Low-Distortion Social Welfare Functions

Gerdus Benadè
Tepper School of Business
Carnegie Mellon University

Ariel D. Procaccia
Computer Science Department
Carnegie Mellon University

Mingda Qiao
Department of Computer Science
Stanford University

Abstract

Work on *implicit utilitarian voting* advocates the design of preference aggregation methods that maximize utilitarian social welfare with respect to latent utility functions, based only on observed rankings of the alternatives. This approach has been successfully deployed in order to help people choose a single alternative or a subset of alternatives, but it has previously been unclear how to apply the same approach to the design of *social welfare functions*, where the desired output is a ranking. We propose to address this problem by assuming that voters' utilities for rankings are induced by unknown weights *and* unknown utility functions, which, moreover, have a combinatorial (subadditive) structure. Despite the extreme lack of information about voters' preferences, we show that it is possible to choose rankings such that the worst-case gap between their social welfare and that of the optimal ranking, called *distortion*, is no larger (up to polylogarithmic factors) than the distortion associated with much simpler problems. Through experiments, we identify practical methods that achieve near-optimal social welfare on average.

1 Introduction

Classical social choice theory typically approaches the problem of preference aggregation from an axiomatic viewpoint, that is, researchers formulate attractive properties, and ask whether there are methods that satisfy them. By contrast, research in *computational social choice* (Brandt et al. 2016) is often guided by optimization, in that researchers specify quantitative measures of the desirability of different alternatives, and construct methods that optimize them.

One such approach is known as *implicit utilitarian voting* (Procaccia and Rosenschein 2006; Boutilier et al. 2015; Caragiannis et al. 2017). In a nutshell, the idea is that each voter i has a utility function u_i that assigns a value to each alternative. However, these utility functions are *implicit*, in the sense they cannot be communicated by the voters (because they are unknown or difficult to pin down). Instead, voters report rankings of the alternatives that are *consistent* with the underlying utility functions, that is, each voter sorts the alternatives in non-increasing order of utility. The goal is to choose an alternative a that maximizes *(utilitarian) social welfare* — the sum of utilities $\sum_i u_i(a)$ — using the reported rankings as a proxy for the latent utility functions.

From that viewpoint, the best method is the one that minimizes a measure called *distortion*, defined by Procaccia and Rosenschein (2006) as the ratio between the social welfare of the best alternative, and the social welfare of the alternative selected by the method, in the worst case over all utility functions that are consistent with the observed rankings. Put another way, this is the approximation ratio to the welfare-maximizing solution, and the need for approximation stems from lack of information about the true utilities.

In recent years, implicit utilitarian voting has emerged as a practical approach for helping groups of people make joint decisions. In particular, optimal *social choice functions* and *social choice correspondences*, based on the implementation of Caragiannis et al. (2017), are deployed on the not-for-profit website `RoboVote.org` for the case where the desired output is a single alternative or a subset of winning alternatives, respectively.

However, RoboVote also has a third output type, a ranking of the alternatives, and for this case the website currently does *not* take the same approach — instead it uses the well-known Kemeny rule (Davenport and Kalagnanam 2004; Conitzer, Davenport, and Kalagnanam 2006). Indeed, it is unclear how to even view the problem of designing a *social welfare function*, which returns a ranking, through the lens of implicit utilitarian voting — if a voter has a utility for each alternative, what is his utility for a ranking of the alternatives? One could assume that a voter i has a *weight* $w_{i,j}$ for each *position* j, so his utility for the ranking $(a_1, a_2, \ldots, a_m)$ would be $\sum_{j=1}^{m} w_{i,j} u_i(a_j)$; but any particular choice of weights would be *ad hoc*.

1.1 Our Approach and Results

The insight underlying our approach is that the worst-case perspective also extends to the choice of weights. That is, when we measure the social welfare of an output ranking given reported input rankings, we consider the worst case over *both* utility functions and weights.

Of course, this is a very conservative approach, and one might worry that it would lead to massive distortion. But our main theoretical result is that the distortion of optimal social welfare functions is asymptotically identical to that of optimal social choice functions (where there are no weights whatsoever), up to a polylogarithmic factor — in both cases it is $\tilde{\Theta}(\sqrt{m})$, where m is the number of alternatives.

In fact, we establish a significantly stronger result, as we allow voters to have *combinatorial* utility functions over subsets of alternatives, and we measure the utility of a voter for a ranking as the weighted sum of his utilities for prefixes of that ranking; the foregoing distortion bound holds when the utility functions are monotonic and subadditive. We find it striking that it is possible to formulate the problem in such generality with no tangible increase in distortion.

Our computational results demonstrate that it is practical to compute deterministic distortion-minimizing rankings for instances with up to 10 alternatives. This constraint on the instance size is not unreasonable, as 98.3% of RoboVote instances have 10 or fewer alternatives. For larger instances we test several heuristics and find that the Borda and Kemeny rules typically lead to low distortion and near-optimal social welfare.

1.2 Related Work

Generally speaking, the implicit utilitarian voting literature can be partitioned into two complementary strands of research. One does not constrain the structure of voters' utility functions (Procaccia and Rosenschein 2006; Caragiannis and Procaccia 2011; Boutilier et al. 2015; Caragiannis et al. 2017; Benade et al. 2017). The other (which is more recent) assumes that utility functions are derived from an underlying metric space, naturally leading to smaller distortion (Anshelevich, Bhardwaj, and Postl 2015; Anshelevich and Postl 2017; Feldman, Fiat, and Golumb 2016; Goel, Krishnaswamy, and Munagala 2017; Gross, Anshelevich, and Xia 2017). Our setup is consistent with the former line of work.

On a technical level, two of the foregoing papers are most closely related to ours. The first is by Boutilier et al. (2015), who study the distortion minimization problem when the output is a distribution over winning alternatives. They prove an upper bound of $\mathcal{O}\left(\sqrt{m} \cdot \log^* m\right)$ on the distortion of optimal social choice functions, and a lower bound of $\Omega(\sqrt{m})$. Their setting coincides with ours when $w_{i,1} = 1$ for each voter i, because in that case social welfare depends only on the utility of each voter for the top-ranked alternative. Achieving low distortion is much more difficult in our setting, and, in particular, their lower bound directly carries over (whereas their upper bound clearly does not).

The second paper, by Benade et al. (2017), studies distortion-minimizing rules for the participatory budgeting problem, where each alternative has a cost, and the goal is to choose a subset of alternatives that satisfies a budget constraint. Voters are assumed to have *additive* utility functions. Their results are incomparable to ours — their problem is 'harder' in that they have to deal with (known) costs and budget constraints, but 'easier' in that they choose a single subset, whereas we, in a sense, choose m nested subsets (the m prefixes of our ranking), which are weighted according to *unknown* weights. Furthermore, our results hold for richer (subadditive) combinatorial utility functions.

2 The Model

Our setting involves a set of voters $[n] = \{1, \ldots, n\}$, and a set of alternatives $[m] = \{1, \ldots, m\}$. We are interested in the set $\mathcal{S}_m$ of rankings, or permutations, over $[m]$. We think of a ranking $\tau \in S_m$ as a function from positions to alternatives, i.e., $\tau(j)$ is the alternative in position j in τ, and $\tau^{-1}(j)$ is the position in which τ places alternative j.

The preferences of each voter i are represented as a ranking $\sigma_i \in \mathcal{S}_m$. A *preference profile* is a vector $\vec{\sigma} = (\sigma_1, \ldots, \sigma_n)$ of the rankings of all voters.

A (randomized) *social choice function* is a function $f : (\mathcal{S}_m)^n \to \Delta([m])$, which takes a preference profile as input, and returns a distribution over winning alternatives. In this paper we focus on (randomized) *social welfare functions*, whose range is instead $\Delta(\mathcal{S}_m)$, i.e., they also take a preference profile as input, but return a distribution over rankings.

A novel component of our model is that we assume that each voter $i \in [n]$ is associated with a *combinatorial* utility function $u_i : 2^{[m]} \to \mathbb{R}_+$ and a *weight vector* $w_i \in \mathbb{R}_+^m$. Following previous work (Caragiannis and Procaccia 2011; Boutilier et al. 2015; Caragiannis et al. 2017; Benade et al. 2017), both are assumed to be *normalized*, that is, for all $i \in [n]$, $u_i(\emptyset) = 0$ and $\sum_{j=1}^m u_i(\{j\}) = \sum_{j=1}^m w_{i,j} = 1$. Moreover, our results make use of the following properties of utility functions:

- *Monotonicity:* $u_i(S) \leq u_i(T)$ for all $S \subseteq T \subseteq [m]$.
- *Subadditivity:* $u_i(S) + u_i(T) \geq u_i(S \cup T)$ for all $S, T \subseteq [m]$.

The utility of voter i for a ranking $\tau \in \mathcal{S}_m$ is given by the weighted sum of his utilities for the prefixes of τ, that is,

$$u_i(\tau) = \sum_{j=1}^{m} w_{i,j} \cdot u_i(\{\tau(1), \tau(2), \ldots, \tau(j)\}).$$

We remark that even *additive* utility functions are able to capture the simpler setting discussed in Section 1, which can be formalized by assigning each voter a utility function $u'_i : [m] \to \mathbb{R}_+$ and weights such that $w'_{i,j} \geq w'_{i,j+1}$ for all $j \in [m-1]$, and letting $u'_i(\tau) = \sum_{j=1}^m w'_{i,j} \cdot u'_i(\tau(j))$. But subadditive utility functions can model realistic settings that are not captured by utility functions that are merely additive. For example, suppose that we want to rank faculty candidates, and plan to make offers to the top candidates in our ranking. Then the weight $w_{i,j}$ reflects the perceived probability that there would be j slots available, and the utility function u_i is naturally subadditive, but not additive, due to complementarities between candidates: The contribution of a candidate is typically small if the set of candidates already contains another candidate working in the exact same area.

We assume that each voter reports a ranking that is *consistent* with his utility function, which, in our general formulation with combinatorial utilities, we take to mean that voter i reports σ_i only if

$$u_i(\{\sigma_i(1)\}) \geq u_i(\{\sigma_i(2)\}) \geq \cdots \geq u_i(\{\sigma_i(m)\}).$$

We denote this notion of consistency by $u_i \triangleright \sigma_i$, and, when σ_i is consistent with u_i for all $i \in [n]$, $\vec{u} \triangleright \vec{\sigma}$.

Our goal is to optimize (utilitarian) *social welfare*, that is, the sum of utilities voters have for the output ranking. Formally,

$$\mathsf{sw}(\tau) \triangleq \sum_{i=1}^{n} u_i(\tau).$$

However, since we only observe the given preference profile, we cannot directly optimize social welfare. To measure how far a social welfare function is from maximizing this objective, we adapt the concept of *distortion* (Procaccia and Rosenschein 2006). Formally, the distortion of a social welfare function f on a preference profile $\vec{\sigma}$ is

$$\mathsf{dist}(f, \vec{\sigma}) \triangleq \max_{\vec{u}:\, \vec{u} \triangleright \vec{\sigma}} \max_{\vec{w}} \frac{\max_{\tau \in \mathcal{S}_m} \mathsf{sw}(\tau)}{\mathbb{E}_{\mu \sim f(\vec{\sigma})}[\mathsf{sw}(\mu)]}.$$

In words, distortion measures the ratio between the social welfare of the welfare-maximizing ranking, and the expected social welfare of the distribution over rankings produced by f, in the worst case over all possible weights $\vec{w} = (w_{i,j})_{i \in [n], j \in [m]}$, and all possible utility profiles that are consistent with the given preference profile. Finally, the distortion of f is the worst case distortion over all possible preference profiles: $\mathsf{dist}(f) \triangleq \max_{\vec{\sigma}} \mathsf{dist}(f, \vec{\sigma})$.

3 Distortion Bound

In this section we establish a tight (up to polylogarithmic factors) bound on the distortion of optimal social welfare functions. As noted under related work, Boutilier et al. (2015) prove a lower bound of $\Omega(\sqrt{m})$ on the distortion of optimal social *choice* functions, which carries over to our setting. Therefore, to show that optimal social welfare functions have distortion $\tilde{\Theta}(\sqrt{m})$, it is sufficient to prove the following theorem, which is our main result.

Theorem 3.1. *Under the monotonicity and subadditivity assumptions, there exists a randomized social welfare function with distortion* $\mathcal{O}\left(\sqrt{m} \ln^{3/2} m\right)$.

The construction of our social welfare function relies on the *harmonic scoring function* (Boutilier et al. 2015), defined as follows. Recall that $\sigma_i^{-1}(j)$ denotes the position of alternative j in the ranking of voter i. The harmonic score of alternative j is $\mathsf{score}(j) \triangleq \sum_{i=1}^{n} 1/\sigma_i^{-1}(j)$.

We will make use of the following two properties of the harmonic scoring function.

Lemma 3.2. *For any* $m \ge 2$, $\sum_{j=1}^{m} \mathsf{score}(j) \le 3n \ln m$.

Proof of Lemma 3.2. By definition,

$$\sum_{j=1}^{m} \mathsf{score}(j) = \sum_{j=1}^{m} \sum_{i=1}^{n} 1/\sigma_i^{-1}(j) = \sum_{i=1}^{n} \sum_{j=1}^{m} 1/j \le n(\ln m + 1) \le 3n \ln m.$$

□

Lemma 3.3. *Under the subadditivity assumption, for any* $S \subseteq [m]$ *it holds that* $\sum_{i=1}^{n} u_i(S) \le \sum_{j \in S} \mathsf{score}(j)$.

Proof of Lemma 3.3. For any voter $i \in [n]$ and alternative $a \in [m]$,

$$1 = \sum_{j=1}^{m} u_i(\sigma_i(j)) \ge \sum_{j=1}^{\sigma_i^{-1}(a)} u_i(\sigma_i(j)) \ge \sigma_i^{-1}(a) \cdot u_i(\{a\}).$$

Thus, $u_i(\{a\}) \le 1/\sigma_i^{-1}(a)$. Moreover, by the subadditivity of u_i, $u_i(S) \le \sum_{j \in S} u_i(\{j\})$. It follows that

$$\sum_{i=1}^{n} u_i(S) \le \sum_{i=1}^{n} \sum_{j \in S} u_i(\{j\}) \le \sum_{j \in S} \sum_{i=1}^{n} 1/\sigma_i^{-1}(j) = \sum_{j \in S} \mathsf{score}(j).$$

□

We require one other lemma that is quite technical. Denote by $T \overset{k}{\leftarrow} S$ the experiment of drawing a subset T of size k from S uniformly at random.

Lemma 3.4. *Suppose* $A \subseteq B \cap C$ *and* $k \le |B| \le |C|$. *Function* $g : 2^A \to \mathbb{R}_+$ *satisfies the monotonicity and subadditivity conditions. Then*

$$|B| \cdot \mathop{\mathbb{E}}_{T \overset{k}{\leftarrow} B} [g(T \cap A)] \le 4|C| \cdot \mathop{\mathbb{E}}_{T \overset{k}{\leftarrow} C} [g(T \cap A)].$$

Proof of Lemma 3.4. Fix set A, integer k, and function $g : 2^A \to \mathbb{R}_+$ that satisfies the monotonicity and subadditivity conditions. For $n \ge \max(k, |A|)$, define $f(n)$ as

$$f(n) \triangleq n \cdot \mathop{\mathbb{E}}_{T \overset{k}{\leftarrow} S_n} [g(T \cap A)],$$

where S_n is a superset of A with n elements.[1] It suffices to prove that $f(n)$ is *approximately non-decreasing* in the sense that for any $n_1 \le n_2$, $f(n_1) \le 4f(n_2)$.

Define $a_j \triangleq \mathbb{E}_{T \overset{j}{\leftarrow} A}[g(T)]$ for $0 \le j \le |A|$, and let $a_j = a_{|A|}$ for $j > |A|$. We require the following property, which we prove in Appendix A.

Lemma 3.5. *The sequence* $(a_j)_{j=0}^{\infty}$ *is monotonic and subadditive, and, moreover,* $2a_k \ge a_{2k}$ *for any* $1 \le k \le |A|$.

For each $j \in [|A|]$, define b_j as $b_j \triangleq a_{j'}/j'$, where $j' = 2^{\lceil \log_2 j \rceil}$. Moreover, let $b_{|A|+1} = 0$. We show that $(b_j)_{j=1}^{|A|}$ is non-increasing and approximates $(a_j/j)_{j=1}^{|A|}$.

By construction,

$$(b_j)_{j=1}^{|A|} = \left(\frac{a_1}{1}, \frac{a_2}{2}, \frac{a_4}{4}, \frac{a_4}{4}, \frac{a_8}{8}, \ldots\right).$$

Since $2a_k \ge a_{2k}$ for any $k \in [|A|]$ by Lemma 3.5, we have

$$\frac{a_1}{1} \ge \frac{a_2}{2} \ge \frac{a_4}{4} \ge \cdots$$

This proves that $(b_j)_{j=1}^{|A|}$ is non-increasing.

Let $j' = 2^{\lceil \log_2 j \rceil}$. Since $j \le j' \le 2j$, it follows from the monotonicity and subadditivity of $\{a_j\}$ (Lemma 3.5) that $a_j \le a_{j'} \le a_{2j} \le 2a_j$. Therefore,

$$\frac{a_j}{2j} \le \frac{a_{j'}}{j'} \le \frac{2a_j}{j}, \tag{1}$$

i.e., b_j approximates a_j/j up to a factor of 2.

[1] Any such set S_n gives the same definition of $f(n)$.

Next, fix $n \geq \max(k, |A|)$. Recall that

$$f(n) = n \cdot \mathop{\mathbb{E}}_{T \xleftarrow{k} S_n} [g(T \cap A)] .$$

For $0 \leq j \leq |A|$, let $\mathcal{E}_j$ denote the event that $|T \cap A| = j$. When conditioning on $\mathcal{E}_j$, $T \cap A$ is uniformly distributed among all subsets of size j in A, i.e., $\mathbb{E}[g(T \cap A)|\mathcal{E}_j] = a_j$. Moreover,

$$\Pr[\mathcal{E}_j] = \frac{\binom{|A|}{j}\binom{n-|A|}{k-j}}{\binom{n}{k}}.$$

By the law of total expectation,

$$\begin{aligned} f(n) &= n \sum_{j=0}^{|A|} \Pr[\mathcal{E}_j] \cdot \mathbb{E}[g(T \cap A)|\mathcal{E}_j] \\ &= n \sum_{j=1}^{|A|} \frac{\binom{|A|}{j}\binom{n-|A|}{k-j}}{\binom{n}{k}} \cdot a_j \\ &= n \sum_{j=1}^{|A|} \frac{\frac{|A|}{j}\binom{|A|-1}{j-1}\binom{n-|A|}{k-j}}{\frac{n}{k}\binom{n-1}{k-1}} \cdot a_j \\ &= k|A| \sum_{j=1}^{|A|} \frac{\binom{|A|-1}{j-1}\binom{n-|A|}{k-j}}{\binom{n-1}{k-1}} \cdot \frac{a_j}{j}, \end{aligned} \tag{2}$$

where the second equality holds because $a_0 = g(\emptyset) = 0$, and the third because $\binom{n}{m} = \frac{n}{m}\binom{n-1}{m-1}$ for $1 \leq m \leq n$.

Let

$$p_{n,j} = \frac{\binom{|A|-1}{j-1}\binom{n-|A|}{k-j}}{\binom{n-1}{k-1}}$$

and $q_{n,j} = \sum_{l=1}^{j} p_{n,l}$. Recall that $b_{|A|+1} = 0$. Using Equations (1) and (2), we have

$$\begin{aligned} f(n) &= k|A| \sum_{j=1}^{|A|} p_{n,j} \cdot \frac{a_j}{j} \\ &\leq 2k|A| \sum_{j=1}^{|A|} p_{n,j} \cdot b_j \\ &= 2k|A| \sum_{j=1}^{|A|} (b_j - b_{j+1})\, q_{n,j}. \end{aligned}$$

Similarly, we have

$$f(n) \geq \frac{1}{2} k|A| \sum_{j=1}^{|A|} (b_j - b_{j+1})\, q_{n,j}.$$

In Appendix B, we prove:

Lemma 3.6. *For any $j \in [|A|]$, $(q_{n,j})_{n=\max(k,|A|)}^{\infty}$ is non-decreasing in n.*

This completes the proof of Lemma 3.4, as for any $\max(k, |A|) \leq n_1 \leq n_2$, we have

$$\begin{aligned} f(n_1) &\leq 2k|A| \sum_{j=1}^{|A|} (b_j - b_{j+1})\, q_{n_1,j} \\ &\leq 2k|A| \sum_{j=1}^{|A|} (b_j - b_{j+1})\, q_{n_2,j} \\ &\leq 4f(n_2). \end{aligned}$$

□

We are now ready to prove the theorem.

Proof of Theorem 3.1. We construct a randomized social welfare function that, given a preference profile $\vec{\sigma}$, proceeds as follows.

- Sort the alternatives into a ranking ν such that $\mathsf{score}(\nu(1)) \geq \mathsf{score}(\nu(2)) \geq \cdots \geq \mathsf{score}(\nu(m))$.
- Let $t_{\max} = \lceil \log_2 m \rceil$ and $\alpha = \sqrt{m \ln m}$. Draw t uniformly at random from $[t_{\max}]$ and set $m'_t = \min(\lfloor 2^t \alpha \rfloor, m)$.
- With probability $1/2$, return a uniformly random permutation of $[m]$. Otherwise, shuffle the first m'_t elements of ν uniformly at random, and return the resulting ordering.

The rest of the proof analyzes the distortion of the foregoing function. By the monotonicity of utility functions, the social welfare of every ranking $\tau \in \mathcal{S}_m$ is at least

$$\begin{aligned} \mathsf{sw}(s) &= \sum_{i=1}^{n} \sum_{j=1}^{m} w_{i,j} \cdot u_i(\{\tau(1), \tau(2), \ldots, \tau(j)\}) \\ &\geq \sum_{i=1}^{n} \sum_{j=1}^{m} w_{i,j} \cdot u_i(\{\tau(1)\}) = \sum_{i=1}^{n} u_i(\{\tau(1)\}), \end{aligned}$$

where the last transition follows from $\sum_{j=1}^{m} w_{i,j} = 1$.

When the mechanism returns a random permutation τ, $\tau(1)$ is uniformly distributed in $[m]$, and thus the expected social welfare is at least

$$\frac{1}{m} \sum_{\tau(1)=1}^{m} \sum_{i=1}^{n} u_i(\{\tau(1)\}) = \frac{1}{m} \sum_{i=1}^{n} \sum_{j=1}^{m} u_i(\{j\}) = \frac{n}{m}.$$

On the other hand, consider the case where the mechanism randomly shuffles the first m'_t elements in ν. Let $m_t \triangleq \min\{2^t, m\}$, and $R_t \triangleq \{\nu(1), \nu(2), \ldots, \nu(m'_t)\}$. The resulting expected social welfare is at least

$$\sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} R_t} [u_i(T)].$$

Let SOL_t denote the expected social welfare conditioning on the value of t. Then the above discussion implies that

$$\mathsf{SOL}_t \geq \frac{n}{2m} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} R_t} [u_i(T)]. \tag{3}$$

Let μ^* denote the welfare-maximizing ranking. Let $S_t \triangleq \{\mu^*(1), \mu^*(2), \ldots, \mu^*(m_t)\}$, and

$$\mathsf{OPT}_t \triangleq \sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} [u_i(T)].$$

In the following, we show that

$$\sum_{t=1}^{t_{\max}} \mathsf{OPT}_t \geq \frac{\mathsf{sw}(\mu^*)}{2}, \tag{4}$$

and for any $t \in [t_{\max}]$,

$$\mathsf{SOL}_t \geq \frac{\mathsf{OPT}_t}{12\sqrt{m \ln m}}. \tag{5}$$

Inequalities (4) and (5) directly imply that the expected social welfare obtained by the mechanism is at least

$$\begin{aligned}\frac{1}{t_{\max}} \sum_{t=1}^{t_{\max}} \mathsf{SOL}_t &\geq \frac{1}{\lceil \log_2 m \rceil} \sum_{t=1}^{t_{\max}} \frac{\mathsf{OPT}_t}{12\sqrt{m \ln m}} \\ &\geq \frac{\mathsf{sw}(\mu^*)}{\mathcal{O}\left(\sqrt{m} \ln^{3/2} m\right)},\end{aligned}$$

which concludes the proof.

Proof of Equation (4). Note that for any $t \in [t_{\max}]$ and $j \in [m_t/2, m_t]$,

$$\begin{aligned}\mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} [u_i(T)] &\geq \mathop{\mathbb{E}}_{T \xleftarrow{m_t/2} S_t} [u_i(T)] \geq \mathop{\mathbb{E}}_{T \xleftarrow{m_t} S_t} [u_i(T)] \cdot \frac{1}{2} \\ &= u_i(S_t) \cdot \frac{1}{2} \\ &\geq u_i(\{\mu^*(1), \mu^*(2), \ldots, \mu^*(j)\}) \cdot \frac{1}{2}, \end{aligned} \tag{6}$$

where the first transition follows from the monotonicity of u_i, the second from its subadditivity, the third from $|S_t| = m_t$, and the last again from monotonicity. Therefore,

$$\begin{aligned}\sum_{t=1}^{t_{\max}} \mathsf{OPT}_t &\geq \sum_{t=1}^{t_{\max}} \sum_{i=1}^{n} \sum_{j=m_t/2}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} [u_i(T)] \\ &\geq \frac{1}{2} \sum_{i=1}^{n} \sum_{t=1}^{t_{\max}} \sum_{j=m_t/2}^{m_t} w_{i,j} \cdot u_i(\{\mu^*(1), \ldots, \mu^*(j)\}) \\ &\geq \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{i,j} \cdot u_i(\{\mu^*(1), \ldots, \mu^*(j)\}) \\ &= \frac{\mathsf{sw}(\mu^*)}{2},\end{aligned}$$

where the second inequality follows from Equation (6), and the third holds because $m_1/2 = 1$ and $m_{t_{\max}} = m$.

Proof of Equation (5). Let $S_t^+ = S_t \cap R_t$ and $S_t^- = S_t \setminus R_t$. The subadditivity of u_i implies that for any $T \subseteq S_t$, $u_i(T) \leq u_i(T \cap S_t^+) + u_i(T \cap S_t^-)$. Thus, we can derive an upper bound on OPT_t as follows:

$$\begin{aligned}\mathsf{OPT}_t &= \sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} [u_i(T)] \\ &\leq \sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} \left[u_i(T \cap S_t^+) + u_i(T \cap S_t^-)\right] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} \left[u_i(T \cap S_t^+)\right] \\ &\quad + \sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} \left[u_i(T \cap S_t^-)\right]. \end{aligned} \tag{7}$$

We establish upper bounds on the two terms on the right hand side of Equation (7) separately. For the first term, note that $S_t^+ \subseteq S_t \cap R_t$ and $|S_t| = m_t \leq m_t' = |R_t|$. For any $i \in [n]$ and $j \in [m_t]$, applying Lemma 3.4 with $g = u_i$, $k = j$, $A = S_t^+$, $B = S_t$ and $C = R_t$ gives

$$|S_t| \mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} \left[u_i(T \cap S_t^+)\right] \leq 4|R_t| \mathop{\mathbb{E}}_{T \xleftarrow{j} R_t} \left[u_i(T \cap S_t^+)\right].$$

It follows that

$$\begin{aligned}\mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} \left[u_i(T \cap S_t^+)\right] &\leq \frac{4m_t'}{m_t} \mathop{\mathbb{E}}_{T \xleftarrow{j} R_t} \left[u_i(T \cap S_t^+)\right] \\ &\leq 4\alpha \mathop{\mathbb{E}}_{T \xleftarrow{j} R_t} \left[u_i(T \cap S_t^+)\right].\end{aligned}$$

Summation over i and j yields

$$\begin{aligned}&\sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} \left[u_i(T \cap S_t^+)\right] \\ &\leq 4\alpha \sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} R_t} \left[u_i(T \cap S_t^+)\right] \\ &\leq 8\alpha \cdot \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} R_t} [u_i(T)]. \end{aligned} \tag{8}$$

We next bound the second term on the right hand side of Equation (7). Note that

$$\begin{aligned}&\sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} \left[u_i(T \cap S_t^-)\right] \\ &\leq \sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot u_i(S_t^-) \leq \sum_{i=1}^{n} u_i(S_t^-) \leq \sum_{j \in S_t^-} \mathsf{score}(j).\end{aligned}$$

Here the first step is due to the monotonicity of u_i, the second step holds since $\sum_{j=1}^{m_t} w_{i,j} \leq \sum_{j=1}^{m} w_{i,j} = 1$, while last step applies Lemma 3.3. For each alternative $a \in S_t^-$, it follows from Lemma 3.2 that

$$3n \ln m \geq \sum_{j=1}^{m} \mathsf{score}(j) \geq \sum_{j=1}^{m_t'} \mathsf{score}(\nu(j)) \geq m_t' \cdot \mathsf{score}(a),$$

so $\mathsf{score}(a) \leq 3n \ln m / m_t'$ for any $a \in S_t^-$. Therefore, we have

$$\sum_{i=1}^{n} \sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \xleftarrow{j} S_t} \left[u_i(T \cap S_t^-)\right] \leq 3n \ln m \cdot \frac{|S_t^-|}{m_t'}.$$

Recall that $m_t' = \min\left(\lfloor 2^t \alpha \rfloor, m\right)$, and $S_t^- = S_t \setminus R_t = S_t \setminus \{\nu(1), \ldots, \nu(m_t')\}$. If $m_t' = m$, we have $S_t^- = \emptyset$ and

$|S_t^-|/m'_t = 0$. When $m'_t < m$, it holds that $m'_t = \lfloor 2^t\alpha \rfloor \geq 2^{t-1}\alpha$ and $m_t = 2^t$. Thus, $|S_t^-|/m'_t \leq m_t/m'_t \leq 2/\alpha$. In either case,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \overset{j}{\leftarrow} S_t}\left[u_i(T \cap S_t^-)\right] &\leq 3n \ln m \cdot \frac{|S_t^-|}{m'_t} \\ &\leq 3n \ln m \cdot \frac{2}{\alpha} \\ &= \frac{n}{2m} \cdot 12\alpha. \qquad (9)\end{aligned}$$

Putting everything together, we have that

$$\begin{aligned}\mathsf{OPT}_t \leq & \sum_{i=1}^{n}\sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \overset{j}{\leftarrow} S_t}\left[u_i(T \cap S_t^+)\right] \\ & + \sum_{i=1}^{n}\sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \overset{j}{\leftarrow} S_t}\left[u_i(T \cap S_t^-)\right] \\ & \leq 8\alpha \cdot \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m_t} w_{i,j} \cdot \mathop{\mathbb{E}}_{T \overset{j}{\leftarrow} R_t}\left[u_i(T)\right] + 12\alpha \cdot \frac{n}{2m} \\ & \leq 12\alpha \cdot \mathsf{SOL}_t = 12\sqrt{m \ln m} \cdot \mathsf{SOL}_t,\end{aligned}$$

where the first inequality follows from Equation (7), the second from (8) and (9), and the third from (3). This proves Equation (5) and completes the proof of the theorem. □

As is the case in previous work (Boutilier et al. 2015; Benade et al. 2017; Caragiannis et al. 2017), the social welfare function we analyze relies on randomization and returns a uniformly random outcome — in our case, a random ranking — with constant probability (note that any constant, even $1/10^6$, will do). It is important to note that this social welfare function should be viewed only as a tool to bound the distortion of the *optimal* social welfare function; we do not propose using it in practice. Instead, we compare the performance of the optimal (deterministic) social welfare function with several common competitors in Section 4.

Finally, we remark that some restriction on the combinatorial structure of the valuation functions, beyond monotonicity, is necessary to achieve sublinear distortion. Indeed, in the following example we construct non-subadditive utility functions such that any social welfare function must have distortion $\Omega(m)$.[2]

Example 3.7. Consider the following utility function: for two distinct alternatives $a, b \in [m]$,

$$u_{a,b}(S) = \begin{cases} 1, & \{a, b\} \subseteq S, \\ |S|/m, & \text{otherwise.} \end{cases}$$

Note that $u_{a,b}$ is monotonic. Moreover, the function is consistent with any ranking of $[m]$, so we can assume that a voter with utility function $u_{a,b}$ reports the same ranking σ regardless of a and b.

[2]Ranking the alternatives uniformly at random achieves distortion $\mathcal{O}(m)$. Thus, in such cases we cannot significantly outperform a random guess.

Let there be a single voter with weight vector $(0, 1, 0, 0, \ldots)$. The utility of a ranking τ is given by $u_{a,b}(\{\tau(1), \tau(2)\})$. In order to achieve a utility of 1 (rather than $2/m$), it is necessary to place a and b in the top two slots. Any randomized welfare function has two alternatives that, given σ as input, are placed in the first two positions with probability at most $2/m(m-1)$. By choosing these two alternatives to be a and b, we can guarantee that the function achieves expected social welfare at most $\frac{2}{m(m-1)} \cdot 1 + (1 - \frac{2}{m(m-1)}) \cdot \frac{2}{m}$, whereas the optimum is 1. The ratio is $\Omega(m)$.

4 Empirical Results

For our computational experiments we focus on deterministic social welfare functions, and on additive utility functions — but we generalize the position-weighted model slightly. Let voter $i \in [n]$ have ranking σ_i consistent with the utility *matrix* $U^i \in \mathbb{R}^{m \times m}$, where U^i_{ap} is the utility voter i has for alternative a appearing in position p. As before, voter i's preferences impose constraints on U^i. Specifically, higher ranked alternatives have utility at least as large as lower ranked alternatives, for any specific position, that is, $U^i_{\sigma_i(p),j} \geq U^i_{\sigma_i(p+1),j}$ for all $p \in [m-1]$, $j \in [m]$, and $U^i_{a,p} \geq U^i_{a,p+1}$ for all $a \in [m]$, $p \in [m-1]$. Utilities are normalized to have $\sum_{a \in [m]} \sum_{p \in [m]} U^i_{ap} = 1$. The *utility profile* is $\vec{U} = (U^1, \ldots, U^n)$.

Let us represent a ranking τ by a permutation matrix $X(\tau) \in \Pi_m$. The social welfare of a ranking τ is $\sum_{i \in [n]} \langle U^i, X(\tau) \rangle$, where $\langle A, B \rangle = \sum_{ij} A_{ij} B_{ij}$ is the *Frobenius inner product*. We can now write the mathematical program that finds the (deterministic) ranking X with minimum distortion z given an input profile $\vec{\sigma}$ as

$$\begin{aligned}\min_{z, \tau \in \mathcal{S}_m} \quad & z \\ z \quad & \geq \frac{\sum_{i=1}^n \langle U^i, X(\rho) \rangle}{\sum_{i=1}^n \langle U^i, X(\tau) \rangle} \quad \forall \vec{U} \triangleright \vec{\sigma}\,,\ \rho \in \mathcal{S}_m \qquad (10)\end{aligned}$$

This formulation has intractably many constraints in Equation (10). But these constraints may be omitted and added as needed, by solving the subproblem

$$\min_{\vec{U} \triangleright \vec{\sigma}, \rho \in \mathcal{S}_m} \qquad \bar{z} \cdot \sum_{i=1}^{n} \langle U^i, X(\bar{\tau}) \rangle - \sum_{i=1}^{n} \langle U^i, X(\rho) \rangle$$

where $\bar{z}, \bar{\tau}$ are the current optimal solutions to the master problem. A violated constraint is found if the objective function value of the subproblem is strictly less than 0. The procedure terminates with the optimal z, τ when no violated constraints are found.

The subproblem is nonconvex even when the integrality constraints are relaxed, and, therefore, finding violated constraints is computationally expensive. Nevertheless, Figure 1 shows that it is currently practical to compute distortion-minimizing rankings exactly for instances with up to 10 alternatives within a couple of minutes. We expect that this will be sufficient for the vast majority of instances seen in practice. Indeed, 98.3% of the instances submitted to RoboVote (as of January 19, 2018) have 10 or fewer alternatives.

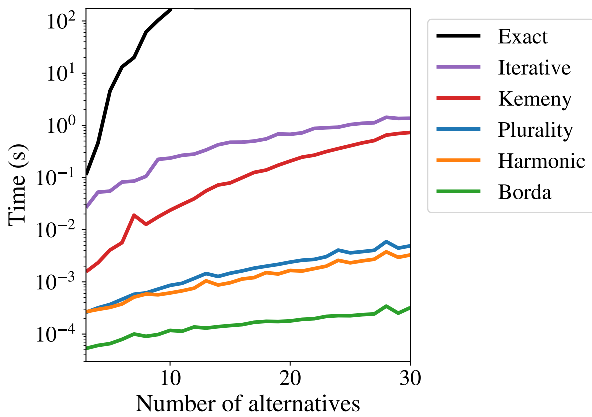

Figure 1: Runtime (in seconds) for increasing instance size, on an a machine with an Intel Core i5-4200U CPU and 8 GB RAM.

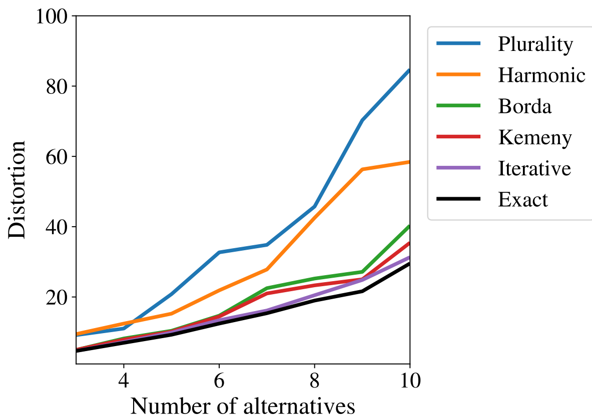

Figure 2: Average distortion of heuristic and exact methods.

For larger instances, we evaluate the performance of the following, more scalable, heuristics:

1. *Kemeny*: Return the ranking that minimizes the total number of disagreements on pairs of alternatives with the input profile.
2. *Borda*: Rank alternatives by their Borda scores, defined as $\sum_{i=1}^{n}(m - \sigma_i^{-1}(a))$.
3. *Plurality*: Rank alternatives based on the number of times they are ranked first. Break ties by considering subsequent positions.
4. *Harmonic*: Return a ranking according to Theorem 3.1.
5. *Iterative*: Iteratively find and remove the alternative that minimizes distortion for the problem of returning a single alternative with maximum social welfare.

We evaluate these heuristics on instances with $n = 10$ and $m \in \{3, \dots, 30\}$. Every alternative a is assigned a *quality* c_a, and $u_i(\{a\})$ is drawn from a truncated normal distribution around c_a. Vector $u_i = (u_i(\{a\}))_{a\in[m]}$ induces σ_i. Weights w_i are drawn uniformly at random in $[0, 1]$, and ordered. Voter i's utility matrix $U^i = w_i u_i$ is normalized.

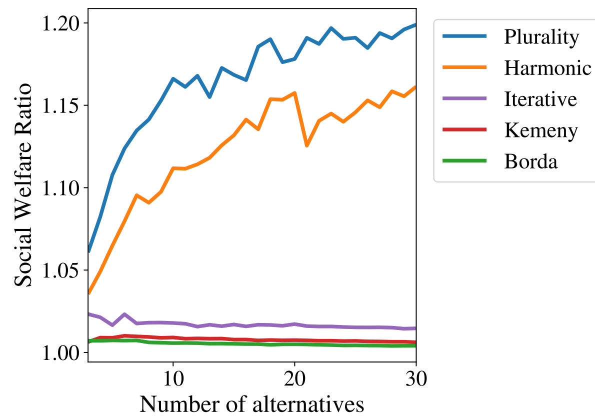

Figure 3: Average social welfare ratio of different heuristics.

Every social welfare function f only has access to $\vec{\sigma}$ and is evaluated on two metrics: the distortion of the returned ranking $\rho = f(\vec{\sigma})$, and the *social welfare ratio* $\max_{\tau\in\mathcal{S}_m} \mathsf{sw}(\tau, \vec{U})/\mathsf{sw}(\rho, \vec{U})$. Note that the latter measure estimates the *average case* with respect to utility profiles.

The distortion and social welfare ratios of the proposed heuristics are shown in Figures 2 and 3. Distortion is reported for $m \le 10$, where it is possible to compare to the optimal distortion, and 100 repetitions; social welfare for $m \le 30$ and 200 repetitions.

The distortion of Borda, Kemeny and especially Iterative compares well with the optimal distortion. Kemeny and Borda also lead to very high efficiency, with average social welfare within 1% of optimal.

5 Discussion

Much like previous papers on implicit utilitarian voting (Caragiannis et al. 2017; Benade et al. 2017), there is a certain gap between the theoretical and empirical results, in the sense that the theoretical distortion bound of Theorem 3.1 holds for *randomized* social welfare functions, whereas the empirical results hold for deterministic functions. The value of theoretical distortion bounds is that they tell us whether rankings inherently provides useful information for optimizing social welfare. The fact that the bound is essentially no worse than for the case of a single winner means that the implicit utilitarian voting approach does extend to the design of social welfare functions.

On a practical level, our empirical results suggest that classic methods like the Kemeny rule (which is currently deployed on RoboVote) and Borda count provide near-optimal performance from the viewpoint of implicit utilitarian voting. Alternatively, it is possible to compute the distortion-minimizing social welfare function if instances are restricted to at most ten alternatives. Although almost all instances arising from small-group decisions (of the type made on RoboVote) are of that size, some high-stakes decisions, such as ranking applicants for a job or candidates for a PhD program, involve a much larger number of alternatives, and motivate the development of faster algorithms.

References

Anshelevich, E., and Postl, J. 2017. Randomized social choice functions under metric preferences. *Journal of Artificial Intelligence Research* 58:797–827.

Anshelevich, E.; Bhardwaj, O.; and Postl, J. 2015. Approximating optimal social choice under metric preferences. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*, 777–783.

Benade, G.; Nath, S.; Procaccia, A. D.; and Shah, N. 2017. Preference elicitation for participatory budgeting. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 376–382.

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

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

Caragiannis, I., and Procaccia, A. D. 2011. Voting almost maximizes social welfare despite limited communication. *Artificial Intelligence* 175(9–10):1655–1671.

Caragiannis, I.; Nath, S.; Procaccia, A. D.; and Shah, N. 2017. Subset selection via implicit utilitarian voting. *Journal of Artificial Intelligence Research* 58:123–152.

Conitzer, V.; Davenport, A.; and Kalagnanam, H. 2006. Improved bounds for computing Kemeny rankings. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence (AAAI)*, 620–626.

Davenport, A., and Kalagnanam, J. 2004. A computational study of the Kemeny rule for preference aggregation. In *Proceedings of the 19th AAAI Conference on Artificial Intelligence (AAAI)*, 697–702.

Feldman, M.; Fiat, A.; and Golumb, I. 2016. On voting and facility location. In *Proceedings of the 17th ACM Conference on Economics and Computation (EC)*, 269–286.

Goel, A.; Krishnaswamy, A. K.; and Munagala, K. 2017. Metric distortion of social choice rules: Lower bounds and fairness properties. In *Proceedings of the 18th ACM Conference on Economics and Computation (EC)*, 287–304.

Gross, S.; Anshelevich, E.; and Xia, L. 2017. Vote until two of you agree: Mechanisms with small distortion and sample complexity. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 544–550.

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

A Proof of Lemma 3.5

Before proving our claim that the sequence $(a_j)_{j=0}^{\infty}$ is monotonic and subadditive, with $2a_k \ge a_{2k}$ for any $1 \le k \le |A|$, we introduce two lemmas that establish some properties of monotonic and subadditive functions. The following lemma states that, given a monotonic utility function defined on set S, when a subset of size k is drawn from S uniformly at random, the expected utility of the resulting subset is non-decreasing in k.

Lemma A.1. *Suppose function $g : 2^S \to \mathbb{R}_+$ satisfies the monotonicity condition. Define $(a_k)_{k=0}^{|S|}$ as $a_k \triangleq \mathbb{E}_{T \xleftarrow{k} S}[g(T)]$. Then,*

$$a_0 \le a_1 \le \cdots \le a_{|S|}.$$

Proof of Lemma A.1. Fix integer k between 1 and $|S|$. Suppose we draw $X \xleftarrow{k} S$, choose x from X uniformly at random, and then let $Y = X \setminus \{x\}$. Note that X and Y are uniformly (yet not independently) distributed in all subsets of S of size k and $k-1$, respectively. Since $Y \subseteq X$, the monotonicity of g implies that $g(Y) \le g(X)$. Taking the expectation yields $a_{k-1} \le a_k$. □

The following lemma states that the sequence $(a_k)_{k=0}^{|S|}$ defined as in Lemma A.1 is subadditive, assuming the subadditivity of function g.

Lemma A.2. *Suppose function $g : 2^S \to \mathbb{R}_+$ satisfies the subadditivity condition. Let $a_k = \mathbb{E}_{T \xleftarrow{k} S}[g(T)]$. Then for any integers $n, m \ge 0$ that satisfy $n + m \le |S|$,*

$$a_n + a_m \ge a_{n+m}.$$

Proof of Lemma A.2. Draw $X \xleftarrow{n} S$ and $Y \xleftarrow{m} S \setminus X$. Clearly, X and Y are uniformly distributed among all subsets of S with n elements and m elements, respectively. Moreover, $X \cup Y$ is also a uniformly random subset of size $n + m$. For each realization of X and Y, it follows from the subadditivity of g that

$$g(X) + g(Y) \ge g(X \cup Y).$$

Taking the expectation over the randomness in (X, Y) yields

$$a_n + a_m \ge a_{n+m}.$$

□

Proof of Lemma 3.5. Lemma A.1 implies that the finite sequence $(a_0, a_1, \ldots, a_{|A|})$ is non-decreasing. Since $a_j = a_{|A|}$ for any $j > |A|$, the complete sequence $(a_j)_{j=0}^{\infty}$ is also non-decreasing. Moreover, we claim that $2a_k \ge a_{2k}$ for any $1 \le k \le |A|$. In fact, if $2k \le |A|$, the inequality directly follows from Lemma A.2. If $2k > |A|$, by Lemmas A.1 and A.2,

$$2a_k \ge a_k + a_{|A|-k} \ge a_{|A|} = a_{2k}.$$

□

B Proof of Lemma 3.6

Fix set A, $j \in [|A|]$ and k. Recall that

$$p_{n,j} = \frac{\binom{|A|-1}{j-1}\binom{n-|A|}{k-j}}{\binom{n-1}{k-1}}$$

and $q_{n,j} = \sum_{l=1}^{j} p_{n,l}$. For every $n \ge \max(k, |A|)$, define random variable $T_n \xleftarrow{k-1} [n-1]$, i.e., T_n is a random subset of size $k-1$ drawn from $[n-1]$. It can be verified that

$$p_{n,j} = \Pr[|T_n \cap [|A|-1]| = j-1],$$

and thus,

$$q_{n,j} = \Pr[|T_n \cap [|A| - 1]| < j].$$

To show that $(q_{n,j})_{n=\max(k,|A|)}^{\infty}$ is non-decreasing in n, we consider the following experiment:

- Draw $X \xleftarrow{k-1} [n]$.
- Let $Y = X$ if $n \notin X$; otherwise, let $Y = X \setminus \{n\} \cup \{x\}$, where x is drawn uniformly from $[n-1] \setminus X$.

By construction, the marginal distributions of X and Y are identical to those of T_{n+1} and T_n, respectively. Moreover, as Y is either equal to X, or obtained from X by replacing n with a smaller element, we have

$$|X \cap [|A| - 1]| \leq |Y \cap [|A| - 1]|.$$

Therefore,

$$\mathbb{I}\left[|X \cap [|A| - 1]| < j\right] \geq \mathbb{I}\left[|Y \cap [|A| - 1]| < j\right],$$

where $\mathbb{I}\left[\cdot\right]$ denotes the indicator function.

Taking the expectation over the randomness in (X, Y) yields

$$\Pr[|T_{n+1} \cap [|A| - 1]| < j] \geq \Pr[|T_n \cap [|A| - 1]| < j],$$

i.e., $q_{n+1,j} \geq q_{n,j}$. This proves the monotonicity of $(q_{n,j})$, and thus completes the proof of the lemma. $\square$