Modal Ranking: A Uniquely Robust Voting Rule

Ioannis Caragiannis
University of Patras
caragian@ceid.upatras.gr

Ariel D. Procaccia
Carnegie Mellon University
arielpro@cs.cmu.edu

Nisarg Shah
Carnegie Mellon University
nkshah@cs.cmu.edu

Abstract

Motivated by applications to crowdsourcing, we study voting rules that output a correct ranking of alternatives by quality from a large collection of noisy input rankings. We seek voting rules that are supremely robust to noise, in the sense of being correct in the face of any "reasonable" type of noise. We show that there is such a voting rule, which we call the *modal ranking* rule. Moreover, we establish that the modal ranking rule is the unique rule with the preceding robustness property within a large family of voting rules, which includes a slew of well-studied rules.

Introduction

The emergence of crowdsourcing platforms and human computation systems (Law and von Ahn 2011) motivates a reexamination of an approach to voting that dates back to the Marquis de Condorcet (1785). He suggested that voters should be viewed as noisy estimators of a ground truth — a ranking of the candidates by their true quality. A *noise* model governs how voters make mistakes. For example, under the noise model suggested by Condorcet — also known today as the Mallows (1957) noise model — each voter ranks each pair of alternatives in the correct order with probability $p > 1/2$, and in the wrong order with probability $1 - p$ (roughly speaking). This specific noise model is quite unrealistic, and, more generally, the very idea of objective noise is arguable in the context of political elections, where opinions are subjective and there is no ground truth. However, the noisy voting setting is a perfect fit for crowdsourcing, where objective estimates provided by workers — often as votes (Little et al. 2010; Mao, Procaccia, and Chen 2013) — must be aggregated.

From this viewpoint, Condorcet and, more eloquently, Young (1988), argued that a voting rule — which aggregates input rankings into a single output ranking — should output the ranking that is most likely to be the ground truth ranking, under the given noise model. This approach has inspired a significant number of recent papers by AI researchers (Conitzer and Sandholm 2005; Conitzer, Rognlie, and Xia 2009; Elkind, Faliszewski, and Slinko 2010; Xia, Conitzer, and Lang 2010; Xia and Conitzer 2011; Lu and Boutilier 2011; Procaccia, Reddi, and Shah 2012; Mao, Procaccia, and Chen 2013), some of which aim to design voting rules that are *maximum likelihood estimators (MLEs)* specifically for crowdsourcing settings.

But the maximum likelihood estimation requirement may be too stringent. Caragiannis et al. (2013) point out that a voting rule may be an MLE for a specific noise model, but in realistic settings the noise can take unpredictable forms (Mao, Procaccia, and Chen 2013). Instead, they propose the following robustness property, called *accuracy in the limit*: as the number of votes grows, the voting rule should output the ground truth ranking with high probability, i.e., with probability approaching one.[1] This allows a single voting rule to be robust against multiple noise models. Moreover, the focus on a large number of votes is natural in the context of crowdsourcing systems — the whole point is to aggregate information provided by a massive crowd!

For example, social networks are enabling organizations to solicit noisy information from millions of users. Indeed, think of a technology company that asks fans to rank product prototypes by their perceived chance of success. While a large company can expect millions of votes, these votes are noisy and the type of noise is unpredictable.

In this paper, we seek voting rules that are robust against such unpredictable noise. Our research challenge is to

> *... find voting rules that are robust (in the accuracy in the limit sense) against* any *"reasonable" noise model.*

Our results. We give a rather clear-cut solution to the preceding research challenge: There is a voting rule that is robust against any "reasonable" noise model, and it is unique within a huge family of voting rules. We call this supremely robust voting rule the *modal ranking* rule. Given a collection of input rankings, the modal ranking rule simply selects the most frequent ranking as the output. To the best of our knowledge, this strikingly basic voting rule has not received any attention in the literature, and for good reason: when the number of voters is not huge compared to the number of alternatives, it is likely that every ranking would appear at most once, so the modal ranking rule does not provide

[1] In statistics, this property is known as *consistency*, but we avoid this terminology as it has completely different interpretations in social choice theory.

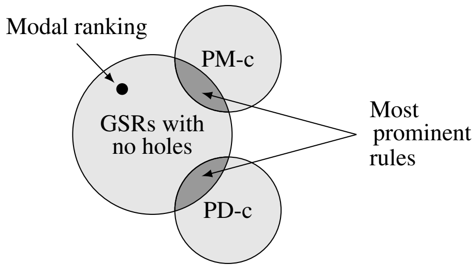

Figure 1: The modal ranking rule is uniquely robust within the union of three families of rules.

any useful guidance. However, when the number of voters is very large, the modal ranking rule is quite sensible; we will prove this intuitive claim formally.

To better understand this result (still on an informal level), we need to clarify two points: What do we mean by "reasonable" noise model? And what is the huge family of voting rules? Starting from the noise model, we employ some additional notions introduced by Caragiannis et al. (2013). We are interested in noise models that are *d-monotone* with respect to a distance function d on rankings, in the sense that the probability of a ranking increases as its distance according to d from the ground truth ranking decreases. A voting rule that is accurate in the limit with respect to any d-monotone noise model is said to be *monotone-robust* with respect to d. So, slightly more formally, the requirement is that the rule be monotone-robust with respect to any distance metric d.

Regarding the family of voting rules in which we prove that modal ranking is the unique robust rule, it is formed by the union of three families of rules: generalized scoring rules (GSRs) (Xia and Conitzer 2008; 2009; Xia 2013) with the no holes property, pairwise majority consistent (PM-c) rules, and position dominance consistent (PD-c) rules (Caragiannis, Procaccia, and Shah 2013). GSRs are a large family of voting rules that is known to capture almost all commonly-studied voting rules. Theorem 1 asserts that a GSR with "no holes" is monotone-robust with respect to all distance metrics if and only if it is the modal ranking rule. The no holes property is a technical restriction, but in Theorem 2 we show that it is quite mild by establishing that all prominent rules that are known to be GSRs have the no holes property.

PM-c and PD-c rules together also contain most prominent voting rules (Caragiannis, Procaccia, and Shah 2013). These two families are disjoint, and neither is contained in the family of GSRs. Theorem 3 asserts that PM-c and PD-c rules do not have the desired robustness property, thereby further extending the scope of the modal ranking rule's uniqueness to all rules that are PM-c or PD-c but not GSRs (with no holes). See Figure 1 for a Venn diagram that illustrates the relation between these families of rules.

Related work. Our paper is most closely related to the work of Caragiannis, Procaccia, and Shah (2013), who introduced the classes of PM-c and PD-c rules as well as the notions of d-monotone noise models, accuracy in the limit, and monotone-robustness. Their main result is a characterization of the distance metrics d for which all PM-c and PD-c rules are monotone-robust. In other words, they *fixed the family of voting rules* to be PM-c or PD-c rules, and asked which distance metrics induce noise models for which *all* the rules in these families are robust. While the answer is a family of distance metrics that contains three popular distance metrics, it does not contain several other prominent distance metrics — moreover, it is by no means clear that natural distance metrics are the ones that induce the noise one encounters in practice. In contrast, instead of fixing the family of rules, we *fix the family of distances* to be all possible distance metrics d, and characterize the "family" of voting rules that are monotone-robust with respect to any d (this family turns out to be a singleton).

On a technical level, we view vectors of rankings as points in $\mathbb{Q}^{m!}$ ($m!$ is the number of possible rankings), where each coordinate represents the fraction of times a ranking appears in the profile. This geometric approach to the analysis of voting rules was initiated by Young (1975), and used by various other authors (Saari 1995; 2008; Xia and Conitzer 2009; Conitzer, Rognlie, and Xia 2009; Obraztsova et al. 2013; Mossel, Procaccia, and Rácz 2013).

Preliminaries

Let A be the set of alternatives, where $|A| = m$. Let $\mathcal{L}(A)$ be the set of rankings (linear orders) over A, and $\mathcal{D}(\mathcal{L}(A))$ be the set of distributions over $\mathcal{L}(A)$. A vote σ is a ranking in $\mathcal{L}(A)$, and a profile π is a collection of votes. A voting rule (sometimes also known as a "rank aggregation rule") is formally a deterministic (resp., randomized) *social welfare function (SWF)* that maps every profile to a ranking (resp., a distribution over rankings). We focus on randomized SWFs. Deterministic SWFs are a special case where the output distributions are centered at a single ranking. In this paper we do not study social choice functions (SCFs), which map each profile to a (single) selected alternative.

Families of SWFs. In order to capture many SWFs simultaneously, our results employ the definitions of three broad families of SWFs.

- *PM-c rules* (Caragiannis, Procaccia, and Shah 2013): For a profile π, the *pairwise-majority (PM) graph* is a directed graph whose vertices are the alternatives, and there exists an edge from $a \in A$ to $b \in A$ if a strict majority of the voters prefer a to b. A randomized SWF f is called *pairwise-majority consistent (PM-c)* if for every profile π with a complete acyclic PM graph whose vertices are ordered according to $\sigma \in \mathcal{L}(A)$, we have $\Pr[f(\pi) = \sigma] = 1$.
- *PD-c rules* (Caragiannis, Procaccia, and Shah 2013): In a profile π, alternative a is said to *position-dominate* alternative b if for every $k \in \{1, \ldots, m-1\}$, (strictly) more voters rank a in first k positions than b. The *position-dominance (PD) graph* is a directed graph whose vertices are the alternatives, and there exists an edge from a to b if a position-dominates b. A randomized SWF f is called *position-dominance consistent (PD-c)* if for every profile π with a complete acyclic PD graph whose vertices are ordered according to $\sigma \in \mathcal{L}(A)$, we have $\Pr[f(\pi) = \sigma] = 1$.

- *GSRs* (Xia and Conitzer 2008): We say that two vectors $y, z \in \mathbb{R}^k$ are *equivalent* (denoted $y \sim z$) if for every $i, j \in [k]$ we have $y_i \geq y_j \Leftrightarrow z_i \geq z_j$. We say that a function $g : \mathbb{R}^k \to \mathcal{D}(\mathcal{L}(A))$ is *compatible* if $y \sim z$ implies $g(y) = g(z)$. A *generalized scoring rule (GSR)* is given by a pair of functions (f, g), where $f : \mathcal{L}(A) \to \mathbb{R}^k$ maps every ranking to a k-dimensional vector, a compatible function $g : \mathbb{R}^k \to \mathcal{D}(\mathcal{L}(A))$ maps every k-dimensional vector to a distribution over rankings, and the output of the rule on a profile $\pi = (\sigma_1, \ldots, \sigma_n)$ is given by $g\left(\sum_{i=1}^n f(\sigma_i)\right)$. GSRs are characterized by two social choice axioms (Xia and Conitzer 2009), and have interesting connections to machine learning (Xia 2013). While GSRs were originally introduced as deterministic SCFs, the definition naturally extends to (possibly) randomized SWFs.

Noise models. A *noise model* G is a collection of distributions over rankings. For every $\sigma^* \in \mathcal{L}(A)$, $G(\sigma^*)$ denotes the distribution from which noisy estimates are generated when the ground truth is σ^*. The probability of sampling $\sigma \in \mathcal{L}(A)$ from this distribution is denoted by $\Pr_G[\sigma; \sigma^*]$.

In order to rule out noise models that are completely outlandish, we focus on d-monotonic noise models with respect to a distance metric d, using definitions from the work of Caragiannis et al. (2013). In more detail, a *distance metric* over $\mathcal{L}(A)$ is a function $d(\cdot, \cdot)$ that satisfies the following properties for all $\sigma, \sigma', \sigma'' \in \mathcal{L}(A)$:

- $d(\sigma, \sigma') \geq 0$, and $d(\sigma, \sigma') = 0$ if and only if $\sigma = \sigma'$.
- $d(\sigma, \sigma') = d(\sigma', \sigma)$.
- $d(\sigma, \sigma'') + d(\sigma'', \sigma') \geq d(\sigma, \sigma')$.

A noise model G is called *d-monotone* for a distance metric d if for all $\sigma, \sigma', \sigma^* \in \mathcal{L}(A)$, $\Pr_G[\sigma; \sigma^*] \geq \Pr_G[\sigma'; \sigma^*]$ if and only if $d(\sigma, \sigma^*) \leq d(\sigma', \sigma^*)$. That is, the closer a ranking is to the ground truth, the higher its probability.

Robust SWFs. We are interested in SWFs that can recover the ground truth from a large number of i.i.d. noisy estimates. Formally, an SWF f is called *accurate in the limit with respect to a noise model G* if, given an arbitrarily large number of samples from G with any ground truth σ^*, the rule outputs σ^* with arbitrarily high accuracy. That is, for every $\sigma^* \in \mathcal{L}(A)$, $\lim_{n\to\infty} \Pr[f(\pi^n) = \sigma^*] = 1$, where π^n denotes a profile consisting of n i.i.d. samples from $G(\sigma^*)$. A voting rule f is called *monotone-robust* with respect to a distance metric d if it is accurate in the limit for all d-monotonic noise models.

Modal Ranking is Unique Within GSRs

In this section, we characterize the modal ranking rule — which selects the most common ranking in a given profile — as the unique rule that is monotone-robust with respect to all distance metrics, among a wide sub-family of GSRs. For this, we use a geometric equivalent of GSRs introduced by Mossel, Procaccia, and Rácz (2013) called "hyperplane rules". Like GSRs, hyperplane rules were also originally defined as deterministic SCFs. Below, we give the natural extension of the definition to (possibly) randomized SWFs.

Given a profile π, let x^π_σ denote the fraction of times the ranking $\sigma \in \mathcal{L}(A)$ appears in π. Hence, the point $x^\pi = (x^\pi_\sigma)_{\sigma \in \mathcal{L}(A)}$ lies in a probability simplex $\Delta^{m!}$. This allows us to use rankings from $\mathcal{L}(A)$ to index the $m!$ dimensions of every point in $\Delta^{m!}$. Formally,

$$\Delta^{m!} = \left\{ x \subseteq \mathbb{Q}^{m!} \;\middle|\; \sum_{\sigma \in \mathcal{L}(A)} x_\sigma = 1 \right\}.$$

Importantly, note that $\Delta^{m!}$ contains only points with rational coordinates. Weights $w_\sigma \in \mathbb{R}$ for all $\sigma \in \mathcal{L}(A)$ define a hyperplane H where $H(x) = \sum_{\sigma \in \mathcal{L}(A)} w_\sigma \cdot x_\sigma$ for all $x \in \Delta^{m!}$. This hyperplane divides the simplex into three regions; the set of points on each side of the hyperplane, and the set of points on the hyperplane.

Definition 1 (Hyperplane Rules)**.** A hyperplane rule is given by $r = (\mathcal{H}, g)$, where $\mathcal{H} = \{H_i\}_{i=1}^l$ is a finite set of hyperplanes, and $g : \{+, 0, -\}^l \to \mathcal{D}(\mathcal{L}(A))$ is a function that takes as input the signs of all the hyperplanes at a point and returns a distribution over rankings. Thus, $r(\pi) = g(\mathtt{sgn}(\mathcal{H}(x^\pi)))$, where

$$\mathtt{sgn}(\mathcal{H}(x^\pi)) = (\mathtt{sgn}(H_1(x^\pi)), \ldots, \mathtt{sgn}(H_l(x^\pi))),$$

and $\mathtt{sgn} : \mathbb{R} \to \{+, 0, -\}$ is the sign function given by

$$\mathtt{sgn}(x) = \begin{cases} + & x > 0 \\ 0 & x = 0 \\ - & x < 0 \end{cases}$$

Next, we state the equivalence between hyperplane rules and GSRs in the case of randomized SWFs. This equivalence was established by Mossel et al. (2013) for deterministic SCFs; it uses the output of a given GSR for each set of compatible vectors to construct the output of its corresponding hyperplane rule in each region, and vice-versa. Simply changing the output of the g functions of both the GSR and the hyperplane rule from a winning alternative (for deterministic SCFs) to a distribution over rankings (for randomized SWFs) and keeping the rest of the proof intact shows the equivalence for randomized SWFs.

Lemma 1. *For randomized social welfare functions, the class of generalized scoring rules coincides with the class of hyperplane rules.*

We impose a technical restriction on GSRs that has a clear interpretation under the geometric hyperplane equivalence. Intuitively, it states that if the rule outputs the same ranking (without ties) almost everywhere around a point x^π in the simplex, then the rule must output the same ranking (without ties) on π as well. More formally, consider the regions in which the simplex is divided by a set of hyperplanes $\mathcal{H}$. We say that a region is *interior* if none of its points lie on any of the hyperplanes in $\mathcal{H}$, that is, if for every point x in the region, $\mathtt{sgn}(\mathcal{H}(x))$ does not contain any zeros.

For $x \in \Delta^{m!}$, let

$$\mathcal{S}(x) = \{y \in \Delta^{m!} | \forall \sigma \in \mathcal{L}(A), x_\sigma = 0 \Rightarrow y_\sigma = 0\}$$

denote the subspace of points that are zero in every coordinate where x is zero. We say that an interior region is *adjacent* to x if its intersection with $\mathcal{S}(x)$ contains points arbitrarily close to x.

Definition 2 (No Holes Property)**.** We say that a hyperplane rule (generalized scoring rule) has no holes if it outputs a ranking σ with probability 1 on a profile π whenever it outputs σ with probability 1 in all interior regions adjacent to x^π.

When this property is violated, we have a point x^π such that the output of the rule on x^π is different from the output of the rule almost everywhere around x^π, creating a hole at x^π. We later show (Theorem 2) that the no holes property is a very mild restriction on GSRs.

We are now ready to formally state our main result.

Theorem 1. *Let r be a (possibly) randomized generalized scoring rule without holes. Then, r is monotone-robust with respect to all distance metrics if and only if r coincides with the modal ranking rule on every profile with no ties (i.e., r outputs the most frequent ranking with probability 1 on every profile where it is unique).*

Before proving the theorem, we wish to point out three subtleties. First, our assumption of accuracy in the limit imposes a condition on the rule as the number of votes *goes to infinity*. This has to be translated into a condition on all *finite profiles*; we do this by leveraging the structure of generalized scoring rules.

Second, if there are several rankings that appear the same number of times, a monotone-robust rule can actually output any ranking with impunity, because in the limit this event happens with probability zero.

Third, every noise model G that is monotone with respect to some distance metric satisfies $\Pr_G[\sigma^*;\sigma^*] > \Pr_G[\sigma;\sigma^*]$ for all pairs of different rankings $\sigma, \sigma^* \in \mathcal{L}(A)$. It seems intuitive that the converse holds, i.e., if a noise model satisfies $\Pr_G[\sigma^*;\sigma^*] > \Pr_G[\sigma;\sigma^*]$ for all $\sigma \neq \sigma^*$ then there exists a distance metric d such that G is monotone with respect to d — but this is false. Hence, our condition asks for accuracy in the limit for noise models that are monotone with respect to some metric, instead of just assuming accuracy in the limit with respect to *all* noise models where the ground truth is the unique mode.

Proof of Theorem 1. Let r be a (possibly) randomized generalized scoring rule without holes. Using Lemma 1, we represent r as a hyperplane rule. Let $r = (\mathcal{H}, f)$ where $\mathcal{H} = \{H_i\}_{i=1}^l$ is the set of hyperplanes.

First, we show the simpler forward direction. Let r output the most frequent ranking with probability 1 on every profile where it is unique. We want to show that r is monotone-robust with respect to all distance metrics. Take a distance metric d, a d-monotonic noise model G, and a true ranking σ^*. We need to show that r outputs σ^* with probability 1 given infinitely many samples from $G(\sigma^*)$.

Note that d satisfies $d(\sigma^*, \sigma^*) = 0 < d(\sigma, \sigma^*)$ for all $\sigma \neq \sigma^*$. Hence, G must satisfy $\Pr_G[\sigma^*;\sigma^*] > \Pr_G[\sigma;\sigma^*]$ for all $\sigma \neq \sigma^*$. Now, given infinite samples from $G(\sigma^*)$, σ^* becomes the unique most frequent ranking with probability 1. Thus, r outputs σ^* with probability 1 in the limit, as required.

For the reverse direction, let r be d-monotone-robust for all distance metrics d. Take a profile π^* with a unique most frequent ranking σ^*. Recall that $x_\sigma^{\pi^*}$ denotes the fraction of times σ appears in π^* and note that $x_{\sigma^*}^{\pi^*} > x_\sigma^{\pi^*}$ for all $\sigma \neq \sigma^*$. We also denote by $X_\sigma^{\pi^*}$ the number of times σ appears in π^*.

The rest of the proof is organized in three steps. First, we define a distance metric d, a d-monotonic noise model G, and a true ranking. Second, we show that given samples from $G(\sigma^*)$, in the limit r outputs σ^* with probability 1 in every interior region adjacent to x^{π^*}. Finally, we use the no holes property of r to argue that $\Pr[r(\pi^*) = \sigma^*] = 1$.

Step 1: We define d as

$$d(\sigma, \sigma') = \begin{cases} \max(1, |X_\sigma^{\pi^*} - X_{\sigma'}^{\pi^*}|) & \text{if } \sigma \neq \sigma', \\ 0 & \text{otherwise.} \end{cases}$$

We claim that d is a distance metric. Indeed, the first two axioms are easy to verify. The triangle inequality $d(\sigma, \sigma') \leq d(\sigma, \sigma'') + d(\sigma'', \sigma')$ holds trivially if any two of the three rankings are equal. When all three rankings are distinct,

$$\begin{aligned} & d(\sigma, \sigma'') + d(\sigma'', \sigma') \\ & = \max(1, |X_\sigma^{\pi^*} - X_{\sigma''}^{\pi^*}|) + \max(1, |X_{\sigma''}^{\pi^*} - X_{\sigma'}^{\pi^*}|) \\ & \geq \max(1 + 1, |X_\sigma^{\pi^*} - X_{\sigma''}^{\pi^*}| + |X_{\sigma''}^{\pi^*} - X_{\sigma'}^{\pi^*}|) \\ & \geq \max(1, |X_\sigma^{\pi^*} - X_{\sigma'}^{\pi^*}|) = d(\sigma, \sigma'). \end{aligned}$$

Now, define the noise model G where

$$\Pr_G[\sigma;\sigma'] = \frac{1/(1 + d(\sigma, \sigma'))}{\sum_{\tau \in \mathcal{L}(A)} 1/(1 + d(\tau, \sigma'))} \text{ for } \sigma' \neq \sigma^*.$$

and $\Pr_G[\sigma;\sigma^*] = x_\sigma^{\pi^*}$. Note that G is trivially d-monotone for true rankings other than σ^*. Denoting the number of votes in π^* by n^*, since σ^* is the unique most frequent ranking, we have that $d(\sigma, \sigma^*) = n^*(x_{\sigma^*}^{\pi^*} - x_\sigma^{\pi^*})$ for all $\sigma \neq \sigma^*$. Hence, $\Pr_G[\sigma_1;\sigma^*] \geq \Pr_G[\sigma_2;\sigma^*]$ if and only if $d(\sigma_1, \sigma^*) \leq d(\sigma_2, \sigma^*)$ and G is also d-monotone for the true ranking σ^*. We conclude that G is a d-monotonic noise model.

Step 2: Let π_n denote a profile consisting of n i.i.d. samples from $G(\sigma^*)$. Since r is monotone-robust for every distance metric, we have

$$\lim_{n \to \infty} \Pr[r(\pi_n) = \sigma^*] = 1. \tag{1}$$

If π^* has only one ranking, then only that ranking will ever be sampled. Hence, we will have $\Pr[x^{\pi_n} = x^{\pi^*}] = 1$, and Equation (1) would imply that the rule must output σ^* with probability 1 on π^*.

Assume that π^* has at least two distinct votes. We want to show that r outputs σ^* with probability 1 in every interior region adjacent to x^{π^*}. As $n \to \infty$, the distribution of x^{π_n} tends to a Gaussian with mean x^{π^*} and concentrated on the hyperplane

$$\sum_{\sigma \in \mathcal{L}(A) | x_\sigma^{\pi^*} > 0} x_\sigma^{\pi_n} = 1.$$

This follows from the multivariate central limit theorem; see (Mossel, Procaccia, and Rácz 2013) for a detailed explanation. Note that the sum ranges only over the rankings that appear in π^* because in the distribution $G(\sigma^*)$, the probability of sampling a ranking σ that does not appear in π^* is zero.

Since the Gaussian lies in the subspace $\mathcal{S}(x^{\pi^*})$, we set the coordinates corresponding to rankings that do not appear in π^* to zero in all the hyperplanes, and remove the hyperplanes that become trivial. Hereinafter we only consider the rest of the hyperplanes, and the regions they form around x^{π^*}, all in the subspace $\mathcal{S}(x^{\pi^*})$.

If none of the hyperplanes pass through x^{π^*}, then there is a unique interior region K which actually contains x^{π^*} as its interior point. In this case, the limiting probability of x^{π_n} falling in K will be 1, as the Gaussian becomes concentrated around x^{π^*}. Thus, Equation (1) implies that r outputs σ^* with probability 1 in K, and therefore on π^*.

If there exists a hyperplane passing through x^{π^*}, then each interior region K adjacent to x^{π^*} is the intersection of finitely many halfspaces whose hyperplanes pass through x^{π^*}. Let $\overline{K}$ and $\overline{\mathcal{S}(x^{\pi^*})}$ denote the closures of K and $\mathcal{S}(x^{\pi^*})$ respectively in $\mathbb{R}^{m!}$.[2] Thus, $\overline{K}$ is a pointed convex cone with its apex at x^{π^*}, and must subtend a positive solid angle (in $\overline{\mathcal{S}(x^{\pi^*})}$) at its apex since the hyperplanes are distinct. By definition, the solid angle $\overline{K}$ forms at x^{π^*} is the fraction of volume (the Lebesgue measure in $\overline{\mathcal{S}(x^{\pi^*})}$) covered by $\overline{K}$ in a ball of radius ρ (again in $\overline{\mathcal{S}(x^{\pi^*})}$) centered at x^{π^*}, as $\rho \to 0$ (see, e.g., Section 2 in (Desario and Robins 2011)).

Since the Gaussian is symmetric in $\overline{\mathcal{S}(x^{\pi^*})}$ around x^{π^*}, and the limiting distribution of x^{π_n} converges to the Gaussian, the limiting probability of x^{π_n} lying in K is positive. This holds for every interior region K adjacent to x^{π^*}. Thus, Equation (1) again implies that r outputs σ^* with probability 1 in every interior region adjacent to x^{π^*}.

Step 3: Finally, because r has no holes and it outputs σ^* with probability 1 in every interior region adjacent to x^{π^*}, we conclude that r must also output σ^* with probability 1 on π^*. □(Proof of Theorem 1)

To complete the picture, we wish to show that the no holes condition that Theorem 1 imposes on GSRs is indeed unrestrictive, by establishing that many prominent voting rules (in the sense of receiving attention in the computational social choice literature) are GSRs with no holes. One issue that must be formally addressed is that the definitions of prominent voting rules typically do not address how ties are broken. For example, the plurality rule ranks the alternatives by their number of voters who rank them first; but what should we do in case of a tie? Below we adopt uniformly random tie-breaking, which is almost always used in political elections (e.g., by throwing dice or drawing cards in small municipal elections where ties are not unlikely to occur). From a theoretical point of view, randomized tie breaking is necessary in order to achieve neutrality with respect to the alternatives (Moulin 1983). In fact, we have proven the following theorem for a wide family of randomized tie-breaking schemes, but here we focus on uniformly random tie-breaking for ease of exposition.

Theorem 2. *Under uniformly random tie-breaking, all positional scoring rules (including plurality and Borda count), the Kemeny rule, single transferable vote (STV), Copeland's method, Bucklin's rule, the maximin rule, Slater's rule, and the ranked pairs method are generalized scoring rules without holes.*

The rather intricate proof of Theorem 2 appears in the full version of the paper.[3] The comprehensive list of GSRs with no holes includes all prominent rules that are known to be GSRs (Xia and Conitzer 2008; Mossel, Procaccia, and Rácz 2013) — suggesting that the no holes property does not impose a significant restriction beyond the assumption that the rule is a GSR. One prominent rule is conspicuously missing — the fascinating but peculiar Dodgson rule (Dodgson 1876), which is indeed not a GSR (Xia and Conitzer 2008).

Impossibility for PM-c and PD-c Rules

Theorem 1 establishes the uniqueness of the modal ranking rule within a large family of voting rules (GSRs with no holes). Next we further expand this result by showing that no PM-c or PD-c rule is monotone-robust with respect to all distance metrics. Thus, the modal ranking rule is the unique rule that is monotone-robust with respect to all distance metrics in the union of GSRs with no holes, PM-c rules, and PD-c rules. Crucially, as shown in Figure 1, the families of PM-c and PD-c rules are disjoint, and neither one is a strict subset of GSRs.

Theorem 3. *For $m \geq 3$ alternatives, no PM-c rule or PD-c rule is monotone-robust with respect to all distance metrics.*

In the proof of Theorem 3 we employ the following intuitive but somewhat technical statement, whose proof appears in the full version of the paper.

Lemma 2. *Given a specific ranking $\sigma^* \in \mathcal{L}(A)$ and a probability distribution D over the rankings of $\mathcal{L}(A)$ such that*

$$\arg\max_{\tau \in \mathcal{L}(A)} \Pr_D[\tau] = \{\sigma^*\},$$

there exists a distance metric d over $\mathcal{L}(A)$ and a d-monotonic noise model G with $\Pr_G[\sigma; \sigma^*] = \Pr_D[\sigma]$ for every $\sigma \in \mathcal{L}(A)$.

Proof of Theorem 3. Let $A = \{a_1, \ldots, a_m\}$ be the set of alternatives. We use a_{4-m} as shorthand for $a_4 \succ \ldots \succ a_m$. Fix

$$\tau = a_1 \succ \ldots \succ a_m,$$

and

$$\sigma^* = a_2 \succ a_1 \succ a_3 \succ a_{4-m}.$$

[2] We remark that considering the closures is necessary since $\Delta^{m!}$ contains only points with rational coordinates; hence it (as well as any subset of it) has measure zero.

[3] Available at: `www.cs.cmu.edu/~arielpro/papers`.

First, we prove that no PM-c rule is monotone-robust with respect to all distance metrics. In particular, using Lemma 2, we will construct a distance metric d and a d-monotonic noise model G such that no PM-c rule is accurate in the limit for G.

Consider the distribution D over $\mathcal{L}(A)$ defined as follows:

$$\Pr_D[a_2 \succ a_1 \succ a_3 \succ a_{4-m}] = \frac{4}{9},$$
$$\Pr_D[a_1 \succ a_2 \succ a_3 \succ a_{4-m}] = \frac{3}{9},$$
$$\Pr_D[a_1 \succ a_3 \succ a_2 \succ a_{4-m}] = \frac{2}{9},$$
$$\Pr_D[\sigma] = 0, \text{ for all } \sigma \text{ not covered above.}$$

By Lemma 2, there exist a distance metric d and a d-monotonic noise model G such that $\Pr_G[\sigma; \sigma^*] = \Pr_D[\sigma]$ for every $\sigma \in \mathcal{L}(A)$.

Given infinite samples from $G(\sigma^*)$, a $5/9$ fraction — a majority — of the votes have a_1 in the top position. A $7/9$ fraction of the votes prefer a_2 to a_3, while all votes prefer a_2 and a_3 to any other alternative besides a_1. Clearly, a_i is preferred to a_{i+1} for $i \geq 4$. Hence, in the PM graph, the alternatives are ordered according to $\tau = a_1 \succ a_2 \succ a_3 \succ a_{4-m}$. Therefore, every PM-c rule outputs τ in the limit, which is not the ground truth. Thus, no PM-c rule is accurate in the limit for G.

The construction for PD-c rules is more complex. Here, we will show that there is a noise model such that, given infinite samples for a specific ground truth, the PD graph of the profile induces a ranking that is different from the ground truth. The distribution D above is not sufficient for our purposes since there are pairs of alternatives (e.g., a_2 and a_3) that have the same probability of appearing in the first three positions of the outcome; hence, the PD graph of profiles with infinite samples may not be complete. Instead, we will use a distribution D' so that all probability values of this kind are different.

Let
$$0 = \delta_1 < \delta_2 < ... < \delta_m$$
so that $\sum_{i=1}^{m} \delta_i = 1$. Define the probability distribution D'' as follows. Pick one out of the m alternatives so that alternative a_i is picked with probability δ_i. Rank alternative a_i last and complete the ranking by a uniformly random permutation of the alternatives in $\mathcal{L}(A) \setminus \{a_i\}$. Now, the distribution D' is defined as follows: With probability $9/10$ (resp., $1/10$), the output ranking is sampled from the distribution D (resp., D'').

The important property of distribution D'' is that for every $k \in [m-1]$, the probability that alternative a_i is ranked in the first k positions is exactly $\frac{(1-\delta_i)k}{m-1}$, i.e., strictly decreasing in i. On the other hand, distribution D has the property that for every $k \in [m-1]$, the probability that alternative a_i is ranked in the first k positions is non-increasing in i. Hence, their linear combination D' has the property that for every $k \in [m-1]$, the probability that alternative a_i is ranked in the first k positions is strictly decreasing in i. Additionally,
$$\arg\max_{\tau \in \mathcal{L}(A)} \Pr_{D'}[\tau] = \{\sigma^*\}.$$

Hence, we can apply Lemma 2 to obtain a distance metric d' and a d'-monotonic noise model G' so that an infinite number of samples from $G'(\sigma^*)$ induce a complete PD graph corresponding to the ranking $\tau = a_1 \succ a_2 \succ a_3 \succ a_{4-m}$, which is different from the ground truth σ^*. Thus, no PD-c rule is accurate in the limit for G'.

We conclude that no PM-c rule or PD-c rule is monotone-robust with respect to all distance metrics. $\square$

The restriction on the number of alternatives in Theorem 3 is indeed necessary. For two alternatives, $\mathcal{L}(A)$ contains only two rankings, and all reasonable voting rules coincide with the majority rule that outputs the more frequent of the two rankings. It can be shown that, in this case, the majority rule is monotone-robust with respect to all distance metrics.

Caragiannis et al. (2013) show that the union of PM-c and PD-c rules includes all positional scoring rules, Bucklin's rule, the Kemeny rule, ranked pairs, Copeland's method, and Slater's rule. Two prominent SWFs that are neither PM-c nor PD-c are the maximin rule and STV. In the example given in the proof of Theorem 3, the maximin rule and STV would also rank the wrong alternative (a_1) in the first position with probability 1 in the limit. Thus, Theorem 3 gives another proof that prominent voting rules are not monotone-robust with respect to all distance metrics.

Discussion

Perhaps our main conceptual contribution is the realization that the modal ranking rule — a natural voting rule that was previously disregarded — can be exceptionally useful in crowdsourcing settings. Interestingly, from a classic social choice viewpoint the modal ranking rule would appear to be a poor choice. It does satisfy some axiomatic properties, such as Pareto efficiency — if all voters rank x above y, the output ranking places x above y (indeed, the rule always outputs one of the input rankings). But the modal ranking rule fails to satisfy many other basic desiderata, such as monotonicity — if a voter pushes an alternative upwards, and everything else stays the same, that alternative's position in the output should only improve. So our uniqueness result implies an impossibility: a voting rule that is monotone-robust with respect to any distance metric d and is a GSR with no holes, PD-c rule, or PM-c rule, cannot satisfy the monotonicity property. A similar statement is true for any social choice axiom not satisfied by the modal ranking rule. That said, social choice axioms like monotonicity were designed with subjective opinions, and notions of social justice, in mind. These axioms are incompatible with the settings that motivate our work on a conceptual level, and — as our results show — on a technical level.

Acknowledgments

Procaccia and Shah were partially supported by the NSF under grants NSF IIS-1350598 and NSF CCF-1215883. Caragiannis was partially supported by COST Action IC1205 on Computational Social Choice.

References

Caragiannis, I.; Procaccia, A. D.; and Shah, N. 2013. When do noisy votes reveal the truth? In *Proceedings of the 14th ACM Conference on Electronic Commerce (EC)*, 143–160.

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

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

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

Desario, D., and Robins, S. 2011. Generalized solid-angle theory for real polytopes. *The Quarterly Journal of Mathematics* 62(4):1003–1015.

Dodgson, C. L. 1876. *A Method for Taking Votes on More than Two Issues*. Clarendon Press.

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

Law, E., and von Ahn, L. 2011. *Human Computation*. Morgan & Claypool.

Little, G.; Chilton, L. B.; Goldman, M.; and Miller, R. C. 2010. Turkit: Human computation algorithms on Mechanical Turk. In *Proceedings of the 23rd Annual ACM symposium on User interface software and technology (UIST)*, 57–66.

Lu, T., and Boutilier, C. 2011. Learning Mallows models with pairwise preferences. In *Proceedings of the 28th International Conference on Machine Learning (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. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI)*, 1142–1148.

Mossel, E.; Procaccia, A. D.; and Rácz, M. Z. 2013. A smooth transition from powerlessness to absolute power. *Journal of Artificial Intelligence Research* 48:923–951.

Moulin, H. 1983. *The Strategy of Social Choice*, volume 18 of *Advanced Textbooks in Economics*. North-Holland.

Obraztsova, S.; Elkind, E.; Faliszewski, P.; and Slinko, A. 2013. On swap-distance geometry of voting rules. In *Proceedings of the 12th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 383–390.

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

Saari, D. G. 1995. *Basic geometry of voting*. Springer.

Saari, D. G. 2008. Complexity and the geometry of voting. *Mathematical and Computer Modelling* 48(9–10):1335–1356.

Xia, L., and Conitzer, V. 2008. Generalized scoring rules and the frequency of coalitional manipulability. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC)*, 109–118.

Xia, L., and Conitzer, V. 2009. Finite local consistency characterizes generalized scoring rules. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, 336–341.

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

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

Xia, L. 2013. Generalized scoring rules: a framework that reconciles Borda and Condorcet. *SIGecom Exchanges* 12(1):42–48.

Young, H. P. 1975. Social choice scoring functions. *SIAM Journal of Applied Mathematics* 28(4):824–838.

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