \in \pi$ such that $a \succ_\sigma b$. Hence, $n_{ab}+n_{ba} = n$ for every $a, b \in A$. The PM graph of π reduces to σ^* if and only if for every $a, b \in A$ such that $a \succ_{\sigma^*} b$, we have $n_{ab} - n_{ba} \geq 1$. Hence, we want an upper bound on n such that

$$\Pr\left[\forall a, b \in A, a \succ_{\sigma^*} b \Rightarrow n_{ab} - n_{ba} \geq 1\right] \geq 1-\epsilon.$$

For any $a, b \in A$ with $a \succ_{\sigma^*} b$, define $\delta_{ab} = \mathbb{E}[(n_{ab}-n_{ba})/n]$. Let $p_{a\succ b}$ denote the probability that $a \succ_\sigma b$ in a random sample σ. Then, by linearity of expectation, we have $\delta_{ab} = p_{a\succ b} - p_{b\succ a}$. Thus,

$$\begin{aligned}\Pr\left[n_{ab}-n_{ba} \leq 0\right] = \Pr\left[\frac{n_{ab}-n_{ba}}{n} \leq 0\right] &\leq \Pr\left[\left|\frac{n_{ab}-n_{ba}}{n} - \mathbb{E}\left[\frac{n_{ab}-n_{ba}}{n}\right]\right| \geq \delta_{ab}\right] \\ &\leq 2\cdot e^{-2\cdot\delta_{ab}^2\cdot n} \leq 2\cdot e^{-2\cdot\delta_{\min}^2\cdot n},\end{aligned}$$

where the third transition is due to Hoeffding's inequality and in the last transition $\delta_{\min} = \min_{a,b\in A: a\succ_{\sigma^*} b} \delta_{ab}$. Applying the union bound, we get

$$\Pr\left[\exists a, b \in A,\ \{(a \succ_{\sigma^*} b) \wedge (n_{ab}-n_{ba} \leq 0)\}\right] \leq \binom{m}{2}\cdot 2\cdot e^{-2\cdot\delta_{\min}^2\cdot n} \leq m^2\cdot e^{-2\cdot\delta_{\min}^2\cdot n}$$

It is easy to check that the right-most quantity above is at most ϵ when $n \geq \frac{1}{2\cdot\delta_{\min}^2}\cdot\log\left(\frac{m^2}{\epsilon}\right)$. To complete the proof we need to show that $\delta_{\min} = \Omega(1)$, that is, it is lower bounded by a constant independent of m. This is quite intuitive since the process of generating a sample from Mallows' model maintains the order between every pair of alternatives with a constant probability $p > 1/2$. However, the fact that we restart the process if a cycle is formed makes the probabilities as well as this analysis a bit more involved. For any $a, b \in A$ such that $a \succ_{\sigma^*} b$, we have

$$\begin{aligned}\delta_{ab} = p_{a\succ b} - p_{b\succ a} &= \sum_{\sigma\in L(A)|a\succ_\sigma b} \Pr[\sigma|\sigma^*] - \sum_{\sigma\in L(A)|b\succ_\sigma a} \Pr[\sigma|\sigma^*] \\ &= \sum_{\sigma\in L(A)|a\succ_\sigma b} \left(\Pr[\sigma|\sigma^*] - \Pr[\sigma_{a\leftrightarrow b}|\sigma^*]\right) = \sum_{\sigma\in L(A)|a\succ_\sigma b} \frac{\varphi^{d_{KT}(\sigma,\sigma^*)} - \varphi^{d_{KT}(\sigma_{a\leftrightarrow b},\sigma^*)}}{Z_\varphi^m} \\ &\geq \sum_{\sigma\in L(A)|a\succ_\sigma b} \frac{\varphi^{d_{KT}(\sigma,\sigma^*)}\cdot(1-\varphi)}{Z_\varphi^m} = (1-\varphi)\cdot p_{a\succ b} = (1-\varphi)\cdot\left(\frac{1+\delta_{ab}}{2}\right),\end{aligned} \tag{1}$$

where the third transition follows since $\sigma \leftrightarrow \sigma_{a\leftrightarrow b}$ is a bijection between all rankings where $a \succ b$ and all rankings where $b \succ a$, the fifth transition follows using $\varphi < 1$

and Lemma 3.5, and the last transition follows by the equalities $p_{a\succ b} - p_{b\succ a} = \delta_{ab}$ and $p_{a\succ b} + p_{b\succ a} = 1$. Solving Equation (1), we get $\delta_{ab} \geq (1-\varphi)/(1+\varphi)$ for all $a, b \in A$ with $a \succ_{\sigma^*} b$. Hence, $\delta_{\min} \geq (1-\varphi)/(1+\varphi)$, as required. □ (Theorem 3.6)

We have seen that PM-c rules have logarithmic sample complexity; it turns out that no rule can do better, i.e., we prove a matching lower bound that holds for any randomized voting rule.

THEOREM 3.7. *For any $\epsilon \in (0, 1/2]$, any (randomized) voting rule requires $\Omega(\log(m/\epsilon))$ samples from Mallows' model to determine the true ranking with probability at least $1-\epsilon$.*

PROOF. Consider any voting rule r. Assume $\mathrm{Acc}^r(n) \geq 1-\epsilon$ for some $n \in \mathbb{N}$. We want to show that $n = \Omega(\log(m/\epsilon))$. For any $\sigma \in \mathcal{L}(A)$, we have $\mathrm{Acc}^r(n,\sigma) \geq 1-\epsilon$. Consider an arbitrary $\sigma \in \mathcal{L}(A)$, and let $\mathcal{N}(\sigma) = \{\sigma' \in \mathcal{L}(A) | d_{KT}(\sigma',\sigma) = 1\}$ denote the set of all rankings at distance 1 from σ. Then, for any ranking $\sigma' \in \mathcal{N}(\sigma)$ and any profile $\pi = (\sigma_1, \ldots, \sigma_n) \in \mathcal{L}(A)^n$, we have

$$\Pr[\pi|\sigma] = \prod_{i=1}^{n} \frac{\varphi^{d_{KT}(\sigma_i,\sigma)}}{Z_\varphi^m} \geq \prod_{i=1}^{n} \frac{\varphi^{d_{KT}(\sigma_i,\sigma')+1}}{Z_\varphi^m} = \varphi^n \cdot \Pr[\pi|\sigma'], \tag{2}$$

where the second transition holds since for any $\tau \in \mathcal{L}(A)$,

$$d_{KT}(\tau,\sigma) \leq d_{KT}(\tau,\sigma') + d_{KT}(\sigma,\sigma') = d_{KT}(\tau,\sigma') + 1$$

due to triangle inequality of distance functions. Now,

$$\begin{aligned}
\mathrm{Acc}^r(n,\sigma) &= \sum_{\pi\in\mathcal{L}(A)^n} \Pr[\pi|\sigma]\cdot\Pr[r(\pi)=\sigma] = \sum_{\pi\in\mathcal{L}(A)^n} \Pr[\pi|\sigma]\cdot(1-\Pr[r(\pi)\neq\sigma]) \\
&= 1 - \sum_{\pi\in\mathcal{L}(A)^n} \Pr[\pi|\sigma]\cdot\Pr[r(\pi)\neq\sigma] \\
&\leq 1 - \sum_{\pi\in\mathcal{L}(A)^n} \Pr[\pi|\sigma]\cdot\left(\sum_{\sigma'\in\mathcal{N}(\sigma)} \Pr[r(\pi)=\sigma']\right) \\
&\leq 1 - \sum_{\sigma'\in\mathcal{N}(\sigma)} \sum_{\pi\in\mathcal{L}(A)^n} \varphi^n\cdot\Pr[\pi|\sigma']\cdot\Pr[r(\pi)=\sigma'] \\
&= 1 - \varphi^n \cdot \sum_{\sigma'\in\mathcal{N}(\sigma)} \mathrm{Acc}^r(n,\sigma') \leq 1 - \varphi^n\cdot(m-1)\cdot(1-\epsilon),
\end{aligned}$$

where the fifth transition follows by changing the order of summation and Equation (2), and the last transition follows since $\mathrm{Acc}^r(n) \geq 1-\epsilon$ and $|\mathcal{N}(\sigma)| = m-1$. Thus, to achieve an accuracy of at least $1-\epsilon$, we need $\varphi^n\cdot(m-1)\cdot(1-\epsilon) \leq \epsilon$, and the theorem follows by solving for n. □ (Theorem 3.7)

3.2. Scoring rules may require exponentially many samples

While Theorems 3.6 and 3.7 show that every PM-c rule requires an asymptotically optimal (and in particular, logarithmic) number of samples to determine the true ranking with high probability, some classical voting rules such as plurality fall short. In particular, we establish that plurality requires at least exponentially many samples to determine the true ranking with high probability. Since plurality relies on the number of appearances of various alternatives in the first position, our analysis crucially relies on the probability of different alternatives coming first in a random vote, i.e., $p_{i,1}$.

LEMMA 3.8. *$p_{i,1} = \varphi^{i-1} / \left(\sum_{j=1}^{m} \varphi^{j-1} \right)$ for all $i \in \{1, \ldots, m\}$.*

PROOF. Recall that a_i denotes the alternative at position i in the true ranking σ^*. First we prove that for any $i \in \{1, \ldots, m-1\}$, we have $p_{i+1,1} = \varphi \cdot p_{i,1}$. To see this,

$$\begin{aligned} p_{i,1} - p_{i+1,1} &= \frac{\sum_{\sigma \in \mathcal{L}(A) | \sigma(a_i) = 1} \varphi^{d_{KT}(\sigma,\sigma^*)} - \sum_{\sigma \in \mathcal{L}(A) | \sigma(a_{i+1}) = 1} \varphi^{d_{KT}(\sigma,\sigma^*)}}{Z_\varphi^m} \\ &= \frac{\sum_{\sigma \in \mathcal{L}(A) | \sigma(a_i) = 1} \left(\varphi^{d_{KT}(\sigma,\sigma^*)} - \varphi^{d_{KT}(\sigma_{a_i \leftrightarrow a_{i+1}},\sigma^*)} \right)}{Z_\varphi^m} \\ &= \sum_{\sigma \in \mathcal{L}(A) | \sigma(a_i) = 1} \frac{\varphi^{d_{KT}(\sigma,\sigma^*)} \cdot (1-\varphi)}{Z_\varphi^m} = (1-\varphi) \cdot p_{i,1}, \end{aligned}$$

where the second transition follows since $\sigma \leftrightarrow \sigma_{a_i \leftrightarrow a_{i+1}}$ is a bijection between the set of all rankings where a_i is first and the set of all rankings where a_{i+1} is first, and the third transition follows due to Lemma 3.5. Hence, $p_{i,1} - p_{i+1,1} = (1-\varphi) \cdot p_{i,1}$, which implies that $p_{i+1,1} = \varphi \cdot p_{i,1}$. Applying this repeatedly, we have that $p_{i,1} = p_{1,1} \cdot \varphi^{i-1}$, for every $i \in \{1, \ldots, m\}$. Summing over $1 \le i \le m$ and observing that $\sum_{i=1}^{m} p_{i,1} = 1$, we get the desired result. □ (Lemma 3.8)

Lemma 3.8 directly implies that the probability of sampling votes in which a_{m-1} or a_m (the two alternatives that are ranked at the bottom of σ^*) are at the top is exponentially small, hence plurality requires an exponential number of samples to "see" these alternatives and distinguish between them. However, what makes the proof more difficult is that in theory the tie-breaking scheme may help plurality return the true ranking; indeed it is known that the choice of tie breaking scheme can significantly affect a rule's performance [Obraztsova et al. 2011]. However, we show that here this is not the case, i.e., our lower bound works for any natural (randomized) tie-breaking scheme.

THEOREM 3.9. *For any $\epsilon \in (0, 1/4]$, plurality (with any possibly randomized tie-breaking scheme that depends on the top-ranked alternatives of the input votes) requires $\Omega((1/\varphi)^m)$ samples from Mallows' model to output the true ranking with probability at least $1-\epsilon$.*

PROOF. We first note that instead of operating on a profile $\pi \in \mathcal{L}(A)^n$, plurality (and its tie-breaking scheme) operates on the vector of its plurality votes $v \in A^n$ (we call it a top-vote) which consists of the top-ranked alternatives of the different votes of π. The probability of observing a top-vote v given a true ranking σ^* is the sum of the probabilities of observing profiles whose top-vote is v; we denote this by $\Pr[v|\sigma^*]$. The accuracy of the plurality rule (denoted PL) with n samples on a true ranking σ can now equivalently be written as

$$\mathrm{Acc}^{\mathrm{PL}}(n, \sigma) = \sum_{v \in A^n} \Pr[v|\sigma] \cdot \Pr[\mathrm{PL}(v) = \sigma]. \tag{3}$$

Fix $\epsilon \in (0, 1/4]$ and suppose we have $\mathrm{Acc}^{\mathrm{PL}}(n) \ge 1 - \epsilon$, i.e., $\mathrm{Acc}^{\mathrm{PL}}(n, \sigma) \ge 1 - \epsilon$ for all $\sigma \in \mathcal{L}(A)$. We want to show that $n = \Omega((1/\varphi)^m)$. Let the set of alternatives be $A = \{a_1, \ldots, a_m\}$. Consider two distinct rankings: $\sigma_1 = (a_1 \succ \ldots \succ a_{m-2} \succ a_{m-1} \succ a_m)$ and $\sigma_2 = (a_1 \succ \ldots \succ a_{m-2} \succ a_m \succ a_{m-1})$ (where the last two alternatives are swapped compared to σ_1). Let $\hat{A} = A \setminus \{a_{m-1}, a_m\}$. We can decompose Equation (3) into two parts: (i) a summation over $v \in \hat{A}^n$ (when plurality does not "see" alternatives a_{m-1}

and a_m); denote this by $f(\sigma)$, and (ii) a summation over $v \in A^n \setminus \hat{A}^n$ (when plurality "sees" at least one of them); denote this by $g(\sigma)$.

For any $v \in \hat{A}^n$, we have $\Pr[v|\sigma_1] = \Pr[v|\sigma_2]$. To see this, observe that in any profile π with top-vote v we can swap alternatives a_{m-1} and a_m in all the votes to obtain (the unique) profile π' which importantly also has top-vote v and $\Pr[\pi|\sigma_1] = \Pr[\pi'|\sigma_2]$. Summing over all profiles with top-vote v, this yields $\Pr[v|\sigma_1] = \Pr[v|\sigma_2]$. Therefore, we have

$$f(\sigma_1) + f(\sigma_2) = \sum_{v \in \hat{A}^n} \Pr[v|\sigma_1] \cdot (\Pr[\mathrm{PL}(v) = \sigma_1] + \Pr[\mathrm{PL}(v) = \sigma_2]) \leq \sum_{v \in \hat{A}^n} \Pr[v|\sigma_1] \leq 1.$$

Further,

$$g(\sigma_1) = \sum_{v \in A^n \setminus \hat{A}^n} \Pr[v|\sigma_1] \cdot \Pr[\mathrm{PL}(v) = \sigma_1] \leq \sum_{v \in A^n \setminus \hat{A}^n} \Pr[v|\sigma_1],$$

where the right hand side is the probability that at least one of the two alternatives a_{m-1} and a_m comes first in at least one vote. Let $t_{i,j}$ denote the number of votes in which alternative a_i appears in position j. Then we have

$$g(\sigma_1) \leq \Pr[(t_{m-1,1} > 0) \vee (t_{m,1} > 0)] \leq \Pr[t_{m-1,1} > 0] + \Pr[t_{m,1} > 0],$$

where the last transition is due to the union bound.

The probability that alternative a_{m-1} appears first in a vote is $p_{m-1,1}$. Therefore, the probability that it appears first in at least one vote is at most $n \cdot p_{m-1,1}$ by the union bound. Similarly, $\Pr[t_{m,1} > 0] \leq n \cdot p_{m,1}$. Therefore, $g(\sigma_1) \leq n \cdot (p_{m-1,1} + p_{m,1})$. In the same way, we can obtain $g(\sigma_2) \leq n \cdot (p_{m-1,1} + p_{m,1})$. Finally, using the bounds obtained on f and g, we have

$$\mathrm{Acc}^{\mathrm{PL}}(n, \sigma_1) + \mathrm{Acc}^{\mathrm{PL}}(n, \sigma_2) = (f(\sigma_1) + f(\sigma_2)) + g(\sigma_1) + g(\sigma_2) \leq 1 + 2 \cdot n \cdot (p_{m-1,1} + p_{m,1}).$$

We assumed that $\mathrm{Acc}^{\mathrm{PL}}(n, \sigma) \geq 1 - \epsilon$ for every $\sigma \in \mathcal{L}(A)$. Therefore, we need $1 + 2 \cdot n \cdot (p_{m-1,1} + p_{m,1}) \geq 2 \cdot (1 - \epsilon)$, i.e.,

$$n \geq \frac{1 - 2 \cdot \epsilon}{2 \cdot (p_{m-1,1} + p_{m,1})} \geq \frac{1}{8 \cdot p_{m-1,1}} = \frac{\sum_{j=0}^{m-1} \varphi^j}{8 \cdot \varphi^{m-2}} \geq \frac{1}{8 \cdot \varphi^{m-2}},$$

where the second transition follows since $\epsilon \in (0, 1/4]$ and $p_{m,1} < p_{m-1,1}$, and the third transition follows by Lemma 3.8. Thus, plurality requires $\Omega((1/\varphi)^m)$ samples to output the true ranking with high probability. □ (Theorem 3.9)

Plurality has terrible performance because it ranks alternatives by just observing their number of appearances in the first positions of the input votes. In contrast, consider the veto rule that essentially ranks alternatives in the ascending order of their number of appearances at the bottom of input votes. By symmetry we have $p_{m,m} = p_{1,1}$ and $p_{m-1,m} = p_{2,1}$, both of which are lower bounded by constants due to Lemma 3.8. Hence, veto requires only constantly many samples to distinguish between a_{m-1} and a_m. Nevertheless, it is difficult for both plurality and veto to distinguish between alternatives $a_{m/2}$ and $a_{m/2+1}$ that are far from both ends. Certain scoring rules, such as Borda count or the harmonic scoring rule, take into consideration the number of appearances of an alternative at *all* positions. We show that a positional scoring rule that gives different weights to all positions and does not give some position exponentially higher weight than any other position would require only polynomially many samples. The proof is given in Appendix B.

THEOREM 3.10. *Consider a positional scoring rule r given by scoring vector $(\alpha_1, \ldots, \alpha_m)$. For $i \in \{1, \ldots, m-1\}$, define $\beta_i = \alpha_i - \alpha_{i+1}$. Let $\beta_{\max} = \max_{i<m} \beta_i$ and $\beta_{\min} = \min_{i<m} \beta_i$. Assume $\beta_{\min} > 0$ and let $\beta^* = \beta_{\max}/\beta_{\min}$. Then for any $\epsilon > 0$, rule r requires $O((\beta^*)^2 \cdot m^2 \cdot \log(m/\epsilon))$ samples to output the true ranking with probability at least $1-\epsilon$.*

While Theorem 3.10 shows that scoring rules such as Borda count and the harmonic rule have polynomial sample complexity, it does not apply to scoring rules such as plurality and veto since they have $\beta_{\min} = 0$. Note that in Borda count all β_i's are equal, hence it is the rule with the lowest possible $\beta^* = 1$.

4. MOVING TOWARDS GENERALIZATIONS

Section 3 focused on Mallows' model and sample complexity. In Section 5 we will consider a much higher level of abstraction, including much more general noise models and infinitely many samples. This section serves as a mostly conceptual interlude where we gradually introduce some new ideas.

4.1. From finite to infinitely many samples and the family of PD-c rules

While the exact or asymptotic sample complexity — as analyzed in Section 3 — can help us distinguish between various voting rules, here we take a normative point of view and argue that voting rules need to meet a basic requirement: given *infinitely many* samples, the rule should be able to reproduce the true ranking with probability 1. Formally, a voting rule r is *accurate in the limit* for a noise model G if given votes from G, $\lim_{n\to\infty} \mathrm{Acc}^r(n) = 1$.

For Mallows' model, achieving accuracy-in-the-limit is very easy. Theorem 3.6 shows that given $O(\log(m/\epsilon))$ samples, every PM-c rule outputs the true ranking with probability at least $1-\epsilon$. Thus, every PM-c rule is accurate in the limit for Mallows' model. While plurality requires at least exponentially many samples to determine the true ranking with high probability (Theorem 3.9), a matching upper bound (up to logarithmic factors) can trivially be established showing that plurality is accurate in the limit for Mallows' model as well. In fact, it can be argued that all scoring rules are accurate in the limit for Mallows' model. We prove a more general statement by introducing a novel family of voting rules that generalizes scoring rules and showing that all rules in this family are accurate in the limit for Mallows' model.

Definition 4.1 (*Position-Dominance*). Given a profile $\pi = (\sigma_1, \ldots, \sigma_n) \in \mathcal{L}(A)^n$, alternative $a \in A$ and $j \in \{1, \ldots, m-1\}$, define $s_j(a) = |\{i : \sigma_i(a) \leq j\}|$, i.e., the number of votes in which alternative a is among first j positions. For $a, b \in A$, we say that a *position-dominates* b if $s_j(a) > s_j(b)$ for all $j \in \{1, \ldots, m-1\}$. The position-dominance graph (PD graph) of π is defined as the directed graph $G = (V, E)$ where alternatives are vertices ($V = A$) and there is an edge from alternative a to alternative b if a position-dominates b.

The concept of position-dominance is reminiscent of the notion of first-order stochastic dominance in probability theory: informally, a random variable (first-order) stochastically dominates another random variable over the same domain if for any value in the domain the former random variable has higher probability of being above the value than the latter random variable. Also note that position-dominance is a transitive relation; for alternatives $a, b, c \in A$ if a position-dominates b and b position-dominates c, then a position-dominates c. However, it is possible that for some alternatives $a, b \in A$, neither a position-dominates b nor b position-dominates a. Thus, the PD graph is always acyclic, but not always complete. When the PD graph is complete, it reduces to a ranking, similarly to the case of the PM graph.

Definition 4.2 (*Position-Dominance Consistent Rules*). A deterministic voting rule r is called *position-dominance consistent* (PD-c) if $r(\pi) = \sigma$ whenever the PD graph of profile π reduces to ranking σ. For randomized voting rules, we require that $\Pr[r(\pi) = \sigma] = 1$.

This novel family of rules captures voting rules that give higher preference to alternatives that appear at earlier positions. It is quite intuitive that all positional scoring rules are PD-c because they score alternatives purely based on their positions in the rankings and give higher weight to alternatives at earlier positions. PD-c rules also capture another classical voting rule — the Bucklin rule. The definition of the Bucklin rule and the proof of Theorem 4.3 appear in Appendix D.

THEOREM 4.3. *All positional scoring rules and the Bucklin rule are PD-c rules.*

It is easy to argue that all PD-c rules are accurate in the limit for Mallows' model. Let σ^* be the true ranking and a_i be the alternative at position i in σ^*. If we construct a profile by sampling n votes from Mallows' model, then $\mathbb{E}[s_j(a_i)] = n \cdot q_{i,j}$. Recall that $q_{i,j}$ is the probability of alternative a_i appearing among the first j positions in a random vote. Clearly in Mallows' model, $q_{i,j} > q_{l,j}$ for any $i < l$. Therefore, as $n \to \infty$, we will have $\Pr[s_j(a_i) > s_j(a_l)] = 1$ for all $j \in \{1, \ldots, m-1\}$ and $i < l$. Hence, the PD graph of the profile would reduce to σ^* (so any PD-c rule will output σ^*) with probability 1 as $n \to \infty$. We conclude that all PD-c rules are accurate in the limit for Mallows' model.

4.2. PM-c rules are disjoint from PD-c rules

In Theorem 3.3 we saw various classical voting rules that are PM-c, and Theorem 4.3 describes well-known voting rules that are PD-c. At first glance, the definitions of PM-c and PD-c may seem unrelated. However, it turns out that no voting rule can be both PM-c and PD-c. To show this we give a carefully constructed profile where both the PM graph and the PD graph are acyclic and complete, but they reduce to different rankings. Hence, a rule that is both PM-c and PD-c must output two different rankings with probability 1, which is impossible. For our example, let $A = \{a, b, c\}$ be the set of alternatives. The profile π consisting of 11 votes is given below.

4 votes	2 votes	3 votes	2 votes
a	b	b	c
b	a	c	a
c	c	a	b

It is easy to check that the PM graph of π reduces to $a \succ b \succ c$ and the PD graph of π reduces to $b \succ a \succ c$. Thus, we have the following result.

THEOREM 4.4. *No (randomized) voting rule can be both PM-c and PD-c.*

The theorem is not entirely surprising, as it is known that there is no positional scoring rule that is Condorcet consistent [Fishburn 1974]. Note that in addition to PM-c rules and PD-c rules, we can construct numerous simple rules that are also accurate in the limit for Mallows' model, such as the rule that ranks alternatives according to their most frequent position in the input votes and the rule that outputs the most frequent ranking.

4.3. Generalizing the noise model

While being accurate in the limit for Mallows' model can be seen as a necessity for voting rules, the assumption that the noise observed in practice would perfectly (or even approximately) fit Mallows' model is unrealistic. For example, Mao et al. [2013] show that, in certain real-world scenarios, the noise observed is far from what Mallows predicts. While voting rules cannot be expected to have low sample complexity in all types of noise models that arise in practice, it is reasonable to expect them to be at least accurate in the limit for such noise models. Indeed, it is not hard to construct voting rules that are accurate in the limit for Mallows' model but not for other reasonable noise models.

Unfortunately, it is not clear what noise models can be expected to arise in practice and little attention has been given to characterizing reasonable noise models in the literature. To address this issue we impose a structure, parametrized by distance functions, on the noise models to make them well-behaved. As noted in Section 1.2, this approach is related to the work of Flinger and Verducci [1986], but we further generalize the structure of the noise model by removing their assumption of exponentially decreasing probabilities.

Definition 4.5 (*d-Monotonic Noise Models*). Let σ^* denote the true underlying ranking. Let $d : \mathcal{L}(A) \times \mathcal{L}(A) \rightarrow \mathbb{R}_{\geq 0}$ be a distance function over rankings. A noise model is called *monotonic* with respect to d (or d-monotonic) if for any $\sigma, \sigma' \in \mathcal{L}(A)$, $d(\sigma, \sigma^*) < d(\sigma', \sigma^*)$ implies $\Pr[\sigma|\sigma^*] > \Pr[\sigma'|\sigma^*]$ and $d(\sigma, \sigma^*) = d(\sigma', \sigma^*)$ implies $\Pr[\sigma|\sigma^*] = \Pr[\sigma'|\sigma^*]$.

In words, given a distance function d we expect that rankings closer to the true ranking would have higher probability of being observed. Note that Mallows' model is monotonic with respect to the KT distance. Any noise model that arises in practice can be expected to be monotonic, and we require that a voting rule be accurate in the limit for any monotonic noise model.

Definition 4.6. A voting rule r is called *monotone-robust* with respect to distance function d (or d-monotone-robust) if r is accurate in the limit for all d-monotonic noise models.

We saw that all PM-c and PD-c rules are accurate in the limit for Mallows' model. In fact, it can be shown that they are accurate in the limit for all d_{KT}-monotonic noise models, i.e., they are d_{KT}-monotone-robust. However, we omit the proof as the theorem will follow from the even more general results of Section 5.

THEOREM 4.7. *All PM-c and PD-c rules are d_{KT}-monotone-robust.*

5. GENERAL CHARACTERIZATIONS

For any given distance function d, we proposed d-monotonic noise models in an attempt to capture noise models that may arise in practice. However, until now we only focused on one specific distance function — the KT distance. Noise models parametrized by other distance functions have been studied in the literature starting with Mallows [1957] himself. In fact, all our previous proofs relied only on the fact that the KT distance is swap-increasing and Theorem 4.7 can also be shown to hold when the KT distance is replaced by any swap-increasing distance. Alas, among the three most popular distance functions that we consider, only the KT distance is swap-increasing.

In this section we ask whether the families of PM-c and PD-c rules are monotone-robust with respect to distance functions other than swap-increasing distances. We fully characterize all distance functions with respect to which all PM-c and/or all PD-c rules are monotone-robust. Given any distance function d, it is easy to construct an equivalent integer-valued distance function d' such that properties like d-monotone-robustness, MC and PC (the latter two are yet to be introduced) are preserved. Thus, without loss of generality we henceforth restrict our distance functions to be integer-valued.

5.1. Distances for which all PM-c rules are monotone-robust

We first characterize the distance functions for which all PM-c rules are monotone-robust. This leads us to the definition of a rather natural family of distance functions, which may be of independent interest.

Definition 5.1 (*Majority-Concentric (MC) Distances*). For any distance function d, ranking $\sigma \in \mathcal{L}(A)$ and integer $k \in \mathbb{N} \cup \{0\}$, let $\mathcal{N}^k(\sigma) = \{\tau \in \mathcal{L}(A) | d(\tau, \sigma) \leq k\}$ be the set of all rankings at distance at most k from σ. Furthermore, for any alternatives $a, b \in A$, let $\mathcal{N}^k_{a \succ b}(\sigma) = \{\tau \in \mathcal{N}^k(\sigma) | a \succ_\tau b\}$. A distance function d is called *majority-concentric* (MC) if for any $\sigma \in \mathcal{L}(A)$ and $a, b \in A$ such that $a \succ_\sigma b$, $|\mathcal{N}^k_{a \succ b}(\sigma)| \geq |\mathcal{N}^k_{b \succ a}(\sigma)|$ for every $k \in \mathbb{N} \cup \{0\}$.

Consider a ranking σ and imagine *concentric* circles around σ where the k^{th} circle from the center represents the neighbourhood $\mathcal{N}^k(\sigma)$. Then, the MC criterion requires that for every pair of alternatives, a (weak) *majority* of rankings in each neighbourhood (which can be viewed as a set of votes) agree with σ, hence the name majority-concentric.

There is an alternative and perhaps more intuitive characterization of MC distances. Fix any MC distance d, base ranking σ and alternatives $a, b \in A$ such that $a \succ_\sigma b$. Let $\mathcal{L}_{a \succ b}(A) = \{\tau \in \mathcal{L}(A) | a \succ_\tau b\}$ denote the set of all rankings where $a \succ b$ and let $\mathcal{L}_{b \succ a}(A) = \mathcal{L}(A) \setminus \mathcal{L}_{a \succ b}(A)$. Let us sort all rankings in both sets in increasing order of their distance from σ, and map the i^{th} ranking (in the sorted order) in $\mathcal{L}_{a \succ b}(A)$ to the i^{th} ranking in $\mathcal{L}_{b \succ a}(A)$. We can show that this mapping takes every ranking to a ranking at equal or greater distance from σ. We call such a mapping *weakly-distance-increasing* with respect to σ. To see this, suppose for contradiction that (say) the i^{th} ranking of $\mathcal{L}_{a \succ b}(A)$ at distance k from σ is mapped to the i^{th} ranking of $\mathcal{L}_{b \succ a}(A)$ at distance $k' < k$ from σ. Then clearly, $|\mathcal{N}^{k'}_{a \succ b}(\sigma)| < i$ and $|\mathcal{N}^{k'}_{b \succ a}(\sigma)| \geq i$, which is a contradiction since we assumed the distance to be MC. In the other direction, again fix any distance d, $\sigma \in \mathcal{L}(A)$ and $a, b \in A$ such that $a \succ_\sigma b$. Suppose there exists a bijection $f : \mathcal{L}_{a \succ b}(A) \to \mathcal{L}_{b \succ a}(A)$ that is weakly-distance-increasing with respect to σ. Then for any $k \in \mathbb{N} \cup \{0\}$ we have $\mathcal{N}^k_{b \succ a}(\sigma) \subseteq \{f(\tau) | \tau \in \mathcal{N}^k_{a \succ b}(\sigma)\}$, so $|\mathcal{N}^k_{a \succ b}(\sigma)| \geq |\mathcal{N}^k_{b \succ a}(\sigma)|$. If this holds for every $\sigma \in \mathcal{L}(A)$ and $a, b \in A$ such that $a \succ_\sigma b$, then the distance is MC. In conclusion, we have proved the following lemma.

LEMMA 5.2. *A distance function d is MC if and only if for every $\sigma \in \mathcal{L}(A)$ and every $a, b \in A$ such that $a \succ_\sigma b$, there exists a bijection $f : \mathcal{L}_{a \succ b}(A) \to \mathcal{L}_{b \succ a}(A)$ which is weakly-distance-increasing with respect to σ.*

We are now ready to prove our first main result of this section: the distance functions with respect to which all PM-c rules are monotone-robust are exactly MC distances.

THEOREM 5.3. *All PM-c rules are d-monotone-robust for a distance function d if and only if d is MC.*

PROOF. First, we assume that d is MC and show that all PM-c rules are d-monotone-robust. Specifically, consider any d-monotonic noise model G; we wish to show that all PM-c rules are accurate in the limit for G. Let σ^* be an arbitrary true ranking and $a, b \in A$ be two arbitrary alternatives with $a \succ_{\sigma^*} b$.

Using Lemma 5.2, there exists an injection $f : \mathcal{L}_{a \succ b}(A) \to \mathcal{L}_{b \succ a}(A)$ which is weakly-distance-increasing with respect to σ^*. Hence, for every $\sigma \in \mathcal{L}_{a \succ b}(A)$, $d(\sigma, \sigma^*) \leq d(f(\sigma), \sigma^*)$, so $\Pr[\sigma | \sigma^*] \geq \Pr[f(\sigma) | \sigma^*]$ since G is d-monotonic. Crucially, $\sigma^* \in \mathcal{L}_{a \succ b}(A)$ and $d(\sigma^*, \sigma^*) = 0 < d(f(\sigma^*), \sigma^*)$, so $\Pr[\sigma^* | \sigma^*] > \Pr[f(\sigma^*) | \sigma^*]$. Recall that f is a bijection, hence its range is the whole of $\mathcal{L}_{b\succ a}(A)$. By summing over all $\sigma \in \mathcal{L}_{a\succ b}(A)$, we get

$$\Pr[a \succ b|\sigma^*] = \sum_{\sigma\in\mathcal{L}_{a\succ b}(A)} \Pr[\sigma|\sigma^*] > \sum_{\sigma\in\mathcal{L}_{a\succ b}(A)} \Pr[f(\sigma)|\sigma^*]$$
$$= \sum_{\sigma\in\mathcal{L}_{b\succ a}(A)} \Pr[\sigma|\sigma^*] = \Pr[b \succ a|\sigma^*].$$

It follows that given infinitely many samples from G, there would be an edge from a to b in the PM graph with probability 1. Since this holds for all $a, b \in A$, the PM graph would reduce to σ^* with probability 1. Therefore, any PM-c rule would output σ^* with probability 1, as required.

In the other direction, consider any distance function d that is not MC. We show that there exists a PM-c rule that is not accurate in the limit for some d-monotonic noise model G. Since d is not MC, there exists a $\sigma^* \in \mathcal{L}(A)$, an integer k and alternatives $a, b \in A$ with $a \succ_{\sigma^*} b$ such that $|\mathcal{N}^k_{a\succ b}(\sigma^*)| < |\mathcal{N}^k_{b\succ a}(\sigma^*)|$. Now we construct the noise model G as follows. Let $M = \max_{\sigma\in\mathcal{L}(A)} d(\sigma, \sigma^*)$ and let $T > M$ (we will set T later). Define a weight w_σ for each ranking σ as follows: if $d(\sigma, \sigma^*) \le k$ (i.e., $\sigma \in \mathcal{N}^k(\sigma^*)$), then $w_\sigma = T - d(\sigma, \sigma^*)$ else $w_\sigma = M - d(\sigma, \sigma^*)$. Now construct G by assigning probabilities to rankings proportionally to their weights, i.e., $\Pr[\sigma|\sigma^*] = w_\sigma / \sum_{\tau\in\mathcal{L}(A)} w_\tau$. First, by the definition of M and the fact that $T > M$, it is easy to check that G is indeed a probability distribution and that G is d-monotone.

Next, we set T such that $\Pr[a \succ b|\sigma^*] < \Pr[b \succ a|\sigma^*]$. Since the probabilities are proportional to the weights, we want to obtain: $\sum_{\sigma\in\mathcal{L}(A)|a\succ_\sigma b} w_\sigma < \sum_{\sigma\in\mathcal{L}(A)|b\succ_\sigma a} w_\sigma$. Let $|\mathcal{N}^k_{a\succ b}(\sigma^*)| = l$, hence $|\mathcal{N}^k_{b\succ a}(\sigma^*)| \ge l + 1$. Now, on the one hand,

$$\sum_{\sigma\in\mathcal{L}_{a\succ b}(A)} w_\sigma \le \sum_{\sigma\in\mathcal{N}^k_{a\succ b}(\sigma^*)} T + \sum_{\sigma\in\mathcal{L}_{a\succ b}(A)\setminus\mathcal{N}^k_{a\succ b}(\sigma^*)} M \le l \cdot T + m! \cdot M.$$

On the other hand,

$$\sum_{\sigma\in\mathcal{L}_{b\succ a}(A)} w_\sigma \ge \sum_{\sigma\in\mathcal{N}^k_{b\succ a}(\sigma^*)} (T - k) + \sum_{\sigma\in\mathcal{L}_{b\succ a}(A)\setminus\mathcal{N}^k_{b\succ a}(\sigma^*)} 0 \ge (l + 1) \cdot (T - k).$$

Now we set T such that $(l + 1) \cdot (T - k) > l \cdot T + m! \cdot M$, i.e., $T > (l + 1) \cdot k + m! \cdot M$. Noting that $l + 1 \le m!$ and $k \le M$, we can achieve this by simply setting $T = 2 \cdot m! \cdot M$.

Since we have obtained $\Pr[a \succ b|\sigma^*] < \Pr[b \succ a|\sigma^*]$ under G, given infinitely many samples there would be an edge from b to a in the PM graph with probability 1. Therefore, with probability 1 the PM graph would not reduce to σ^*. We can easily construct a PM-c rule r that outputs a ranking σ whenever the PM graph reduces to σ, and outputs an arbitrary ranking with $b \succ a$ when the PM graph does not reduce to any ranking. With probability 1, such a rule would output a ranking where $b \succ a$. Hence, r is not accurate in the limit for G, as required. □ (Theorem 5.3)

5.2. Distances for which all PD-c rules are monotone-robust

We next characterize the distance functions for which all PD-c rules are monotone-robust. This leads us to define another natural family of distance functions.

Definition 5.4 (*Position-Concentric (PC) Distances*). For any ranking $\sigma \in \mathcal{L}(A)$, integer $k \in \mathbb{N} \cup \{0\}$, integer $j \in \{1, \ldots, m-1\}$ and alternative $a \in A$, let $\mathcal{S}^k_j(\sigma, a) = \{\tau \in \mathcal{N}^k(\sigma) | \tau(a) \le j\}$ be the set of rankings at distance at most k from σ where alternative a is ranked in the first j positions. A distance function d is called *position-concentric*

(PC) if for any $\sigma \in \mathcal{L}(A)$, $j \in \{1, \ldots, m-1\}$, and $a, b \in A$ such that $a \succ_\sigma b$, we have that $|\mathcal{S}_j^k(\sigma, a)| \geq |\mathcal{S}_j^k(\sigma, b)|$ for all $k \in \mathbb{N} \cup \{0\}$, and strict inequality holds for some $k \in \mathbb{N} \cup \{0\}$.

While MC distances are defined by matching aggregate pairwise comparisons of alternatives in every circle that is centered on the base ranking, PC distances focus on matching pairwise comparisons of aggregate positions of alternatives in every concentric circle. Similarly to Lemma 5.2 for MC distances, PC distances also admit an equivalent characterization. We use this equivalence and show that PC distances are exactly the distance functions with respect to which all PD-c rules are monotone-robust. The proofs appear in Appendix F.

Let $\mathcal{S}_j(a) = \{\sigma \in \mathcal{L}(A) | \sigma(a) \leq j\}$ denote the set of all rankings where alternative a is ranked among the first j positions. Call a distance function $d : X \to Y$ *distance-increasing* with respect to a ranking σ if $d(f(\tau), \sigma) \geq d(\tau, \sigma)$ for every $\tau \in X$ (i.e., d is weakly-distance-increasing) and strict inequality holds for at least one $\tau \in X$.

LEMMA 5.5. *A distance function d is PC if and only if for every $\sigma \in \mathcal{L}(A)$, every $a, b \in A$ such that $a \succ_\sigma b$ and every $j \in \{1, \ldots, m-1\}$, there exists a bijection $f : \mathcal{S}_j(a) \to \mathcal{S}_j(b)$ which is distance-increasing with respect to σ.*

THEOREM 5.6. *All PD-c rules are d-monotone-robust for a distance function d if and only if d is PC.*

We proved that MC and PC are exactly the distance functions with respect to which all PM-c rules and all PD-c rules, respectively, are monotone-robust. If a distance function d is both MC and PC, then it follows that all PM-c as well as all PD-c rules are d-monotone-robust. On the other hand, if d is not MC (resp., not PC), then there exists a PM-c rule (resp., a PD-c rule) that is not d-monotone-robust. We therefore have the following corollary.

COROLLARY 5.7. *All rules in the union of PM-c rules and PD-c rules are d-monotone-robust for a distance function d if and only if d is both MC and PC.*

Fix any true ranking $\sigma^* \in \mathcal{L}(A)$ and alternatives $a, b \in A$ such that $a \succ_{\sigma^*} b$. Consider any swap-increasing distance function d. By definition, the mapping which maps every ranking σ with $a \succ_\sigma b$ to the ranking $\sigma_{a \leftrightarrow b}$ increases the distance by at least 1. Therefore it is clearly weakly-distance-increasing with respect to σ^*. Such a mapping is also a bijection from $\mathcal{L}_{a \succ b}(A)$ to $\mathcal{L}_{b \succ a}(A)$. Using Lemma 5.2, it follows that d is MC. While the mapping is also a bijection from $\mathcal{S}_j(a)$ to $\mathcal{S}_j(b)$, it may decrease the distance on $\sigma \in \mathcal{S}_j(a)$ where $b \succ a$. Using additional arguments, however, it is possible to show that d is PC as well. The proof of the following lemma is given in Appendix E.

LEMMA 5.8. *Any swap-increasing distance function is both MC and PC.*

Corollary 5.7 and Lemma 5.8 imply that all PM-c rules and all PD-c rules are d-monotone-robust for any swap-increasing distance d, which implies Theorem 4.7.

5.3. Did we generalize the distance functions enough?

How strong are the characterization results of this section? We saw that all PM-c and PD-c rules are d-monotone-robust for any swap-increasing distance d. However, we remarked at the beginning of this section that we need to widen our family of distances as two of the three popular distances that we study are not swap-increasing. We went ahead and characterized all distance functions for which all PM-c rules or all PD-c rules or both are monotone-robust; respectively, these are all MC distances, all PC distances, and their intersection. Are these families wide enough or do we need to search for better voting rules that work for a bigger family of distance functions? Fortunately, we show that even the intersection of the families of MC and PC distances is sufficiently general to include all three popular distance functions.

THEOREM 5.9. *The KT distance, the footrule distance, and the maximum displacement distance are both MC and PC.*

The proof of Theorem 5.9 appears in Appendix E. Together with Corollary 5.7, it implies that all PM-c rules and all PD-c rules are monotone-robust with respect to all three popular distance functions that we study. We have established that our new families of distance functions are wide enough; this further justifies our focus on PM-c rules and PD-c rules, as they are monotone-robust with respect to all MC and PC distances, respectively.

6. DISCUSSION

While we study three popular distance functions over rankings, we exclude some other distances such as the Cayley distance and the Hamming distance; even the most prominent voting rules such as plurality are not accurate in the limit for any noise models that are monotonic with respect to these distances (see Appendix E). On the one hand, this motivates a study of distance functions over rankings that are more appropriate in the social choice context. On the other hand, one may ask: Which voting rules are monotone-robust even with respect to such distance functions?

Furthermore, we have seen that all PM-c rules and all PD-c rules are accurate in the limit for Mallows' model. We later argued that being accurate in the limit for Mallows' model is a very mild requirement, and there are numerous other voting rules that satisfy it. Is it possible to define a much wider class (possibly within the framework of generalized scoring rules [Xia and Conitzer 2008]) that is accurate in the limit for Mallows' model?

On the conceptual level, we analyze the sample complexity of voting rules as the number of alternatives grows, but our analysis assumes (as is traditionally the case in the literature) that the input to the voting rule is total orders over alternatives. As argued in the introduction, the issue of sample complexity of voting rules directly translates to the problem of estimating the required budget in crowdsourcing tasks. When the number of alternatives is large, obtaining total orders is unrealistic, and inputs with partial information such as pairwise comparisons, partial orders or top-k-lists are employed in practice. Several noise models have been proposed in the literature for the generation of such partial information (see, e.g., [Xia and Conitzer 2011]). Going one step further, Procaccia et. al. [2012] proposed a noise model that can incorporate multiple input formats simultaneously given a true underlying ranking. It would be of great practical interest to extend our sample complexity analysis to such noise models.

Finally, we mentioned several points of view on the comparison of voting rules: social choice axioms, maximum likelihood estimators, and the distance rationalizability framework. Elkind et. al. [2010b] point out the weakness of the connection between the MLE framework and the DR framework by showing that the Kemeny rule is the only rule that is both MLE and distance rationalizable. We argued that asking for a voting rule to be the *maximum* likelihood estimator is too restrictive, and proposed quantifying the sample complexity instead. This begs the question: How does the relaxed framework of sample complexity relate to the DR framework?

REFERENCES

ARROW, K. 1951. *Social Choice and Individual Values*. John Wiley and Sons.

BOUTILIER, C., CARAGIANNIS, I., HABER, S., LU, T., PROCACCIA, A. D., AND SHEFFET, O. 2012. Optimal social choice functions: A utilitarian view. In *Proc. of 13th EC*. 197–214.

BOUTILIER, C. AND PROCACCIA, A. D. 2012. A dynamic rationalization of distance rationalizability. In *Proc. of 26th AAAI*. 1278–1284.

BRAVERMAN, M. AND MOSSEL, E. 2008. Noisy sorting without resampling. In *Proc. of 19th SODA*. 268–276.

CHIERICHETTI, F. AND KLEINBERG, J. 2012. Voting with limited information and many alternatives. In *Proc. of 23rd SODA*. 1036–1055.

CONITZER, V., ROGNLIE, M., AND XIA, L. 2009. Preference functions that score rankings and maximum likelihood estimation. In *Proc. of 21st IJCAI*. 109–115.

CONITZER, V. AND SANDHOLM, T. 2005. Common voting rules as maximum likelihood estimators. In *Proc. of 21st UAI*. 145–152.

CRITCHLOW, D. E., FLIGNER, M. A., AND VERDUCCI, J. S. 1991. Probability models on rankings. *Journal of Mathematical Psychology 35,* 3, 294–318.

ELKIND, E., FALISZEWSKI, P., AND SLINKO, A. 2009. On distance rationalizability of some voting rules. In *Proc. of 12th TARK*. 108–117.

ELKIND, E., FALISZEWSKI, P., AND SLINKO, A. 2010a. Good rationalizations of voting rules. In *Proc. of 24th AAAI*. 774–779.

ELKIND, E., FALISZEWSKI, P., AND SLINKO, A. 2010b. On the role of distances in defining voting rules. In *Proc. of 9th AAMAS*. 375–382.

FISHBURN, P. C. 1974. Paradoxes of voting. *American Political Science Review 68,* 2, 537–546.

FLIGNER, M. A. AND VERDUCCI, J. S. 1986. Distance based ranking models. *Journal of the Royal Statistical Society B 48,* 3, 359 – 369.

LEBANON, G. AND LAFFERTY, J. 2002. Cranking: Combining rankings using conditional probability models on permutations. In *Proc. of 9th ICML*. 363 – 370.

LU, T. AND BOUTILIER, C. 2011. Learning Mallows models with pairwise preferences. In *Proc. of 28th ICML*. 145–152.

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

MAO, A., PROCACCIA, A. D., AND CHEN, Y. 2013. Better human computation through principled voting. Manuscript.

MEILĂ, M., PHADNIS, K., PATTERSON, A., AND BILMES, J. A. 2012. Consensus ranking under the exponential model. arXiv:1206.5265.

MESKANEN, T. AND NURMI, H. 2008. Closeness counts in social choice. In *Power, Freedom, and Voting*, M. Braham and F. Steffen, Eds. Springer-Verlag.

OBRAZTSOVA, S., ELKIND, E., AND HAZON, N. 2011. Ties matter: Complexity of voting manipulation revisited. In *Proc. of 10th AAMAS*. 71–78.

PROCACCIA, A. D., REDDI, S. J., AND SHAH, N. 2012. A maximum likelihood approach for selecting sets of alternatives. In *Proc. of 28th UAI*. 695–704.

XIA, L. AND CONITZER, V. 2008. Generalized scoring rules and the frequency of coalitional manipulability. In *Proc. of 9th EC*. 109–118.

XIA, L. AND CONITZER, V. 2011. A maximum likelihood approach towards aggregating partial orders. In *Proc. of 22nd IJCAI*. 446–451.

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

Online Appendix to: When Do Noisy Votes Reveal the Truth?

IOANNIS CARAGIANNIS, University of Patras & CTI
ARIEL D. PROCACCIA, Carnegie Mellon University
NISARG SHAH, Carnegie Mellon University

A. KEMENY IS OPTIMAL FOR MALLOWS' MODEL

We show that for any number of alternatives m and any accuracy level ϵ, the Kemeny rule (with uniform tie-breaking) requires the minimum number of sample from Mallows' model. Note that the Kemeny rule is the maximum likelihood estimator (MLE) for the true ranking given samples from Mallows' model. Formally, given a profile $\pi = (\sigma_1, \ldots, \sigma_n)$ from Mallows' model, the MLE estimator of the true ranking is

$$\begin{aligned}\arg\max_{\sigma \in \mathcal{L}(A)} \Pr[\pi|\sigma] &= \arg\max_{\sigma \in \mathcal{L}(A)} \prod_{i=1}^{n} \frac{\varphi^{-d_{KT}(\sigma_i,\sigma)}}{Z_\varphi^m} \\ &= \arg\max_{\sigma \in \mathcal{L}(A)} \frac{\varphi^{-\sum_{i=1}^n d_{KT}(\sigma_i,\sigma)}}{(Z_\varphi^m)^n} = \arg\min_{\sigma \in \mathcal{L}(A)} \sum_{i=1}^{n} d_{KT}(\sigma_i, \sigma),\end{aligned}$$

where the last expression denotes the Kemeny ranking. While at first glance it may seem that this directly implies optimality of Kemeny rule in terms of its sample complexity, we demonstrate via an example that the MLE rule for a noise model need not always be the rule with the optimal sample complexity in general.

Example A.1. Consider a scenario where there are 3 possible underlying ground truths — σ_1, σ_2 and σ_3. These map to underlying true ranking in the voting context. Let there be 4 possible outcomes — π_1 through π_4. The outcomes map to samples from Mallows' model in our voting context. In the table below, entry in row i and column j gives the probability of observing outcome π_j given that the ground truth is σ_i, i.e., $\Pr[\pi_j|\sigma_i]$.

	π_1	π_2	π_3	π_4
σ_1	$^1/_5$	$^1/_5$	$^1/_5$	$^2/_5$
σ_2	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_2$
σ_3	$^1/_4$	$^1/_4$	$^1/_4$	$^1/_4$

Take $\epsilon = 4/5$, so the accuracy requirement is $1 - \epsilon = 1/5$. Given just one sample from the noise model, the circled entries in the table show the ground truth returned by the MLE rule for various outcomes. It is clear that the MLE rule never returns σ_1, thus it does not achieve the minimum (over all ground truths) accuracy of $1/5$. In contrast, consider the rule which is identical to the MLE rule except that it returns σ_1 when observing π_3. It is clear that given one sample, this rule returns the ground truth with probability at least $1/5$ no matter what the ground truth is. Hence, the sample complexity of the new rule is strictly less than that of the MLE rule for $\epsilon = 4/5$. This shows that the MLE rule need not always be optimal in terms of its sample complexity.

DOI: http://dx.doi.org/10.1145/0000000.0000000

While Example A.1 shows that the MLE rule need not always have the optimal sample complexity, we show that the Kemeny rule (which is MLE for Mallows' model) indeed has the optimal sample complexity. Let KEM denote the Kemeny rule where ties are broken uniformly at random. That is, for any profile $\pi = (\sigma_1, \ldots, \sigma_n) \in \mathcal{L}(A)^n$, let $\textsc{Tie-Kem}(\pi) = \arg\min_{\sigma \in \mathcal{L}(A)} \sum_{i=1}^{n} d_{KT}(\sigma_i, \sigma)$ denote the set of all rankings that are tied to be the Kemeny ranking. Then $\Pr[\mathrm{KEM}(\pi) = \sigma] = 1/|\textsc{Tie-Kem}(\pi)|$ for every $\sigma \in \textsc{Tie-Kem}(\pi)$.

Proof of Theorem 3.1. Note that by definition of $N^r(\epsilon)$, it is sufficient to show that the Kemeny rule has the greatest accuracy among all voting rules for any number of samples, that is, $\mathrm{Acc}^{\mathrm{KEM}}(k) \geq \mathrm{Acc}^r(k)$ for all rules r and all $k > 0$. To show that KEM has the greatest accuracy, we need two lemmas. Define $\mathrm{TotAcc}^r(k) = \sum_{\sigma \in \mathcal{L}(A)} \mathrm{Acc}^r(k, \sigma)$.

Lemma A.2. $\mathrm{Acc}^{\mathrm{KEM}}(k, \sigma) = \mathrm{Acc}^{\mathrm{KEM}}(k, \sigma'), \forall \sigma, \sigma' \in \mathcal{L}(A), \forall k \in \mathbb{N}$.

Lemma A.3. $\mathrm{TotAcc}^{\mathrm{KEM}}(k) \geq \mathrm{TotAcc}^r(k), \forall$ *rule* $r, \forall k \in \mathbb{N}$.

First, it is easy to derive the final result using Lemmas A.2 and A.3. Fix any $\epsilon > 0$ and let $N^{\mathrm{KEM}}(\epsilon) = k$. Then, there exists $\hat{\sigma} \in \mathcal{L}(A)$ such that $\mathrm{Acc}^{\mathrm{KEM}}(k-1, \hat{\sigma}) < 1 - \epsilon$, hence $\mathrm{Acc}^{\mathrm{KEM}}(k-1, \sigma) < 1 - \epsilon$ for every $\sigma \in \mathcal{L}(A)$ due to Lemma A.2. Hence, $\mathrm{TotAcc}^{\mathrm{KEM}}(k-1) < m! \cdot (1-\epsilon)$. Now for any voting rule r, Lemma A.3 implies $\mathrm{TotAcc}^r(k-1) \leq \mathrm{TotAcc}^{\mathrm{KEM}}(k-1) < m! \cdot (1-\epsilon)$ and hence by pigeonhole principle, there exists $\sigma \in \mathcal{L}(A)$ such that $\mathrm{Acc}^r(k-1, \sigma) < 1 - \epsilon$. Therefore, $N^r(\epsilon) \geq k = N^{\mathrm{KEM}}(\epsilon)$, as required. Now we prove Lemmas A.2 and A.3.

Proof of Lemma A.2. Take any $k \in \mathbb{N}$ and $\sigma, \sigma' \in \mathcal{L}(A)$. Let $\omega : A \to A$ be the (unique) bijection that when applied on σ gives σ'. That is, $\omega(\sigma(i)) = \sigma'(i)$ for all $1 \leq i \leq m$. We abuse the notation and extend ω to a bijection $\omega : \mathcal{L}(A) \to \mathcal{L}(A)$ where for any $\tau \in \mathcal{L}(A)$, we have $(\omega(\tau))(i) = \omega(\tau(i))$. Essentially, we apply ω on each element of a ranking. So $\omega(\sigma) = \sigma'$. Finally, we further extend ω to operate on profiles where we apply ω to each ranking in the profile individually. Then,

$$
\begin{aligned}
\mathrm{Acc}^{\mathrm{KEM}}(k, \sigma') &= \sum_{\pi \in \mathcal{L}(A)^k} \Pr[\pi|\sigma'] \cdot \Pr[\mathrm{KEM}(\pi) = \sigma'] \\
&= \sum_{\omega(\pi) \in \mathcal{L}(A)^k} \Pr[\omega(\pi)|\sigma'] \cdot \Pr[\mathrm{KEM}(\omega(\pi)) = \sigma'] \\
&= \sum_{\pi \in \mathcal{L}(A)^k} \Pr[\pi|\omega^{-1}(\sigma')] \cdot \Pr[\mathrm{KEM}(\pi) = \omega^{-1}(\sigma')] \\
&= \sum_{\pi \in \mathcal{L}(A)^k} \Pr[\pi|\sigma] \cdot \Pr[\mathrm{KEM}(\pi) = \sigma] = \mathrm{Acc}^{\mathrm{KEM}}(k, \sigma).
\end{aligned}
$$

The second transition follows since ω is a bijection, the third transition follows since Mallows' model and Kemeny rule with uniform tie-breaking are anonymous with respect to the alternatives (note that uniform tie-breaking plays an important role), and the fourth transition follows since $\omega^{-1}(\sigma') = \sigma$. □ (Lemma A.2)

PROOF OF LEMMA A.3. For any rule r and any $k \in \mathbb{N}$,

$$\begin{aligned}
\text{TotAcc}^r(k) &= \sum_{\sigma \in \mathcal{L}(A)} \sum_{\pi \in \mathcal{L}(A)^k} \Pr[\pi|\sigma] \cdot \Pr[r(\pi) = \sigma] = \sum_{\pi \in \mathcal{L}(A)^k} \sum_{\sigma \in \mathcal{L}(A)} \Pr[\pi|\sigma] \cdot \Pr[r(\pi) = \sigma] \\
&\le \sum_{\pi \in \mathcal{L}(A)^k} \sum_{\sigma \in \mathcal{L}(A)} \Pr[r(\pi) = \sigma] \cdot \left(\max_{\sigma' \in \mathcal{L}(A)} \Pr[\pi|\sigma'] \right) = \sum_{\pi \in \mathcal{L}(A)^k} \max_{\sigma' \in \mathcal{L}(A)} \Pr[\pi|\sigma'] \\
&= \sum_{\pi \in \mathcal{L}(A)^k} \max_{\sigma' \in \mathcal{L}(A)} \Pr[\pi|\sigma'] \cdot \sum_{\sigma \in \text{TIE-KEM}(\pi)} \frac{1}{|\text{TIE-KEM}(\pi)|} \\
&= \sum_{\pi \in \mathcal{L}(A)^k} \sum_{\sigma \in \text{TIE-KEM}(\pi)} \Pr[\pi|\sigma] \cdot \Pr[\text{KEM}(\pi) = \sigma] = \text{TotAcc}^{\text{KEM}}(k),
\end{aligned}$$

where the sixth transition follows since the Kemeny rule is MLE for Mallows' model, hence $\max_{\sigma' \in \mathcal{L}(A)} \Pr[\pi|\sigma'] = \Pr[\pi|\sigma]$ for every $\sigma \in \text{TIE-KEM}(\pi)$. Also, we have uniform tie-breaking, so $\Pr[\text{KEM}(\pi) = \sigma] = 1/|\text{TIE-KEM}(\pi)|$ for every $\sigma \in \text{TIE-KEM}(\pi)$. Note that uniform tie-breaking plays an important role. □ (Lemma A.3)

Thus, we have established that to output the true underlying ranking with any given probability, the Kemeny rule with uniform tie-breaking requires the minimum number of samples from Mallows' model among all voting rules. □ (Theorem 3.1)

B. UPPER BOUND FOR A FAMILY OF POSITIONAL SCORING RULES

PROOF OF THEOREM 3.10. Recall that a_i denotes the i^{th} alternative in the true ranking σ^*. Consider a profile π consisting of n samples from Mallows' model. Let $t_{i,j}$ denote the number of times a_i appears in position j, and let $s_{i,k} = \sum_{j=1}^{k} t_{i,j}$. First, we note that for any $i \in \{1, \ldots, m\}$,

$$\begin{aligned}
\sum_{k=1}^{m-1} \beta_k \cdot s_{i,k} &= \sum_{k=1}^{m-1} \beta_k \cdot \left(\sum_{j=1}^{k} t_{i,j} \right) = \sum_{j=1}^{m-1} \left(\sum_{k=j}^{m-1} \beta_k \right) \cdot t_{i,j} = \sum_{j=1}^{m-1} (\alpha_j - \alpha_m) \cdot t_{i,j} \\
&= \sum_{j=1}^{m-1} \alpha_j \cdot t_{i,j} - \alpha_m \cdot (n - t_{i,m}) = \sum_{j=1}^{m} \alpha_j \cdot t_{i,j} - n \cdot \alpha_m,
\end{aligned}$$

where the second transition follows by switching the order of summation and the fourth transition follows since $\sum_{j=1}^{m} t_{i,j} = n$ as the total number of appearances of a_i equals the number of votes. Since $n \cdot \alpha_m$ is independent of the alternative, we can equivalently consider $\sum_{k=1}^{m-1} \beta_k \cdot s_{i,k}$ as the score of alternative a_i. Hence, for rule r to output σ^* with high probability we require $\Pr[\forall i \in \{1, \ldots, m\}, \sum_{k=1}^{m-1} \beta_k \cdot (s_{i,k} - s_{i+1,k}) > 0] \ge 1 - \epsilon$. If we had $\Pr[\sum_{k=1}^{m-1} \beta_k \cdot (s_{i,k} - s_{i+1,k}) \le 0] \le \epsilon/m$ for every $i \in \{1, \ldots, m\}$, then we would obtain (using the union bound) that r outputs σ^* with probability at least $1 - \epsilon$. Observe that

$$\Pr\left[\sum_{k=1}^{m-1} \beta_k \cdot (s_{i,k} - s_{i+1,k}) \le 0\right] \le e^{-\frac{2 \cdot n \cdot \left(\sum_{k=1}^{m-1} \beta_j \cdot (q_{i,k} - q_{i+1,k})\right)^2}{4 \cdot m^2 \cdot \beta_{\max}^2}} \le e^{-\frac{n \cdot \beta_{\min}^2 \cdot \left(\sum_{k=1}^{m-1} (q_{i,k} - q_{i+1,k})\right)^2}{2 \cdot m^2 \cdot \beta_{\max}^2}},$$

where $q_{i,k} = \sum_{j=1}^{k} p_{i,j}$ and the first transition follows due to Hoeffding's inequality. It follows that for this probability to be at most ϵ/m, it is sufficient to have

$$n \geq \frac{2 \cdot m^2 \cdot \beta_{\max}^2}{\beta_{\min}^2 \cdot \left(\sum_{k=1}^{m-1} (q_{i,k} - q_{i+1,k})\right)^2} \log(m/\epsilon)$$

Now we only need to prove that the term $\sum_{k=1}^{m-1} (q_{i,k} - q_{i+1,k})$ in the denominator is lower bounded by a constant independent of m and ϵ. Note that $\sum_{k=1}^{m-1} q_{i,k} = \sum_{k=1}^{m} q_{i,k} - 1 = \sum_{k=1}^{m} \sum_{j=1}^{k} p_{i,j} - 1 = \sum_{j=1}^{m} j \cdot p_{i,j} - 1 = \mathbb{E}[Borda(a_i)] - 1$, where $\mathbb{E}[Borda(a_i)]$ denotes the expected Borda score of alternative a_i under one random sample from Mallows' model. Similarly, $\sum_{k=1}^{m-1} q_{i+1,k} = \mathbb{E}[Borda(a_{i+1})] - 1$. Therefore,

$$\begin{aligned}
\sum_{k=1}^{m-1} (q_{i,k} - q_{i+1,k}) &= \mathbb{E}[Borda(a_i)] - \mathbb{E}[Borda(a_{i+1})] = \mathbb{E}[Borda(a_i) - Borda(a_{i+1})] \\
&= \sum_{\sigma \in \mathcal{L}(A)} \Pr[\sigma|\sigma^*] \cdot ((m+1-\sigma(a_i)) - (m+1-\sigma(a_{i+1}))) \\
&= \sum_{\sigma \in \mathcal{L}(A) | a_i \succ_\sigma a_{i+1}} \Pr[\sigma|\sigma^*] \cdot (\sigma(a_{i+1}) - \sigma(a_i)) \\
&\quad + \sum_{\sigma \in \mathcal{L}(A) | a_{i+1} \succ_\sigma a_i} \Pr[\sigma|\sigma^*] \cdot (\sigma(a_{i+1}) - \sigma(a_i)) \\
&= \sum_{\sigma \in \mathcal{L}(A) | a_i \succ_\sigma a_{i+1}} (\Pr[\sigma|\sigma^*] - \Pr[\sigma_{a_i \leftrightarrow a_{i+1}}|\sigma^*]) \cdot (\sigma(a_{i+1}) - \sigma(a_i)) \\
&\geq \sum_{\sigma \in \mathcal{L}(A) | a_i \succ_\sigma a_{i+1}} (1-\varphi) \cdot \Pr[\sigma|\sigma^*] \cdot 1 = (1-\varphi) \cdot \Pr[a_i \succ a_{i+1}|\sigma^*] \\
&\geq 0.5 \cdot (1-\varphi).
\end{aligned}$$

The second transition follows due to linearity of expectation. The fifth transition follows by noting that under the bijective mapping $\sigma \leftrightarrow \sigma_{a_i \leftrightarrow a_{i+1}}$, we have $\sigma_{a_i \leftrightarrow a_{i+1}}(a_i) = \sigma(a_{i+1})$ and $\sigma_{a_i \leftrightarrow a_{i+1}}(a_{i+1}) = \sigma(a_i)$. For the sixth transition, note that in any σ where $a_i \succ_\sigma a_{i+1}$, $\sigma(a_{i+1}) \geq \sigma(a_i) + 1$. Also, Lemma 3.5 implies that $d_{KT}(\sigma_{a_i \leftrightarrow a_{i+1}}, \sigma^*) = d_{KT}(\sigma, \sigma^*) + 1$, so $\Pr[\sigma|\sigma^*] - \Pr[\sigma_{a_i \leftrightarrow a_{i+1}}|\sigma^*] = (\varphi^{d_{KT}(\sigma,\sigma^*)} - \varphi^{d_{KT}(\sigma_{a_i \leftrightarrow a_{i+1}},\sigma^*)})/Z_\varphi^m = (1-\varphi) \cdot \varphi^{d_{KT}(\sigma,\sigma^*)}/Z_\varphi^m = (1-\varphi) \cdot \Pr[\sigma|\sigma^*]$. The last transition holds trivially (see proof of Theorem 3.6 for a tighter bound). Thus, we have the desired result. □ (Theorem 3.10)

C. SEVERAL CLASSICAL VOTING RULES ARE PM-C

In this section, we prove that the Kemeny rule, the ranked pairs method, Copeland's method and Schulze's method are PM-c (Theorem 3.3). The definition of the Kemeny rule is given in Section 2. We define the remaining methods below.

Copeland's method. We say that alternative a beats alternative a' in a profile π if $|\{\sigma \in \pi | a \succ_\sigma a'\}| > |\{\sigma \in \pi | a' \succ_\sigma a\}|$, i.e., if there is an edge from a to a' in the PM graph of π. The Copeland score of an alternative a is the number of alternatives it beats in π and Copeland's method ranks the alternative in the non-increasing order of their Copeland scores.

The ranked pairs method. Under the ranked pairs method, all ordered pairs of alternatives (a, a') are sorted by the number of rankings in the profile in which alternative

a is preferred to a' (in the non-increasing order). Then, starting with the first pair in the list, the method "locks in" the outcome with the result of the pairwise comparison. It proceeds with the next pairs and locks in every pairwise result that does not contradict (form a cycle) the partial ordering established so far. Finally, the method outputs the total order obtained.

Schulze's method. Given any profile π, define $w(a, a') = |\{\sigma \in \pi | a \succ_\sigma a'\}|$ for any $a, a' \in A$. Consider the (directed) weighted pairwise comparison graph $G = (V, E)$ where the alternatives are the vertices ($V = A$) and there is an edge from every $a \in A$ to every other $a' \in A$ with weight $w(a, a')$. A path of strength t from $a \in A$ to $a' \in A$ is a sequence of vertices $v_0 = a, v_1, \ldots, v_{k-1}, v_k = a'$ where $w(v_i, v_{i+1}) \geq t$ for all $i \in \{0, \ldots, k-1\}$. Define the strength of alternative a over alternative a', denoted $s[a, a']$, to be strength of the strongest path from a to a', if one exists, and 0 otherwise. Schulze's method ranks $a \succ a'$ if $s[a, a'] > s[a', a]$. A tie-breaking scheme is used in the case when $s[a, a'] = s[a', a]$.

PROOF OF THEOREM 3.3. Take any ranking $\sigma^* = (a_1 \succ \ldots \succ a_m) \in \mathcal{L}(A)$. We show that for each of the four rules whenever the PM graph of a profile reduces to σ^*, the rule outputs σ^* with probability 1. Consider any profile π with n votes such that its PM graph reduces to σ^*.

First, note that the Kemeny rule returns the ranking that minimizes the total pairwise disagreements with the input votes. If we consider the (directed) weighted pairwise comparison graph described in the definition of Schulze's method above, then the Kemeny score of a ranking σ, denoted $KemSc(\sigma)$, measures the total pairwise disagreements of σ with the input votes, i.e., $KemSc(\sigma) = \sum_{a,a' \in A | a \succ_\sigma a'} w(a', a)$. Note in this summation, exactly one edge from the two edges between any pair of alternatives is added. In our profile π, for any $a, a' \in A$ with $a \succ_\sigma a'$, $w(a, a') > n/2 > w(a', a)$. Hence, $KemSc(\sigma^*)$ adds the lesser of the two edges between any pair of alternatives. Therefore, σ^* has the minimum Kemeny score. Thus, the Kemeny rule returns σ^*.

When the ranked pairs method is applied to a profile that reduces to σ^*, every ordered pair (a, a') with $a \succ_{\sigma^*} a'$ will be placed before every ordered pair (b, b') with $b' \succ_\sigma b$. This is because the former pair would be consistent with more than half of the rankings in the profile, while the latter pair would be consistent with less than half of the rankings in the profile. Hence, the ranked pairs method would lock every pair (a, a') where $a \succ_{\sigma^*} a'$ (and obtain the total order σ^*) before reaching any pair of the opposite direction. Therefore, the ranked pairs method would also output σ^*.

For Copeland's method, note that when the PM graph reduces to σ^*, then alternative a_i has Copeland score $m - i$, for every $i \in \{1, \ldots, m\}$. Therefore, Copeland's method outputs exactly the ranking σ^*.

Finally for Schulze's method, note that for any $a, a' \in A$ with $a \succ_{\sigma^*} a'$, $s[a, a'] > n/2$ (because the edge a to a' itself has weight more than $n/2$), which is clearly greater than $s[a', a]$. Hence, Schulze's method ranks $a \succ a'$ for every $a, a' \in A$ with $a \succ_{\sigma^*} a'$. Therefore, Schulze's method also outputs σ^*.

We have thus established that the Kemeny rule, the ranked pairs method, Copeland's method and Schulze's method are all PM-c. □ (Theorem 3.3)

D. SEVERAL CLASSICAL VOTING RULES ARE PD-C

We first define the Bucklin rule and then prove Theorem 4.3.

The Bucklin rule. The Bucklin score of an alternative a is the minimum k such that a is among the first k positions in the majority of input votes. The Bucklin rule sorts the alternatives in the non-decreasing order of their Bucklin score and breaks ties among alternatives with the same Bucklin score ℓ in terms of the number of rankings that have the alternative in the first ℓ positions.

PROOF OF THEOREM 4.3. Consider a profile π with n rankings such that its PD graph reduces to the ranking σ^*, and let a_i denote the alternative at position i in σ^*. We show that any positional scoring rule as well as the Bucklin rule outputs σ^* on π.

For the Bucklin rule, consider any two alternatives $a, a' \in A$ such that $a \succ_{\sigma^*} a'$. For any $j \in \{1, \ldots, m-1\}$, let $s_j(c)$ denote the number of votes where alternative c is among the first j positions. Let k denote the Bucklin score of a and k' denote the Bucklin score of a'. If $k > k'$, then $s_{k'}(a') > n/2$ and $s_{k'}(a) \leq s_{k-1}(a) < n/2$, which is impossible since the PD graph reduces to σ^*. If $k < k'$, then the Bucklin rule ranks $a \succ a'$, as required. If $k = k'$ and $k \neq m$, then again since the PD graph reduces to σ^*, we have that $s_k(a) > s_k(a')$, so tie is broken in favor of a. Lastly, we note that $k = k' = m$ is not possible since then it would imply that the total number of appearances of a and a' in the last position is $n - s_{m-1}(a) + n - s_{m-1}(a') > 2 \cdot n - 2 \cdot s_{m-1}(a) \geq n$. Thus, for every $a \succ_{\sigma^*} a'$, the Bucklin rule ranks a above a'. Thus, the Bucklin rule outputs σ^*.

For positional scoring rules, we can follow the reasoning of Theorem 3.10 and express the score of alternative a_i as $\sum_{j=1}^{m-1}(\beta_j \cdot s_j(a_i)) + n\alpha_m$. Then, the desired fact that the score of a_i is higher than that of a_k when $1 \leq i < j \leq m$ follows since $s_j(a_i) > s_j(a_k)$ for every $j \in \{1, ..., m-1\}$. □ (Theorem 4.3)

E. DISTANCE FUNCTIONS

In this section, we prove various properties of the three popular distance functions studied in the paper.

E.1. Definitions

We first give the omitted definitions of the footrule distance and the maximum displacement distance.

The Footrule Distance. The footrule distance between two rankings measures the total displacements of all alternatives between the rankings. Formally, $d_{FR}(\sigma_1, \sigma_2) = \sum_{a \in A} |\sigma_1(a) - \sigma_2(a)|$.

The Maximum Displacement Distance. The maximum displacement distance between two rankings measures the maximum displacement of any alternative between the rankings. Formally, $d_{MD}(\sigma_1, \sigma_2) = \max_{a \in A} |\sigma_1(a) - \sigma_2(a)|$.

E.2. The KT distance is swap-increasing

We first prove Lemma 3.5, showing that the KT distance is swap-increasing.

PROOF OF LEMMA 3.5. Let $\sigma^*, \sigma \in \mathcal{L}(A)$ and $a, b \in A$ with $a \succ_{\sigma^*} b$ and $a \succ_\sigma b$. Let $\sigma(a) = i$ and $\sigma(b) = j$, so $i < j$. Define $Y = \{y \in A | i < \sigma(y) < j\}$. Since $\sigma^*(a) < \sigma^*(b)$, the following properties hold:

(1) For every $y \in Y$, $\sigma^*(y) < \sigma^*(a)$ implies that $\sigma^*(y) < \sigma^*(b)$. Hence,
$$\sum_{y \in Y} \mathbb{1}[\sigma^*(y) < \sigma^*(a)] \leq \sum_{y \in Y} \mathbb{1}[\sigma^*(y) < \sigma^*(b)].$$
(2) For every $y \in Y$, $\sigma^*(b) < \sigma^*(y)$ implies that $\sigma^*(a) < \sigma^*(y)$. Hence,
$$\sum_{y \in Y} \mathbb{1}[\sigma^*(b) < \sigma^*(y)] \leq \sum_{y \in Y} \mathbb{1}[\sigma^*(a) < \sigma^*(y)].$$
(3) $\mathbb{1}[\sigma^*(a) < \sigma^*(b)] = 1$.
(4) $\mathbb{1}[\sigma^*(b) < \sigma^*(a)] = 0$.

Now, we can express $d_{KT}(\sigma_{a\leftrightarrow b}, \sigma^*) - d_{KT}(\sigma, \sigma^*)$ as

$$\begin{aligned} & d_{KT}(\sigma_{a\leftrightarrow b}, \sigma^*) - d_{KT}(\sigma, \sigma^*) \\ &= \sum_{y \in Y} \mathbb{1}[\sigma^*(y) < \sigma^*(b)] + \sum_{y \in Y} \mathbb{1}[\sigma^*(a) < \sigma^*(y)] + \mathbb{1}[\sigma^*(a) < \sigma^*(b)] \\ &\quad - \sum_{y \in Y} \mathbb{1}[\sigma^*(y) < \sigma^*(a)] - \sum_{y \in Y} \mathbb{1}[\sigma^*(b) < \sigma^*(y)] - \mathbb{1}[\sigma^*(b) < \sigma^*(a)] \\ &\geq 1, \end{aligned}$$

as desired. When a and b are adjacent in σ^* (i.e., $\sigma^*(b) = \sigma^*(a) + 1$), we show that equality holds. In this case, observe that the implications in properties (1) and (2) are actually equivalences and the inequalities can be replaced by equalities. Then, the sums in the above derivation cancel out, and it can be seen that the distance increases by exactly 1. □ (Lemma 3.5)

We remarked in Section 5 that among the three distance functions that we consider, the KT distance is the only one that is swap-increasing. Below, we give an example showing that the footrule distance and the maximum displacement distance are not swap-increasing.

Example E.1. Let the set of alternatives $A = \{a, b, c\}$. Let $\sigma^* = (a \succ b \succ c)$ and $\sigma = (b \succ c \succ a)$. Note that $b \succ_\sigma c$ and $b \succ_{\sigma^*} c$. Now consider the ranking $\sigma_{b\leftrightarrow c} = (c \succ b \succ a)$. It is easy to verify that $d_{FR}(\sigma, \sigma^*) = d_{FR}(\sigma_{b\leftrightarrow c}, \sigma^*) = 4$ and $d_{MD}(\sigma, \sigma^*) = d_{MD}(\sigma_{b\leftrightarrow c}, \sigma^*) = 2$. Thus, the distance does not increase by swapping two alternatives that were in the correct order, which shows that neither the footrule distance nor the maximum displacement distance is swap-increasing.

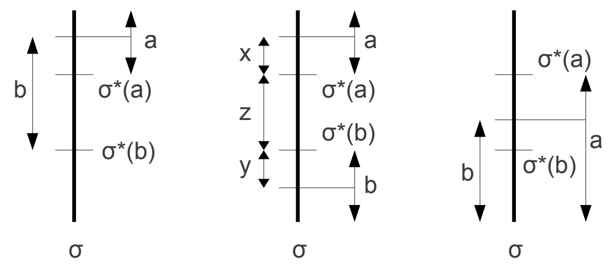
E.3. All three of our popular distance functions are both MC and PC

First we give a proof of Lemma 5.8, showing that any swap-increasing distance is both MC and PD. Lemma 3.5 would then imply that the KT distance is both MC and PC.

PROOF OF LEMMA 5.8. In Section 5.2, we already argued that any swap-increasing distance is MC. Take any $\sigma^* \in \mathcal{L}(A)$ and $a, b \in A$ such that $a \succ_{\sigma^*} b$. The mapping from every σ to $\sigma_{a\leftrightarrow b}$ is a bijection from $\mathcal{L}_{a\succ b}(A)$ to $\mathcal{L}_{b\succ a}(A)$ and it increases the distance by at least 1 (since any ranking in the domain $\mathcal{L}_{a\succ b}(A)$ follows $a \succ b$). Hence, the mapping is weakly-distance-increasing with respect to σ^*. It follows from Lemma 5.2 that any swap-increasing distance is MC.

To show that it is also PC, fix any $\sigma^* \in \mathcal{L}(A)$, $a, b \in A$ such that $a \succ_{\sigma^*} b$ and $j \in \{1, \ldots, m-1\}$. We wish to show that there exists a bijection $f : \mathcal{S}_j(a) \to \mathcal{S}_j(b)$ which is distance-increasing. Note that we cannot use the mapping from every σ to $\sigma_{a\leftrightarrow b}$ as before, since not every ranking in the domain $\mathcal{S}_j(a)$ follows $a \succ b$ and therefore such a mapping would not be guaranteed to increase distance. Instead, we decompose the domain and the range into a total of three parts: $T = \{\sigma \in \mathcal{S}_j(a) | \sigma(b) \leq j\}$, $D_1 = \{\sigma \in \mathcal{S}_j(a) | \sigma(b) > j\}$, and $D_2 = \{\sigma \in \mathcal{S}_j(b) | \sigma(a) > j\}$. Therefore, $\mathcal{S}_j(a) = T \cup D_1$ and $\mathcal{S}_j(b) = T \cup D_2$.

Consider the *identity* bijection $I : T \to T$ which maps every ranking to itself. Clearly, I is weakly-distance-increasing with respect to σ^* since it does not change the distance of any ranking from σ^*. Note that for any $\sigma \in D_1$, $\sigma(a) \leq j$ and $\sigma(b) > j$, so $a \succ_\sigma b$. Further, $\sigma_{a\leftrightarrow b}(a) > j$ and $\sigma_{a\leftrightarrow b}(b) \leq j$. Thus, $\sigma_{a\leftrightarrow b} \in D_2$ and $d(\sigma_{a\leftrightarrow b}, \sigma^*) \geq d(\sigma, \sigma^*) + 1$ (by definition). Therefore, the mapping $E : D_1 \to D_2$ where $E(\sigma) = \sigma_{a\leftrightarrow b}$ is distance-increasing with respect to σ^*. Combining the two, the joint bijection $F : \mathcal{S}_j(a) \to \mathcal{S}_j(b)$ naturally given by $F(\sigma) = I(\sigma)$ when $\sigma \in T$ and $F(\sigma) = E(\sigma)$ when $\sigma \in D_1$ is weakly-distance-increasing with respect to σ^*. Further, it is easy to verify that $D_1 \neq \emptyset$, and F



(a) Case 1 (b) Case 2 (c) Case 3

Fig. 1. Exchanges in the footrule and the maximum displacement distances

increases the distance on any ranking from D_1. Therefore, F is distance-increasing, as required. □ (Lemma 5.8)

Now we are prove Theorem 5.9.

PROOF OF THEOREM 5.9. Lemma 5.8 and Lemma 3.5 already imply that the KT distance is both MC and PC. Now we show that the same holds for the footrule distance (d_{FR}) and the maximum displacement distance (d_{MD}) as well.

Fix any $\sigma^* \in \mathcal{L}(A)$ and $a, b \in A$ such that $a \succ_{\sigma^*} b$. First, we show that for both d_{FR} and d_{MD}, the mapping from every ranking σ with $a \succ_\sigma b$ to $\sigma_{a\leftrightarrow b}$ is weakly-distance-increasing with respect to σ^*. Fix any ranking σ with $a \succ_\sigma b$. We have $\sigma^*(a) < \sigma^*(b)$ and $\sigma(a) < \sigma(b)$. Let $\sigma' = \sigma_{a\leftrightarrow b}$. Recall that $\sigma'(a) = \sigma(b)$ and $\sigma'(b) = \sigma(a)$. For any $c \in A$, let $f(c) = |\sigma(c) - \sigma^*(c)|$ and $f'(c) = |\sigma'(c) - \sigma^*(c)|$ be the displacements of c in σ and σ' respectively. Therefore, $d_{FR}(\sigma, \sigma^*) = \sum_{c\in A} f(c)$, $d_{FR}(\sigma', \sigma^*) = \sum_{c\in A} f'(c)$, $d_{MD}(\sigma, \sigma^*) = \max_{c\in A} f(c)$, $d_{MD}(\sigma', \sigma^*) = \max_{c\in A} f'(c)$. We want to show that $d_{FR}(\sigma', \sigma^*) \geq d_{FR}(\sigma, \sigma^*)$ and $d_{MD}(\sigma', \sigma^*) \geq d_{MD}(\sigma, \sigma^*)$. Note that $f(c) = f'(c)$ for any $c \in A\backslash\{a, b\}$ since exchanging a and b does not change the positions of the other alternatives. Thus, for the footrule distance it is sufficient to show that $f'(a) + f'(b) \geq f(a) + f(b)$, and for the maximum displacement distance it is sufficient to show that $\max(f'(a), f'(b)) \geq \max(f(a), f(b))$. We consider three cases.

Case 1. Let $\sigma(a) \leq \sigma^*(a)$ and $\sigma(a) < \sigma(b) \leq \sigma^*(b)$ as shown in Figure 1(a). Let $x = \sigma(b) - \sigma(a)$. From the figure, it is easy to verify that by exchanging a and b in σ, b moves farther from $\sigma^*(b)$ by exactly x and a may move closer to $\sigma^*(a)$ but by at most x. Formally,

$$f'(b) - f(b) = (\sigma^*(b) - \sigma'(b)) - (\sigma^*(b) - \sigma(b)) = \sigma(b) - \sigma(a) = x, \tag{4}$$

where the second transition follows since $\sigma'(b) = \sigma(a)$. Similarly,

$$f'(a)-f(a) = |\sigma'(a)-\sigma^*(a)|-|\sigma^*(a)-\sigma(a)| \geq -|\sigma'(a)-\sigma(a)| = -(\sigma(b)-\sigma(a)) = -x, \tag{5}$$

where the second transition is due to triangle inequality and the third transition follows since $\sigma'(a) = \sigma(b)$. Adding Equations (4) and (5), we get that $f'(a) + f'(b) \geq f(a) + f(b)$. For the maximum displacement distance, note that the displacement of b in σ' is $\sigma^*(b) - \sigma(a)$, which is clearly at least as much as the displacements of a and b in σ. Formally,

$$\max(f'(a), f'(b)) = \max(|\sigma(b) - \sigma^*(a)|, \sigma^*(b) - \sigma(a)) = \sigma^*(b) - \sigma(a),$$

where the second transition follows since $\sigma(a) < \sigma(b) \leq \sigma^*(b)$ and $\sigma(a) \leq \sigma^*(a) < \sigma^*(b)$. Also, $\sigma^*(b) > \sigma^*(a)$ implies $\sigma^*(b) - \sigma(a) > \sigma^*(a) - \sigma(a) = f(a)$ and $\sigma(a) < \sigma(b)$ implies

$\sigma^*(b)-\sigma(a) > \sigma^*(b)-\sigma(b) = f(b)$. Hence, $\max(f'(a), f'(b)) > f(a)$ and $\max(f'(a), f'(b)) > f(b)$, so $\max(f'(a), f'(b)) > \max(f(a), f(b))$, as required.

Case 2. Let $\sigma(a) \leq \sigma^*(a)$ and $\sigma(b) < \sigma^*(b)$ as shown in Figure 1(b). Let $x = \sigma^*(a) - \sigma(a)$, $y = \sigma(b) - \sigma^*(b)$ and $z = \sigma^*(b) - \sigma(b)$. Then, it is clear that $f(a) = x$, $f(b) = y$, $f'(a) = z + y$, and $f'(b) = z + x$. It is trivial to check that $f'(a) + f'(b) \geq f(a) + f(b)$ and $\max(f'(a), f'(b)) \geq \max(f(a), f(b))$.

Case 3. Let $\sigma(a) \geq \sigma^*(a)$ and $\sigma(b) > \sigma(a)$ as shown in Figure 1(c). This case is very similar to Case 1. For the footrule distance, alternative a (rather than b) moves away by exactly $x = \sigma(b) - \sigma(a)$ and alternative b (rather than a) may move closer by at most x. Similarly, for the maximum displacement distance, alternative a (rather than b) has greater displacement after the exchange compared to the displacements of both alternatives before the exchange. Hence, we again have $f'(a) + f'(b) \geq f(a) + f(b)$ and $\max(f'(a), f'(b)) \geq \max(f(a), f(b))$.

From the above three cases, it follows that the mapping which exchanges a and b in a ranking σ with $a \succ_\sigma b$ is weakly-distance-increasing with respect to σ^* for both the footrule distance and the maximum displacement distance. Similarly to the proof of Theorem 5.9, we can view this mapping to be a weakly-distance-increasing bijection from $\mathcal{L}_{a \succ b}(A)$ to $\mathcal{L}_{b \succ a}(A)$, which shows that both distances are MC (using the equivalent representation of MC distances given in Lemma 5.2). For proving that both distances are PC, we use the same technique that we used in the proof of Theorem 5.9. We want to give a bijection from $\mathcal{S}_j(a)$ to $\mathcal{S}_j(b)$ which is distance-increasing. We map every ranking where both a and b are in the first j position to itself, which does not change the distance of the ranking from σ^*. We map any ranking σ where $\sigma(a) \leq j$ and $\sigma(b) > j$ to the ranking where alternatives a and b are swapped, which does not decrease the distance from σ^* as shown in the three cases above. Therefore, this mapping is at least weakly-distance-increasing. We need to show that it is distance-increasing. That is, the distance must increase for some $\sigma \in \mathcal{S}_j(a)$. Clearly, the identity map does not change the distance.

Thus, we need to show that for any $\sigma^* \in \mathcal{L}(A)$, $a, b \in A$ such that $a \succ_{\sigma^*} b$ and $j \in \{1, \ldots, m-1\}$, there exists a ranking σ such that $\sigma(a) \leq j$, $\sigma(b) > j$ and $d(\sigma_{a \leftrightarrow b}, \sigma^*) > d(\sigma, \sigma^*)$ for both $d = d_{FR}$ and $d = d_{MD}$ (note that we can in principle show different rankings for d_{FR} and d_{MD}, but we give a stronger example that works for both distances). For this, we again take two cases. Let a_1 and a_m denote the first and the last alternatives in σ^*.

Case 1. If $1 \leq j < \sigma^*(b)$, then consider the ranking σ where $\sigma(a) = \sigma^*(a_1) = 1$, $\sigma(a_1) = \sigma^*(a)$ and $\sigma(c) = \sigma^*(c)$ for every $c \in A \setminus \{a_1, a\}$. In particular, $\sigma(b) = \sigma^*(b)$. First, note that $\sigma(a) = 1 \leq j$ and $\sigma(b) = \sigma^*(b) > j$. Now, it is easy to verify that $f(a) = \sigma^*(a) - 1$, $f(b) = 0$, $f'(a) = \sigma^*(b) - \sigma^*(a)$, and $f'(b) = \sigma^*(b) - 1$. Therefore,

$$f'(a)+f'(b) = 2\cdot\sigma^*(b)-\sigma^*(a)-1 \geq 2\cdot(\sigma^*(a)+1)-\sigma^*(a)-1 = \sigma^*(a)+1 > \sigma^*(a)-1 = f(a)+f(b).$$

Therefore, the footrule distance strictly increases. For the maximum displacement distance, note that in the original ranking, only alternatives a and a_1 are displaced, hence $d_{MD}(\sigma, \sigma^*) = f(a)$. Also, in the final ranking, only alternatives a_1, a and b are displaced, among which alternative b has the highest displacement. Thus, $d_{MD}(\sigma', \sigma^*) = f'(b)$. Finally, note that $f'(b) = \sigma^*(b) - 1 > \sigma^*(a) - 1 = f(a)$. Hence, the maximum displacement distance also strictly increases.

Case 2. If $\sigma^*(b) \leq j < m$, then consider the ranking σ where $\sigma(b) = \sigma^*(a_m) = m$, $\sigma(a_m) = \sigma^*(b)$ and $\sigma(c) = \sigma^*(c)$ for every $c \in A \setminus \{a_m, b\}$. In particular, $\sigma(a) = \sigma^*(a)$. Again note that $\sigma(a) = \sigma^*(a) < \sigma^*(b) \leq j$ and $\sigma(b) = m > j$. Now, it is easy to verify that $f(a) = 0$, $f(b) = m - \sigma^*(b)$, $f'(a) = m - \sigma^*(a)$, and $f'(b) = \sigma^*(b) - \sigma^*(a)$. Therefore,

$$f'(a) + f'(b) = m + \sigma^*(b) - 2 \cdot \sigma^*(a) > m + \sigma^*(b) - 2 \cdot \sigma^*(b) = m - \sigma^*(b) = f(a) + f(b).$$

Therefore, the footrule distance strictly increases. For the maximum displacement distance, note that in the original ranking, only alternatives b and a_m are displaced, hence $d_{MD}(\sigma, \sigma^*) = f(b)$. Also, in the final ranking, only alternatives a_m, a and b are displaced, among which alternative a has the highest displacement. Thus, $d_{MD}(\sigma', \sigma^*) = f'(a)$. Finally, note that $f'(a) = m - \sigma^*(a) > m - \sigma^*(b) = f(b)$. Hence, the maximum displacement distance also strictly increases.

From both cases, it is clear that for any $\sigma^* \in \mathcal{L}(A)$, $a, b \in A$ such that $a \succ_{\sigma^*} b$ and $j \in \{1, \ldots, m-1\}$, the bijection we constructed from $\mathcal{S}_j(a)$ to $\mathcal{S}_j(b)$ is distance-increasing. Hence, using the equivalent representation of PC distances given in Lemma 5.5, it follows that both the footrule distance and the maximum displacement distance are PC, as required. $\square$ (Theorem 5.9)

E.4. The case of the Cayley distance and the Hamming distance

In the discussion, we mentioned that we exclude distances such as the Cayley distance and the Hamming distance from our analysis because even the most prominent voting rules such as plurality are not accurate in the limit for any noise models that are monotonic with respect to these distances. We show that this is indeed the case. First, let us define these two distances.

The Cayley Distance. The Cayley distance between two rankings measures the minimum number of (possibly non-adjacent) swaps of alternatives required to convert one ranking into the other. Let us denote it by d_{CY}.

The Hamming Distance. The hamming distance between two rankings is defined as the number of positions where rankings differ. Formally, $d_{HM}(\sigma_1, \sigma_2) = \sum_{a \in A} \mathbb{1}[\sigma_1(a) \neq \sigma_2(a)]$.

Let $A = \{a, b, c\}$ be the set of alternatives. Let $\sigma^* = (a \succ b \succ c)$ be the true ranking. The following table describes the various possible rankings over these three alternatives and their Cayley distances as well as Hamming distances from σ^*.

	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6
	a	a	b	c	b	c
	b	c	a	b	c	a
	c	b	c	a	a	b
d_{HM}	0		2		3	
d_{CY}	0		1		2	

Recall that for any d-monotonic noise model, $d(\sigma, \sigma^*) = d(\tau, \sigma^*)$ implies $\Pr[\sigma|\sigma^*] = \Pr[\tau|\sigma^*]$. Therefore, any noise model that is monotonic with respect to the Hamming distance or the Cayley distance would assign equal probabilities to rankings σ_3 and σ_4, and to rankings σ_5 and σ_6, making the probabilities of alternatives b and c coming first in a random vote equal. It follows that with probability $1/2$, alternative c would be ranked higher than alternative b by plurality. Thus, plurality is not accurate in the limit with respect to any noise model that is monotonic with respect to either the Hamming distance or the Cayley distance.

F. PD-C RULES AND PC DISTANCES: THE CHARACTERIZATION

First, we prove the equivalent representation of PC distances (Lemma 5.5).

PROOF OF LEMMA 5.5. For the forward direction, fix any PC distance d, base ranking σ, alternatives $a, b \in A$ such that $a \succ_\sigma b$ and $j \in \{1, \ldots, m-1\}$. Let us sort all rankings in $\mathcal{S}_j(a)$ and $\mathcal{S}_j(b)$ in the increasing order of their distance from σ, and map the i^{th} ranking (in the sorted order) in $\mathcal{S}_j(a)$ to the i^{th} ranking in $\mathcal{S}_j(b)$. First, we show that this mapping is weakly-distance-increasing with respect to σ. Suppose for contradiction that (say) the i^{th} ranking of $\mathcal{S}_j(a)$ at distance k from σ is mapped to the i^{th} ranking of $\mathcal{S}_j(b)$ at distance $k' < k$ from σ. Then clearly, $|\mathcal{S}_j^{k'}(\sigma,a)| < i$ and $|\mathcal{S}_j^{k'}(\sigma,b)| \geq i$, so $|\mathcal{S}_j^{k'}(\sigma,b)| > |\mathcal{S}_j^{k'}(\sigma,a)|$ which is a contradiction since we assumed the distance to be PC. Now we show that this mapping takes at least one ranking to a ranking at strictly greater distance from σ. Since d is PC, there exists some $k^* \in \mathbb{N}$ such that $|\mathcal{S}_j^{k^*}(\sigma,a)| > |\mathcal{S}_j^{k^*}(\sigma,b)|$. Consider the largest i such that i^{th} ranking of $\mathcal{S}_j(a)$ is at distance at most k^* from σ. If this ranking is mapped to a ranking at equal distance (hence at most k^*) from σ, then we would have $|\mathcal{S}_j^{k^*}(\sigma,b)| \geq |\mathcal{S}_j^{k^*}(\sigma,a)|$, which is a contradiction. Hence, this ranking is mapped to a ranking at strictly greater distance from σ.

In the other direction, take any distance function d. Suppose for every $\sigma \in \mathcal{L}(A)$, $a, b \in A$ such that $a \succ_\sigma b$ and $j \in \{1,\ldots,m-1\}$, there exists a bijection $f : \mathcal{S}_j(a) \to \mathcal{S}_j(b)$ that is weakly-distance-increasing with respect to σ and maps at least one ranking to a ranking at strictly greater distance from σ. Fix any particular $\sigma \in \mathcal{L}(A)$, $a, b \in A$ such that $a \succ_\sigma b$ and $j \in \{1,\ldots,m-1\}$. First, for any $k \in \mathbb{N}$ we have $\mathcal{S}_j^k(b) \subseteq \{f(\tau) | \tau \in \mathcal{S}_j^k(a)\}$, so $|\mathcal{S}_j^k(a)| \geq |\mathcal{S}_j^k(b)|$. Further, take the ranking $\tau \in \mathcal{S}_j(a)$ for which $d(f(\tau),\sigma) > d(\tau,\sigma)$. Let $k^* = d(\tau,\sigma)$. Then, $f(\tau) \in \{f(\gamma) | \gamma \in \mathcal{S}_j^{k^*}(a)\}$ but $f(\tau) \notin \mathcal{S}_j^{k^*}(b)$. Hence, $|\mathcal{S}_j^{k^*}(b)| < |\{f(\gamma)|\gamma \in \mathcal{S}_j^{k^*}(a)\}| = |\mathcal{S}_j^{k^*}(a)|$, as required. Since this holds for every $\sigma \in \mathcal{L}(A)$, $a, b \in A$ such that $a \succ_\sigma b$ and $j \in \{1,\ldots,m-1\}$, the distance is PC. $\square$ (Lemma 5.5)

We are now ready to prove our characterization result (Theorem 5.6).

PROOF OF THEOREM 5.6. First, we assume that d is PC and show that all PD-c rules are d-monotone-robust. Consider any d-monotonic noise model G; we wish to show that all PD-c rules are accurate in the limit for G. Fix any true ranking $\sigma^* \in \mathcal{L}(A)$, $a, b \in A$ such that $a \succ_{\sigma^*} b$, and $j \in \{1,\ldots,m-1\}$.

Since d is PC, Lemma 5.5 implies that there exists a bijection $f : \mathcal{S}_j(a) \to \mathcal{S}_j(b)$ such that i) for every $\sigma \in \mathcal{S}_j(a)$, $d(f(\sigma),\sigma^*) \geq d(\sigma,\sigma^*)$, hence $\Pr[\sigma|\sigma^*] \geq \Pr[f(\sigma)|\sigma^*]$, and ii) for some $\sigma \in \mathcal{S}_j(a)$, $d(f(\sigma),\sigma^*) > d(\sigma,\sigma^*)$, hence $\Pr[\sigma|\sigma^*] > \Pr[f(\sigma)|\sigma^*]$. Recall that f is a bijection, hence its range is the whole of $\mathcal{S}_j(b)$. Now we sum over all $\sigma \in \mathcal{S}_j(a)$ (similarly to the proof of Theorem 5.3) and get that the probability that a appears in first j positions is strictly greater than the probability that b appears in the first j positions in a random vote. It follows that given infinitely many samples from G, a would appear in first j positions in more votes than b does. Since this holds for all $j \in \{1,\ldots,m-1\}$, there would be an edge from a to b in the PD graph with probability 1. Further, since this holds for all $a, b \in A$, the PD graph would reduce to σ^* with probability 1. Hence, any PD-c rule would output σ^* with probability 1, as required.

In the other direction, consider any distance function d that is not PC. We show that there exists a PD-c rule that is not accurate in the limit for some d-monotonic noise model G. Since d is not PC, there exist $\sigma^* \in \mathcal{L}(A)$, $a, b \in A$ with $a \succ_{\sigma^*} b$ and $j \in \{1,\ldots,m-1\}$ such that either i) there exists $k^* \in \mathbb{N}$ with $|\mathcal{S}_j^k(\sigma^*,a)| < |\mathcal{S}_j^k(\sigma^*,b)|$, or ii) for every $k \in \mathbb{N}$, $|\mathcal{S}_j^k(\sigma^*,a)| = |\mathcal{S}_j^k(\sigma^*,a)|$.

In case i), we construct the noise model G exactly as in the proof of Theorem 5.3. We define $M = \max_{\sigma \in \mathcal{L}(A)} d(\sigma,\sigma^*)$ and $T = 2 \cdot m! \cdot M$. Assign weights $w_\sigma = T - d(\sigma,\sigma^*)$ if $d(\sigma,\sigma^*) \leq k^*$ and $w_\sigma = M - d(\sigma,\sigma^*)$ otherwise. The noise model G would consequently assign $\Pr[\sigma|\sigma^*] = w_\sigma / \sum_{\tau \in \mathcal{L}(A)} w_\tau$. It follows that under G, the probability of a coming in first j positions of a random vote, i.e., $\sum_{\sigma \in \mathcal{S}_j(a)} \Pr[\sigma|\sigma^*]$, would be strictly less than the probability of b coming in first j positions in a random vote, i.e., $\sum_{\sigma \in \mathcal{S}_j(b)} \Pr[\sigma|\sigma^*]$. Thus, given infinitely many samples, b would appear more times in first j positions than a. This implies that with probability 1, there would be no edge from a to b in the PD graph. Therefore, with probability 1 the PD graph would not reduce to σ^*. We can easily construct a PD-c rule similarly to the proof of Theorem 5.3 that outputs a ranking σ whenever the PD graph reduces to σ, and outputs an arbitrary ranking with $b \succ a$ when the PM graph does not reduce to any ranking. With probability 1, such a rule would output a ranking where $b \succ a$. Hence, r is not accurate in the limit for G, as required.

Consider case ii). Since $|\mathcal{S}_j^k(\sigma^*, a)| = |\mathcal{S}_j^k(\sigma^*, a)|$ for every $k \in \mathbb{N}$, we also have $|\mathcal{S}_j^k(\sigma^*, a)| - |\mathcal{S}_j^{k-1}(\sigma^*, a)| = |\mathcal{S}_j^k(\sigma^*, a)| - |\mathcal{S}_j^{k-1}(\sigma^*, b)|$ for every $k \in \mathbb{N}$, i.e., the number of rankings at distance *exactly* k in which a is in first j positions is equal to the number of rankings at distance *exactly* k where b is in first j positions. Now consider any d-monotonic noise model G. Since it assigns equal probabilities to rankings at equal distances, we see that the probability of a coming in first j positions of a random vote, i.e., $\sum_{\sigma \in \mathcal{S}_j(a)} \Pr[\sigma|\sigma^*]$, would be exactly equal to the probability of b coming in first j positions in a random vote, i.e., $\sum_{\sigma \in \mathcal{S}_j(b)} \Pr[\sigma|\sigma^*]$. Therefore, with probability $1/2$, a would not come in first j positions more times than b, in which case there would be no edge from a to b in the PD graph and the PD graph would not reduce to σ^*. Now we can easily construct a PD-c rule r such that r outputs σ whenever the PD graph reduces to σ and outputs a fixed ranking $\sigma' \neq \sigma^*$ whenever the PD graph does not reduce to any ranking. Since the PD graph does not reduce to σ^* with probability at least $1/2$, r is clearly not accurate in the limit under such a noise model. Hence, r is not d-monotone-robust, as required. $\square$ (Theorem 5.6)