
Distributional Rank Aggregation, and an Axiomatic Analysis

Adarsh Prasad*
Harsh Pareek*
Pradeep Ravikumar

ADARSH@CS.UTEXAS.EDU
HARSHP@CS.UTEXAS.EDU
PRADEEPR@CS.UTEXAS.EDU

Department of Computer Science, The University of Texas, Austin, TX 78712, USA

Abstract

The rank aggregation problem has been studied with varying desiderata in varied communities such as Theoretical Computer Science, Statistics, Information Retrieval and Social Welfare Theory. We introduce a variant of this problem we call *distributional rank aggregation*, where the ranking data is only available via the induced distribution over the set of all permutations. We provide a novel translation of the usual social welfare theory axioms to this setting. As we show this allows for a more quantitative characterization of these axioms: which then are not only less prone to misinterpretation, but also allow simpler proofs for some key impossibility theorems. Most importantly, these quantitative characterizations lead to natural and novel *relaxations* of these axioms, which as we show, allow us to get around celebrated impossibility results in social choice theory. We are able to completely characterize the class of positional scoring rules with respect to our axioms and show that Borda Count is optimal in a certain sense.

1. Introduction

We study the problem of rank aggregation: a set of agents provide their ranked preferences over a set of alternatives, and we wish to aggregate them into a consensus ranking. Due in part to its generality and importance, this problem has been studied with varying desiderata in varied communities.

In Theoretical Computer Science, this has been framed as a combinatorial optimization problem of finding the ranking that minimizes the average distance (for some appropriate distance measure such as the Kemeny distance) to the set of

agent rankings. These resulting problems are however typically NP-hard, and research effort has been devoted to developing polynomial time algorithms with approximation guarantees (Ailon, 2010; Dwork et al., 2001a;b). These can also be cast as maximizing the average utility (in contrast to minimizing the average distance); Pivato (2013); Conitzer & Sandholm (2005) in particular showed that a large class of existing voting rules could also be interpreted as such average utility maximizers. Lu & Boutilier (2010) design a suitable utility measure that in their context captured the facet of uncertainty about the availability of alternatives.

In the field of Information Retrieval, the problem of rank aggregation arises as a “meta-algorithm” where for instance we are given the output rankings from multiple search engines, and we are to combine these into a single ranking (Dwork et al., 2001a;b). Also in this domain, Cheng et al. (2010) cast the aggregation task in a conditional or supervised setting, where we are given a training set of (feature-vector, ranking) pairs, and the task is to learn a map from the feature vector space to the set of permutations.

In Sociometrics, “social welfare theories” by Condorcet (1785); Arrow (1951) specify normative properties (axioms) that any reasonable rank aggregation should satisfy. Some results (including the celebrated impossibility theorem of Arrow (1951)) then state that some subsets of axioms cannot be jointly compatible. Several voting rules such as Condorcet methods, Borda Count, Runoff procedures, etc. have then been suggested, which satisfy different subsets of these axioms.

In Statistics, the agents’ preferences are viewed as i.i.d. samples from a statistical model and the problem is reformulated to that of learning a distribution over permutations and of computing the Maximum Likelihood Estimate (MLE) of the parameters; the final aggregation returned is then typically just the mode of this distribution. The focus in this line of work has then been on investigating efficiently learnable models for different types of in-

Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37. Copyright 2015 by the author(s).

*Denotes equal contribution by the authors

puts (Critchlow et al., 1991; Lebanon & Lafferty, 2002a;b; Hunter, 2004; Lebanon & Mao, 2008; Guiver & Snelson, 2009; Negahban et al., 2012; Soufiani et al., 2013). There have also been other proposals that go beyond this MLE based approach. In the social choice domain, where the aim is to pick a single winner, Young (1988) proposed to select a winning alternative that is “most likely to be the best (i.e., top-ranked in the true ranking)” and provided formulas to compute it for three alternatives. This idea has been formalized and extended by Procaccia et al. (2012) to choose a given number of alternatives with highest marginal probability under the Mallows model. In recent work, Soufiani et al. (2014) considered the estimation of the Mallows model ($P(\cdot|W) \propto \phi^{\ell_{KT}(\cdot, W)}$), and outputting a point estimate of the distribution based on statistical decision theoretic considerations. The interesting aspect of this work was that they then derived constraints on the dispersion parameter ϕ of the Mallows model for their estimator to satisfy different social-choice theoretic axioms.

We consider rank aggregation from the computer science and social-welfare theoretic viewpoint: and wish to aggregate an arbitrary set of ranking preferences over a set of alternatives into a consensus ranking over the entire set of alternatives. We note that we focus on the more difficult social *welfare* problem of providing a ranking over the entire set of alternatives, in contrast to the simpler social *choice* problem of outputting only the winner, i.e. the top position, in the aggregation. In many cases however, we do not have access to the specific agent or voter identities associated with a particular preference, as for instance with data from a secret ballot. Even given access to such voter identities, the key normative property (i.e. social choice theoretic axiom) of *anonymity* entails that the consensus ranking not make use of this information. Accordingly, one might consider a rank aggregation algorithm that has access to (or makes use of) only the histogram of the ranking preferences, or in other words, the empirical distribution over the set of all possible rankings as entailed by the given set of agent ranking preferences. We term this problem as *distributional rank aggregation*: given any distribution over the set of all possible rankings over a set of alternatives, output a consensus ranking, that in effect serves as a point-estimate of this distribution. (We note that this is distinct from the statistical viewpoint of assuming preferences are i.i.d. samples from some distribution, and estimating this distribution.) Such distributional rank aggregation is also natural when aggregating preferences of a very large population, where we would expect almost all possible rankings to be present in differing relative proportions, and it would be natural to collate these preferences into a distribution over the set of all possible rankings over the set of alternatives.

The first contribution of this paper is in translating the set

of normative axioms studied in the social choice literature to our distributional ranking setting. As we show, our more general distributional setting allows to *quantitatively* characterize these axioms: which addresses a common problem in social choice practice where the qualitatively stated social choice theoretic axioms are frequently misunderstood or misinterpreted. This also provides a potential way to understand the relationships between the varied social welfare axioms. We also provide a counterpart of a classical impossibility theorem stating that it is impossible to satisfy all of a set of key axioms simultaneously, but where our quantitative characterization of the axioms allows a very simple proof.

Another key advantage of a quantitative characterization of the axioms is in potentially *relaxing* them. In our second key contribution, we provide precisely such a novel quantitative characterization of approximately satisfiable or relaxed variants of the set of social choice axioms.

In our third key contribution, we leverage these relaxed axiom variants to *finesse* the celebrated impossibility theorem(s): we show that there exist distributional rank aggregation algorithms that satisfy the relaxed variants of all of a set of key axioms. We build towards this in the following stages: we first deconstruct the set of normative axioms and derive an equivalent characterization in terms of certain margin-like quantities. We then focus on a particular class of rank aggregation algorithms known as positional scoring rules, and exhaustively characterize which members of this class satisfy the set of exact axioms as well as their relaxed counterparts. As we then show, there exist members of this family that do satisfy the relaxed variants of all of a set of key axioms, thus allowing us to finesse a key impossibility theorem.

Our development leads us to a surprising result: that Borda Count has the minimum worst-case margin with respect to approximate satisfiability for certain axioms over all distributions. For specific distributions, our framework also allows us to identify positional losses which can outperform Borda and we present experiments to this effect.

2. Problem Setup

Consider a set of *alternatives* or a *label* set \mathcal{X} of size n ; without loss of generality, we will assume this set is $\mathcal{X} := [n] = \{1, \dots, n\}$. A total ordering over the label set is given by a permutation σ over the set $[n]$, so that $\sigma(x) = j$ indicates that the position of label x is j in σ . Let S_n be the set of all permutations. We also frequently refer to permutations as votes or rankings in this paper. We will refer to the alternatives or labels with alphabets $\{x, y\}$. The alternative x being preferred to y in a permutation σ is equivalent to having $\sigma(x) < \sigma(y)$.

Consider any distribution P over the set of permutations S_n . Let the marginal probability of x being ranked first be denoted by $P^1(x) = \sum_{\sigma:\sigma(x)=1} P(\sigma)$ and the marginal probability of x being ranked above y be denoted as $P_{x < y} = \sum_{\sigma:\sigma(x) < \sigma(y)} P(\sigma)$.

Definition 1. Let \mathcal{P}_n denote the set of all distributions over S_n . We then say that a *distributional rank aggregation function* is any map $\sigma^* : \mathcal{P}_n \mapsto S_n$, that given any distribution P over S_n outputs a “consensus” ranking $\sigma^*(P) \in S_n$.

We note that this differs from the standard rank aggregation algorithm framework in two subtle respects: (a) the aggregation function takes as input only histogram information of the set of agent preferences, so that it does not have access to voter identities, and (b) it does not assume a finite set of voters. Note that an infinite set of voter preferences can be naturally aggregated into a distribution over the set of all rankings and provided as input to a distributional rank aggregation algorithm. Indeed, typical social-choice theoretic settings implicitly assume a very small number of voters, which in a distributional setting would translate to a distribution with support limited to a very small subset of S_n . Another facet of the distributional setting is that all possible ranking preferences are present, in potentially differing proportions (as specified by the distribution over the set of all rankings), so that it also does not assume any notion of consistency among the voter preferences.

But the key advantage of the distributional rank aggregation framework is that it allows us to characterize social-choice theoretic axioms quantitatively in terms of the distribution P over S_n . In the following sections, we first review standard social-choice theoretic axioms which specify normative properties that any rank aggregation algorithm should satisfy, and extend these to our distributional rank aggregation setting.

3. Axiomatic Analysis

Social choice/welfare theory was put on an axiomatic foundation by Arrow (1951) who stated the following axioms that a voting (rank aggregation) procedure should satisfy:

- (S1) **Non-dictatorship:** The social welfare function should account for the wishes of multiple voters. It cannot simply mimic the preferences of a single voter.
- (S2) **Universality:** For any set of individual voter preferences, the social welfare function should yield a unique, deterministic and complete ranking of societal choices.
- (S3) **Transitivity:** The final aggregation should be strictly transitive i.e. if alternative x is preferred over y , y is

preferred over z then x should also be preferred over z .

- (S4) **Pareto-efficiency:** For every pair x and y of alternatives, If everyone prefers x to y , then x is preferred to y in the resulting social preference order.
- (S5) **Independence of irrelevant alternatives (IIA):** The social preference between x and y should depend only on the individual preferences between x and y . i.e. if one or more voters change their preferences, but no one changes their relative positions of x and y , then the relative positions of x and y in the aggregation should still remain the same.

Arrow (1951) proves Arrow’s Impossibility Theorem, i.e. when then the number of alternatives is strictly greater than 2, then no function exists which satisfies the above stated axioms and thus, these axioms are inconsistent. Brief proofs of Arrow’s theorem are presented in Geanakoplos (2005).

In addition to Arrow’s axioms, some other axioms are also well-established in the Social Welfare Theory literature (Soufiani et al., 2014):

- (S6) **Anonymity:** The social welfare procedure is insensitive to permutations over voters.
- (S7) **Monotonicity:** For any alternative, say x , if one of the voters moves the position of x up in his list, the position of x can only improve in the aggregation.
- (S8) **Consistency:** Let $S = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$ and $S' = \{\sigma'_1, \sigma'_2, \dots, \sigma'_n\}$ be two sets of voters/permutations such that the aggregation procedure assigns the same alternative to the top position in each, then when the input is $\hat{S} = \{S \cup S'\}$, the winner should remain the same.
- (S9) **Majority rule:** For any set of voters, if an alternative x is ranked in the top position in strictly more than half the votes, then in the final aggregation, x should be ranked at the highest position.
- (S10) **Condorcet Criterion:** If there exists an alternative say x , such that for every other alternative y , x is preferred over y , by strictly more than half of the voters, then x should be ranked at the highest position in the aggregation.

3.1. Translating Axioms to the Distributional Rank Aggregation Framework

We consider a distributional view of the rank aggregation problem where the individual rankings cannot be accessed

directly and are available only through a distribution, i.e. a histogram, over the permutations.

It is unclear what the social-choice theoretic axioms above imply in such a setting. Further, we would prefer more quantitative characterizations of these axioms, and perhaps even axioms whose preconditions involve quantitative conditions. In this section, we thus discuss a translation of each of the axioms above Section 3 to the distributional setting.

For any distribution P over S_n , we specify the axioms below for a general distributional rank aggregation map $\sigma^*(P)$, and for which we use the shorthand σ_P^* .

- (D1) **Non-dictatorship:** To be a dictatorship, a social welfare function must be able to identify the votes of individual voters (and subsequently follow the vote of the dictator). In the distributional framework, we only observe the aggregate distribution over S_n and hence cannot determine the vote of any individual voter. Thus, the non-dictatorship condition is always satisfied by the distributional framework.
- (D2) **Universality:** Since $\sigma_P^* \in S_n$, the output of our ranking procedure is always a complete ranking. Universality then requires that the output of the aggregation procedure should be unique.
- (D3) **Transitivity:** Since $\sigma_P^* \in S_n$, transitivity is automatically satisfied.
- (D4) **Pareto-efficiency:** The Pareto-efficiency condition can be translated as, for all permutations in the support of P , if alternative x is ranked above y , then x should be ranked above y in the aggregation i.e. if $\forall \sigma : P(\sigma) > 0; \sigma(x) < \sigma(y)$, then Pareto-efficiency is satisfied iff $\sigma_P^*(x) < \sigma_P^*(y)$.
- (D5) **Independence of irrelevant alternatives (IIA):** If voters change their preferences in a way that preserves the relative positions of x and y in each vote, the marginal probability of x being preferred over y remains constant. Let $P_{x < y} = \sum_{\sigma: \sigma(x) < \sigma(y)} P(\sigma)$ be the marginal probability of x being preferred over y . Then, the IIA condition can be translated as: $\forall x, y \in \mathcal{X}$, given any two distributions P and Q such that $P_{x < y} = Q_{x < y}$, then IIA is satisfied iff $\text{sign}[\sigma_P^*(x) - \sigma_P^*(y)] = \text{sign}[\sigma_Q^*(x) - \sigma_Q^*(y)]$.
- (D6) **Anonymity:** As in the non-dictatorship condition, we only have access to the votes through the distribution and thus cannot identify individual voters and anonymity is implicitly satisfied in the distributional framework.
- (D7) **Monotonicity:** If one or more voters improves the position of an alternative x in their respective rank-

ings, this shifts probability mass between certain permutations where the position of x goes up while that of other alternatives goes down or remains the same. Let σ_1 and $\sigma_2 \in S_n$ such that $\sigma_2(x) < \sigma_1(x)$ and $\sigma_2(y) \geq \sigma_1(y) \forall y \in \mathcal{X} \setminus \{x\}$. Then, we can translate the monotonicity condition as: Given two distributions P and Q s.t. $P(\sigma_2) = Q(\sigma_2) + \delta, P(\sigma_1) = Q(\sigma_1) - \delta, \delta > 0$ and $P(\sigma) = Q(\sigma)$ for $\sigma \neq \sigma_1, \sigma_2$ then, monotonicity is satisfied iff $\sigma_P^*(x) \leq \sigma_Q^*(x)$.

- (D8) **Consistency:** Consistency requires that if an alternative wins in two separate sets of votes, it should also win in the (multiset) union of those sets. For distributions the union of the sets is given as $(P + Q)/2$ and we have the condition: Given two distributions P and Q such that $\sigma_P^*(x) = 1, \sigma_Q^*(x) = 1$, then consistency is satisfied iff $\sigma_{\frac{P+Q}{2}}^*(x) = 1$.
- (D9) **Majority rule:** If an alternative x ranks at top of strictly greater than half of the votes, then x must win. Let $P^1(x) = \sum_{\sigma: \sigma(x)=1} P(\sigma)$ measure the marginal probability of x being ranked first. Then, majority rule is satisfied iff $P^1(x) > \frac{1}{2} \Rightarrow \sigma_P^*(x) = 1$.
- (D10) **Condorcet Criterion:** Given a distribution P such that if $P_{x < y} > \frac{1}{2} \forall y \in \mathcal{X} \setminus \{x\}$ then the Condorcet criterion is satisfied iff $\sigma_P^*(x) = 1$. Observe that, if the Condorcet criterion is satisfied, then Majority rule is also satisfied.

This translation is a novel contribution of this paper. The following theorem, which follows from the above discussion, summarizes the equivalence of these axioms to those from social welfare theory:

Theorem 1 (Equivalence of axioms). *If voting data is accessible only via its distribution over S_n , the Social Welfare Axioms (S1-S10) are equivalent to the axioms (D1-D10) introduced above.*

Since non-dictatorship and transitivity are implicit in the distributional framework, Arrow's Impossibility Theorem states that a procedure cannot satisfy Universality, Pareto and IIA simultaneously. However, the quantitative nature of these axioms suggests that they can be relaxed and we will do so in the next section.

3.2. Relaxations of Social Welfare Axioms

We see that in the distributional setting, most axioms can be interpreted in terms of marginals of the distribution P , and represent constraints on σ_P^* based on these marginals. This probabilistic interpretation of the axioms motivates us to account for noise in the observed rankings by introducing a margin-like notion and go beyond impossibility results.

Consider for example, IIA. Let $M_\gamma(x, y) = \{P | P_{x < y} = \gamma\}$ be the family of distributions with marginal probability γ for $\sigma(x) < \sigma(y)$, then IIA is satisfied, if for all $0 \leq \gamma \leq 1$, and distributions $Q_1, Q_2 \in M_\gamma(x, y)$,

$$\text{sign} [\sigma_{Q_1}^*(x) - \sigma_{Q_1}^*(y)] = \text{sign} [\sigma_{Q_2}^*(x) - \sigma_{Q_2}^*(y)] \quad (1)$$

i.e. moving probability mass such that the marginals remain the same does not alter the relationship between x and y in the aggregation.

Intuitively, the hardest case for IIA is when $\gamma = \frac{1}{2}$, as it is unclear which of x and y should be preferred based on the marginal probability. However, as γ moves away from $\frac{1}{2}$, a decision between x and y can be made more easily. This intuition motivates the definition below, where we require IIA to hold only for such distributions where the distinction between x and y can be made more easily:

Definition 2. ϵ -IIA is satisfied, iff for all $Q_1, Q_2 \in \{P | P_{x < y} = \gamma\}$, where $|\gamma - \frac{1}{2}| \geq \epsilon, \epsilon > 0$, then $\text{sign} [\sigma_{Q_1}^*(x) - \sigma_{Q_1}^*(y)] = \text{sign} [\sigma_{Q_2}^*(x) - \sigma_{Q_2}^*(y)]$.

Note that $\epsilon = 0$ corresponds to the usual IIA.

Similarly, the Majority Rule axiom can be parameterized in terms of marginal probability of an alternative x being ranked at the top, $P^1(x) = \sum_{\sigma: \sigma(x)=1} P(\sigma) = \gamma$. As this marginal probability goes to 1, a reasonable rank aggregation procedure would place x at the top. This intuition motivates the following definition:

Definition 3. The ϵ -Majority Rule axiom is satisfied, iff for all $P : P^1(x) \geq \frac{1}{2} + \epsilon ; \epsilon > 0, \sigma_P^*(x) = 1$.

The Condorcet Criterion also allows for a similar extension:

Definition 4. ϵ -Condorcet Rule is satisfied, if for all $P : P_{x < y} \geq \frac{1}{2} + \epsilon ; \epsilon > 0 \forall y \in X \setminus \{x\}, \sigma_P^*(x) = 1$.

While it is also possible to relax the other axioms in a similar manner, it is not as natural; for instance, we expect Pareto to always be satisfied for any reasonable rank aggregation procedure. In addition, as we will see in Section 4, the above relaxations are sufficient to ensure that a ranking procedure satisfying all our axioms exist.

Note that the Condorcet criterion is stronger than Pareto in that Condorcet implies Pareto. It is also possible to strengthen the Condorcet criterion further by dropping the $\forall y$ requirement. We call this the ϵ -Strong Condorcet criterion:

Definition 5. ϵ -Strong Condorcet Rule is satisfied, if for all $P : P_{x < y} \geq \frac{1}{2} + \epsilon ; \epsilon > 0$, then $\sigma_P^*(x) < \sigma_P^*(y)$.

Thus, if x is ranked above y in a significant fraction of the votes, x should be ranked above y in the aggregation. We

note that this may not always be desirable, due to the following proposition:

Proposition 1. For $\epsilon < 1/6$, ϵ -Strong Condorcet is incompatible with universality.

We also have:

Proposition 2. ϵ -Strong Condorcet $\Rightarrow \epsilon$ -IIA.

3.3. Comparative Analysis of Axioms

While the previous sections provided a quantitative characterization of social-choice theoretic axioms, as well as suitable relaxations of these, one might ask if there were commonalities between these different axioms. In this section, we investigate this question further. Before doing so, we first obtain a optimization based characterization of any distributional rank aggregation procedure.

Consider the following class of distributional rank aggregation procedures:

$$\sigma_P^* = \underset{\sigma \in S_n}{\text{argmin}} g(\sigma, P) \quad (2)$$

where $g : S_n \times \mathcal{P}_n \mapsto \mathbb{R}$, and \mathcal{P}_n is the set of all distributions on S_n . This characterization is WLOG as:

Proposition 3. Any distributional rank aggregation procedure can be expressed in the form (2) for some function g .

We now define two quantities for any pair of alternatives x and y based on g and $0 \leq \gamma \leq 1$.

$$L_\gamma(g; x, y) = \min_{P: P_{x < y} = \gamma} \left[\min_{\sigma: \sigma(x) < \sigma(y)} g_P(\sigma) - \min_{\sigma: \sigma(x) > \sigma(y)} g_P(\sigma) \right]$$

$$U_\gamma(g; x, y) = \max_{P: P_{x < y} = \gamma} \left[\min_{\sigma: \sigma(x) < \sigma(y)} g_P(\sigma) - \min_{\sigma: \sigma(x) > \sigma(y)} g_P(\sigma) \right]$$

Thus, $L_\gamma(g; x, y)$ and $U_\gamma(g; x, y)$ are the lower and upper bounds for the quantity in brackets. We now show how most of the axioms can be re-written in terms of these two quantities.

Proposition 4. g satisfies Pareto-efficiency iff $U_\gamma(g; x, y) < 0, \forall x, y \in \mathcal{X}$ for $\gamma = 1$

Proposition 5. g satisfies IIA iff $L_\gamma(g; x, y), U_\gamma(g; x, y) > 0, \forall \gamma : 0 \leq \gamma \leq 1$ and $\forall x, y \in \mathcal{X}$,

Proposition 6. g satisfies the Condorcet criterion iff $U_\gamma(g; x, y) < 0$, for $\frac{1}{2} < \gamma \leq 1, \forall x, y \in \mathcal{X}$.

We define another quantity for any alternative x based on g and $0 \leq \gamma \leq 1$

$$U_\gamma^1(g; x) = \max_{P: P^1(x) = \gamma} \left[\min_{\sigma: \sigma(x)=1} g_P(\sigma) - \min_{\sigma: \sigma(x) \neq 1} g_P(\sigma) \right]$$

Proposition 7. g satisfies Majority Rule iff $U_\gamma^1(g; x) < 0$, for all $\frac{1}{2} < \gamma \leq 1, \forall x \in \mathcal{X}$.

Similar conditions can be obtained for the relaxed definitions of axioms discussed before. Consistency and monotonicity both depend on the structure of g in ways that we believe cannot be captured as easily by the quantities $L_\gamma(g; x, y)$ and $U_\gamma(g; x, y)$ and we leave these for future work.

3.4. An Impossibility Theorem for Distributional Rank Aggregation

We now present an impossibility theorem regarding the ability of any distributional rank aggregation procedure to satisfy both Universality and Pareto axioms simultaneously. Recall that Arrow's impossibility theorem in our setting states that no procedure satisfies Pareto, IIA and Universality. This result then provides an alternative intuitive proof of Arrow's theorem for the special case of g continuous in P .

Theorem 2. *For $n \geq 2$, if $g(\sigma, P)$ in (2) is a continuous function of P for each fixed σ , both Universality and Pareto cannot be satisfied simultaneously.*

The continuity property is desirable for computational purposes and procedures used in practice for rank aggregation use a continuous g . Since Pareto is an important condition that we do not wish to relax, the above result implies that we must abandon the universality assumption, i.e. we must be content with procedures that sometimes return multiple aggregations.

Regarding Arrow's impossibility theorem for general g , we first note that the usual proofs of Arrow's theorem which rely on identifying pivotal voters fail completely in this framework. In fact, Fishburn (1970) gives a *possibility* result in the case of infinite voters, which our framework captures. However, the proof there is not constructive and the g involved will quite likely be pathological — it will certainly not be continuous in P . We hope to investigate this further in future work.

4. Positional Scoring Losses

The impossibility theorems derived in the previous section state that there are no distributional rank aggregation procedures that simultaneously satisfy a set of key axioms. But would the impossibility theorem continue to hold if we only need satisfy our relaxed variants of the exact axioms? We can ask a more general and more difficult question: for each axiom, and its relaxed variant, can we exhaustively characterize which rank aggregation procedures would satisfy or not satisfy these?

We consider a simplification of the above general question, and ask for such an exhaustive characterization for a simpler family of rank aggregation procedures known as *posi-*

tional scoring rules (Kenneth J. Arrow & Suzumura, 2002, Chapter 7).

Recall from Section 3.3 that any distributional rank aggregation function can be expressed in the form in (2) as the minimizer of some loss functional $g(\sigma, P)$. Let us first consider the following specialization of these loss functionals that are commonly used in varied computer scientific domains. Specifically, for any bivariate loss function $\ell : S_n \times S_n \mapsto \mathbb{R}$ that measures the discrepancy between two rankings, consider the loss functional $g(\sigma, P)$ set to the expected value of ℓ over P . We then obtain the following rank aggregation procedure:

$$\sigma_{\ell, P}^* = \operatorname{argmin}_{\sigma \in S_n} g(\sigma, P) = \operatorname{argmin}_{\sigma \in S_n} \mathbb{E}_{\sigma' \sim P}[\ell(\sigma, \sigma')] \quad (3)$$

In general, $\ell(\sigma', \sigma)$ can be any bivariate function but the following are of special interest:

- $\ell_{0-1}(\sigma, \sigma') = \mathbb{1}[\sigma \neq \sigma']$: This corresponds to the zero-one distance between permutations. The corresponding $\sigma_{\ell_{0-1}, P}^*$ is then the mode of the distribution P . Statistical and machine learning approaches to the rank aggregation problem fit a (unimodal) distribution to the P and then return the mode of the estimated distribution as a point estimate. Thus, these techniques implicitly use the zero-one loss.
- $\ell_{KT}(\sigma, \sigma') = \sum_{(x, y) \in [n]} \mathbb{1}[\operatorname{sign}(\sigma(x) - \sigma(y)) \neq \operatorname{sign}(\sigma'(x) - \sigma'(y))]$: This loss measures the Kendall-Tau distance between two permutations, i.e. the number of inversions between them. The minimization problem for this loss has been shown to be NP-hard and this loss has been studied widely in the Theoretical Computer Science literature.
- $\ell(\sigma, \sigma') = \sum_{x \in [n]} |\sigma(x) - \sigma'(x)|$: This loss measures the Spearman-footrule distance and is used as an approximation to the Kendall-Tau.

Positional scoring methods assign a score to each alternative in a permutation based on its position, these scores are aggregated across permutations and the final aggregation is obtained by sorting the alternatives based on their cumulative scores. Equivalently,

Definition 6 (Positional Scoring Loss). *A loss ℓ_h is a Positional Scoring Loss iff it can be decomposed as $\ell_h(\sigma, \sigma') = \sum_x h(\sigma'(x)) \cdot \sigma(x)$, where $h : [n] \mapsto \mathbb{R}$*

Using Equation (3), the corresponding σ_{P, ℓ_h}^* can be written as:

$$\begin{aligned} \sigma_{P, \ell_h}^* &= \operatorname{argmin}_{\sigma \in S_n} \mathbb{E}_{\sigma' \sim P}[\ell_h(\sigma, \sigma')] \\ &= \operatorname{argmin}_{\sigma \in S_n} \sum_x \mathbb{E}_{\sigma' \sim P}[h(\sigma'(x))] \cdot \sigma(x) \end{aligned}$$

| Name | $h : [n] \mapsto \mathbb{R}$ | Ω_h | ω_h | ω_h^1 | ϵ for IIA/Cond. | ϵ for Maj. |
|---------------------|--|------------|-----------------------|--------------|-------------------------------------|-------------------------------------|
| Borda Count | $h(i) = n - i$ | $n - 1$ | 1 | 1 | $1/2 - 1/n$ | $1/2 - 1/n$ |
| Plurality Rule | $h(i) = 1$ if $i = 1$, 0 else | 1 | 0 | 1 | $1/2$ | 0 |
| Anti-Plurality Rule | $h(n) = 0$; $h(i) = 1 \forall i \neq n$ | 1 | 0 | 0 | $1/2$ | $1/2$ |
| Log Rule | $h(i) = -\log(i)$ | $\log(n)$ | $\log(\frac{n}{n-1})$ | $\log(2)$ | $\frac{\log(n)}{2 \log(n^2/(n-1))}$ | $\frac{\log(n)}{2 \log(n^2/(n-1))}$ |
| Squared Rule | $h(i) = -i^2$ | $n^2 - 1$ | 3 | 3 | $\frac{n^2 - 4}{2(n^2 + 2)}$ | $\frac{n^2 - 4}{2(n^2 + 2)}$ |

Table 1. Different Positional Scoring Rules and their key properties

$$= \operatorname{argmin}_{\sigma \in S_n} \sum_x f_P(x) \cdot \sigma(x) \quad (4)$$

where $f_P : \mathcal{X} \mapsto \mathbb{R}$ is given by $f_P(x) = \mathbb{E}_{\sigma' \sim P}[h(\sigma'(x))]$ is the cumulative scoring function.

Proposition 8. Given $f_P(x) = \mathbb{E}_{\sigma' \sim P}[h(\sigma'(x))]$, then σ_{P, ℓ_h}^* is the permutation achieved by sorting the alternatives in descending order of $f_P(x)$

This proposition also shows that the rank aggregation using Positional Scoring rules can be computed efficiently. Table 1 give examples of some well-known positional scoring rules, their corresponding h -functions, and their key-properties.

4.1. Axiomatic Analysis of Positional Scoring Rules

Non-dictatorship, Transitivity and Anonymity are implicit in the distributional framework. Positional scoring rules are continuous in P , so Theorem 2 applies and we will not require Universality.

We define three properties of interest for any positional loss function. These quantities play a key role in establishing several axioms:

- $\Omega_h = \max_{\{i, j \in [n] | i < j\}} h(i) - h(j)$.

This is the maximum variation in possible scores for h .

- $\omega_h = \min_{\{i, j \in [n] | i < j\}} h(i) - h(j)$

This is the minimum variation in possible scores for h .

- $\omega_h^1 = \min_{j \in [n] | j > 1} h(1) - h(j)$

This is the minimum variation in possible scores for h with respect to the first position. This is used in analyzing the Majority Rule.

We observe the following:

- For $n = 2$, $\Omega_h = \omega_h = \omega_h^1$.
- For $n > 3$, $\Omega_h \geq \omega_h$ with equality holding iff $h : [n] \mapsto \mathbb{R}$ is a constant function.

Proposition 9 (IIA). No positional scoring loss ℓ_h satisfies IIA exactly.

Proposition 10 (ϵ -IIA). A positional scoring loss ℓ_h satisfies ϵ -IIA iff $\epsilon > \frac{1}{2} \left(\frac{\Omega_h - \omega_h}{\Omega_h + \omega_h} \right)$.

Proposition 11 (Pareto Efficiency). A positional scoring loss ℓ_h satisfies Pareto-efficiency exactly iff $h : [n] \mapsto \mathbb{R}$ is strictly monotonically decreasing.

Proposition 12 (Monotonicity). A positional scoring loss ℓ_h satisfies Monotonicity iff $h : [n] \mapsto \mathbb{R}$ is non-increasing.

Proposition 13 (Consistency). Every positional loss function ℓ_h satisfies consistency.

Proposition 14 (Majority rule). A positional scoring loss ℓ_h satisfies Majority rule iff $\max(\Omega_h, -\omega_h) - \omega_h^1 = 0$ and $\omega_h^1 + \max(\Omega_h, -\omega_h) > 0$

Proposition 15 (ϵ -Majority rule). A positional scoring loss ℓ_h satisfies ϵ -Majority Rule iff $\epsilon > \frac{1}{2} \left(\frac{\max(\Omega_h, -\omega_h) - \omega_h^1}{\max(\Omega_h, -\omega_h) + \omega_h^1} \right)$.

Proposition 16 (Condorcet Criterion). No positional scoring loss ℓ_h satisfies the Condorcet Criterion

Proposition 17 (ϵ -Strong Condorcet Criterion). A positional scoring loss ℓ_h satisfies the ϵ -Strong Condorcet Criterion iff $\epsilon > \frac{1}{2} \left(\frac{\Omega_h - \omega_h}{\Omega_h + \omega_h} \right)$

Recall from Proposition 2, that ϵ -Strong Condorcet implies ϵ -IIA, consistent with propositions 10 and 17. Thus, we have the same ϵ for both IIA and Condorcet for positional loss functions and they are shown together in Table 1.

4.2. A Possibility Theorem with Relaxed Axioms

Given the theory above, we can now quantify the extent to which a given positional loss satisfies each of the approximate axioms. A natural first question is whether any loss satisfies our relaxed axioms.

Theorem 3. The set of positional loss functions which admits the following axioms is non-empty: exact versions of Pareto and Monotonicity, relaxed axioms ϵ -IIA, ϵ -Strong Condorcet, ϵ' -Majority Rule

A second question is whether we can design losses that satisfy the approximate axioms to the greatest extent possible. The following theorems show a surprising result, that the Borda Count is the optimal positional ranking function in a certain sense.

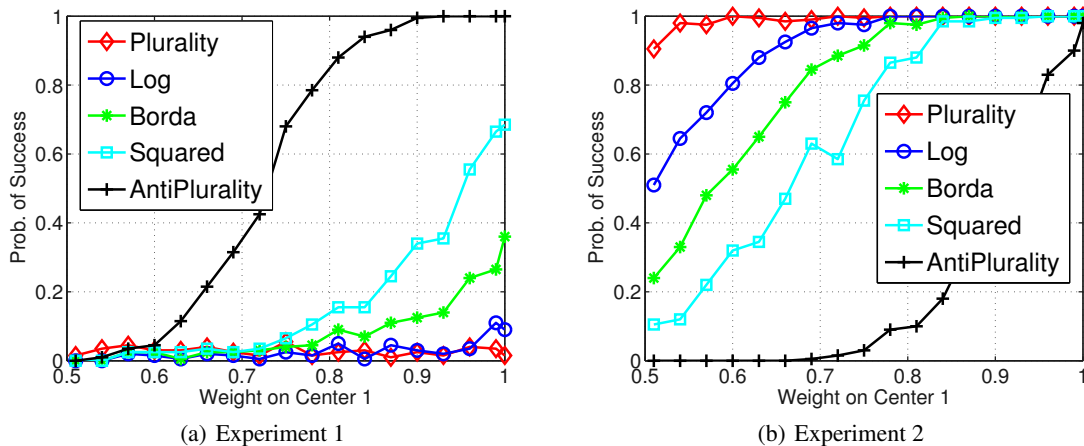


Figure 1. Performance of different positional scoring rules on the mixture of Mallows' models

Lemma 1. For any positional scoring loss ℓ_h that satisfies Pareto, $\frac{\Omega_h}{\omega_h} \geq n - 1$, where n is the number of alternatives.

Theorem 4. For any fixed n , Borda count is optimal w.r.t. the ϵ -Strong Condorcet condition and ϵ -IIA, i.e. has the least ϵ among all positional loss functions.

This fact seems to have been noted previously in the literature. Saari (1990) states “the Borda method is the unique positionalist method to minimize the kinds and number of paradoxes that can occur”. However, the justification given in (Saari, 1990) uses the much more complex machinery that our relaxation-based approach.

5. Experiments

Theorem 4 showed that Borda Count is optimal in a “worst-case” sense in satisfying certain relaxed axioms. However, this does not mean that it dominates all other positional losses over all distributions and we present experiments to demonstrate this fact. Our framework shows that the key quantities to take into account when designing a positional loss are Ω_h and ω_h . For a non-increasing h , h can be scaled so that Ω_h is fixed. Then ω_h , intuitively the smallest (discrete) gradient of h , is the key quantity to compare among losses. A key factor in the following discussion will be the position in the list where this minimum value ω_h is achieved. Our discussion will be terse for lack of space and we discuss the results in detail in Appendix B.

Setup. We consider the following losses: Borda count where h is linear, a “Convex” loss log, a “concave” loss Square and the limiting convex and concave losses Plurality and AntiPlurality. The convex losses achieve minimum $\omega_h = h(n) - h(n - 1)$ at the bottom of the list and thus pay more attention to relative positions at the top. The concave losses achieve minimum $\omega_h = h(1) - h(2)$ at the top of the

list and thus pay more attention to relative positions at the bottom. See Table 1 for details on the losses.

Let $\{A, B, C, D, E\}$ be the set of alternatives \mathcal{X} , $n = 5$. A Mallows model is given by $P(\cdot|W) \propto \phi^{\ell_{KT}(\cdot, W)}$, where W is the central permutation, $\phi \in (0, 1]$ is the dispersion parameter and ℓ_{KT} is the Kendall-Tau distance. A higher ϕ makes the distribution more concentrated around the central permutation. We use a mixture of two Mallows models for our experiments. Let Z_1 and Z_2 be the two central permutations and w and $1 - w$ be their weights in the mixture. We perform two experiments. In *Experiment 1*, we fix centers $Z_1 = \{D, E, A, B, C\}$ and $Z_2 = \{B, C, D, E, A\}$, while in *Experiment 2* we fix centers $Z_1 = \{A, B, C, D, E\}$ and $Z_2 = \{B, C, D, E, A\}$. Then, for $\phi = 0.8$, we vary w from 0.51 to 1.0. Thus, in both experiments, we put higher weight on Z_1 having $\sigma(A) < \sigma(B)$ and as we increase the weight, we are effectively increasing $P_{x < y}$. For a given w , we generate samples from the mixture of Mallows model and aggregate them using the positional scoring rules.

Metric. We call an aggregation a “success” if it ranks A above B . We perform multiple trials for each w to calculate the probability of success.

Results. Figure 1 shows the probability of success against the weight value w for both experiments. As expected, there is an inverse relationship between the two plots w.r.t. the order of the best performing methods.

Acknowledgments

We acknowledge the support of ARO via W911NF-12-1-0390 and NSF via IIS-1149803, IIS-1320894, IIS-1447574, and DMS-1264033.

References

- Ailon, Nir. Aggregation of Partial Rankings, p-Ratings and Top-m Lists. *Algorithmica*, 57(2):284–300, 2010.
- Arrow, Kenneth J. 1963. Social choice and individual values. *Wiley, New York.*, 1951.
- Cheng, W., Hüllermeier, E., and Dembczynski, K. J. Label ranking methods based on the Plackett-Luce model. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 215–222, 2010.
- Condorcet, M de. Essay on the Application of Analysis to the Probability of Majority Decisions. *Paris: Imprimerie Royale*, 1785.
- Conitzer, Vincent and Sandholm, Tuomas. Common Voting Rules as Maximum Likelihood Estimators. In *UAI*, 2005.
- Critchlow, Douglas E, Fligner, Michael A, and Verducci, Joseph S. Probability models on rankings. *Journal of mathematical psychology*, 35(3):294–318, 1991.
- Dwork, C., Kumar, R., Naor, M., and Sivakumar, D. Rank aggregation methods for the web. In *10th international conference on World Wide Web*, pp. 613–622, 2001a.
- Dwork, C., Kumar, R., Naor, M., and Sivakumar, D. Rank aggregation revisited, 2001b.
- Fishburn, Peter C. Arrow’s impossibility theorem: Concise proof and infinite voters. *Journal of Economic Theory*, 2(1):103–106, 1970.
- Geanakoplos, John. Three brief proofs of Arrow’s impossibility theorem. *Economic Theory*, 2005.
- Guiver, J. and Snelson, E. Bayesian inference for Plackett-Luce ranking models. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 377–384, 2009.
- Hunter, D. R. MM algorithms for generalized Bradley-Terry models. *Annals of Statistics*, pp. 384–406, 2004.
- Kenneth J. Arrow, Amartya K. Sen and Suzumura, Kotaro. *Handbook of Social Choice and Welfare*, volume 1. Elsevier, 2002.
- Lebanon, G. and Lafferty, J. D. Cranking: Combining Rankings Using Conditional Probability Models on Permutations. In *ICML*, pp. 363–370, 2002a.
- Lebanon, G. and Lafferty, J. D. Conditional Models on the Ranking Poset. In *NIPS*, pp. 415–422, 2002b.
- Lebanon, G. and Mao, Y. Non-Parametric Modeling of Partially Ranked Data. *Journal of Machine Learning Research*, 9:2401–2429, 2008.
- Lu, Tyler and Boutilier, Craig E. The unavailable candidate model: a decision-theoretic view of social choice. In *ACM conference on Electronic Commerce*, 2010.
- Negahban, S., Oh, S., and Shah, D. Iterative Ranking from Pair-wise Comparisons. *Arxiv*, abs/1209.1688, 2012. URL <http://arxiv.org/abs/1209.1688>.
- Pivato, Marcus. Voting rules as statistical estimators. *Social Choice and Welfare*, 40(2), 2013.
- Procaccia, Ariel D., Reddi, Sashank Jakkam, and Shah, Nisarg. A Maximum Likelihood Approach For Selecting Sets of Alternatives. In *UAI*, 2012. URL <http://dblp.uni-trier.de/db/conf/uai/uai2012.html#ProcacciaRS12>.
- Saari, Donald G. The borda dictionary. *Social Choice and Welfare*, 7(4):279–317, 1990.
- Soufiani, Hossein Azari, Chen, William, Parkes, David C, and Xia, Lirong. Generalized method-of-moments for rank aggregation. In *Advances in Neural Information Processing Systems*, pp. 2706–2714, 2013.
- Soufiani, Hossein Azari, Parkes, David C, and Xia, Lirong. A Statistical Decision-Theoretic Framework for Social Choice. In *Advances in Neural Information Processing Systems*, pp. 3185–3193, 2014.
- Young, H Peyton. Condorcet’s theory of voting. *American Political Science Review*, 1988.