Concentration of Distortion: The Value of Extra Voters in Randomized Social Choice

Brandon Fain[1], **William Fan**[2] and **Kamesh Munagala**[1]
[1]Department of Computer Science, Duke University, USA
[2]North Carolina School of Science and Mathematics, USA
{btfain, kamesh}@cs.duke.edu

Abstract

We study higher statistical moments of Distortion for randomized social choice in a metric implicit utilitarian model. The Distortion of a social choice mechanism is the expected approximation factor with respect to the optimal utilitarian social cost (OPT). The k^{th} moment of Distortion is the expected approximation factor with respect to the k^{th} power of OPT. We consider mechanisms that elicit alternatives by randomly sampling voters for their favorite alternative. We design two families of mechanisms that provide constant (with respect to the number of voters and alternatives) k^{th} moment of Distortion using just k samples if all voters can then participate in a vote among the proposed alternatives, or $2k-1$ samples if only the sampled voters can participate. We also show that these numbers of samples are tight. Such mechanisms deviate from a constant approximation to OPT with probability that drops exponentially in the number of samples, independent of the total number of voters and alternatives. We conclude with simulations on real-world Participatory Budgeting data to qualitatively complement our theoretical insights.

1 Introduction

For many problems in social choice, the number of alternatives is very large. For example, consider the problem of voting over possible budgets in a given municipality, where the number of alternatives is infinite (for a divisible budget) or exponential (for funding integral projects). In such settings, it may be impractical to elicit full rankings over alternatives from every voter. Instead, we may want to design mechanisms that only require voters to rank at most a constant number of alternatives. In this paper, we study such mechanisms.

We consider the standard problem in social choice wherein there is a set N of n voters and a set M of alternatives from which we must select a single winner. However, we assume that $|M|$ is large enough to prohibit eliciting full rankings over the alternatives. We also allow n to be large. We adopt the implicit utilitarian perspective with metric constraints [Boutilier *et al.*, 2015; Cheng *et al.*, 2017; Anshelevich and Postl, 2017; Goel *et al.*, 2017; Anshelevich *et al.*, 2018; Feldman *et al.*, 2016]. That is, we assume that voters have cardinal costs over alternatives, and these costs are constrained to be metric, but voters cannot directly report cardinal costs. We want to design social choice mechanisms to minimize the total social cost by only asking voters to rank at most a constant number of alternatives. We measure the efficiency of a mechanism as its *Distortion* (see Section 2), the worst case approximation to the total social cost.

It is easy to see that randomization is necessary to achieve constant Distortion if we cannot elicit the ordinal preferences of voters over all alternatives. One natural form of randomization is to elicit alternatives by randomly sampling voters and querying them for their favorite alternatives. More generally, in this paper we consider mechanisms of the following type: The set of alternatives will be the favorite alternatives expressed by a subset of the voters. Subsequently, these alternatives are ranked either (i) by the entire population of the voters or (ii) by a small subset of the voters. We refer to these as the full and limited participation models respectively.

These assumptions are not merely of theoretical interest, but model social choice in emergent domains. Assumption (i) is natural in contexts where all voters are entitled to participate in the final election. For instance, in real-world Participatory Budgeting applications (see Section 5), a small subset of individuals propose projects, but a much larger number participate in the subsequent vote. Assumption (ii) models situations where we want a lightweight social choice mechanism that only involves a small number of voters overall such as the many department level decisions made at universities by committees representing samples of the faculty.

Prior work [Anshelevich and Postl, 2017; Gross *et al.*, 2017; Fain *et al.*, 2019] analyzed simple social choice mechanisms for achieving constant Distortion. However, focusing on the *expected* Distortion can yield randomized mechanisms that can deviate significantly from their expectation ex-post, and hence may be risky to implement in practice.

We address this problem by considering higher moments of Distortion. The k^{th} moment of Distortion is the expected approximation factor with respect to the k^{th} power of the optimal utilitarian social cost. The goal of bounding higher moments of Distortion is directly analogous to providing high probability bounds on approximation guarantees with respect to the total social cost. We note that obtaining such a bound does not follow in a trivial manner from standard sampling

Participation Model	Lower bound	Upper bound
Full	k (Thm. 1)	k (Thm. 2)
Limited	$2k-1$ (Thm. 3)	$2k-1$ (Thm. 4)

Table 1: The number of samples of favorite alternatives of voters for achieving constant normalized k^{th} moment of Distortion.

arguments: The higher moments depend on the entire distribution of the Distortion obtained by the mechanism, and if this distribution has unbounded variance, then it is not possible to bound the second moment by a constant with any number of samples, let alone higher moments. Moreover, it is initially unclear how to take the "best" result out of many randomly sampled alternatives. Our key insight is that the metric assumption enables us to derive tight bounds on higher moments with only a few samples by using existing deterministic social choice rules to take the "best" from many randomly sampled alternatives.

1.1 Summary of Results

Our primary contribution is the development and analysis of randomized social choice mechanisms that achieve constant k^{th} moment of Distortion in the metric implicit utilitarian model while requiring each voter to rank at most $O(k)$ sampled alternatives, regardless of the total number of voters and alternatives. The normalized k^{th} moment of Distortion is defined formally in Section 2, and our results are summarized in Table 1. In particular, we design two families of mechanisms that have constant k^{th} moment of Distortion. The first asks just k randomly chosen voters for their favorite alternatives, assuming all n voters can subsequently participate in a vote among these alternatives. The second asks $2k-1$ voters for their favorite alternatives, and only these sampled voters participate in a vote among their favorite alternatives. To the best of our knowledge, these are the first results in implicit utilitarian social choice providing guarantees for arbitrarily high moments of Distortion and approximating the optimal social cost with high probability.

Additionally, we show that our upper bounds on the number of samples needed are tight. We show that the k^{th} moment of Distortion is unbounded in the following two settings: First, when we only sample $k-1$ favorite alternatives and all n voters can subsequently compare these alternatives, and secondly, when we only sample $2k-2$ voters and the entire mechanism uses only their favorite alternatives and their comparisons between these alternatives. From a practical perspective, we demonstrate the value of using additional voters and alternatives: At most two additional samples guarantee that another higher moment of Distortion can be bounded. Finally, in Section 5, we present simulations on real-world Participatory Budgeting data to qualitatively complement our theoretical insights.

1.2 Related Work

Metric Distortion

The Distortion of randomized social choice mechanisms in metrics is well studied [Boutilier *et al.*, 2015; Anshelevich and Postl, 2017; Goel *et al.*, 2017; Gross *et al.*, 2017]. The Random Dictatorship mechanism samples the favorite alternative of a single voter, and the 2-Agree mechanism [Gross *et al.*, 2017] samples at most $\min(n+1, m+1)$ favorite alternatives of voters. Random Dictatorship has Distortion at most 3 [Anshelevich and Postl, 2017], and 2-Agree improves this when m is small. Nothing better than Random Dictatorship is known if the goal is to minimize the Distortion. However, it is easy to show that such mechanisms do not have constant second (or higher) moment of Distortion [Fain *et al.*, 2019].

Using the second moment of Distortion as a proxy for risk was introduced in [Fain *et al.*, 2017; Fain *et al.*, 2019], where it was shown that making one sampled voter compare the favorite alternatives of two randomly sampled voters bounds the second moment of Distortion. In this paper, we consider the natural question: *What is the value of each additional voter in how well the Distortion concentrates?* We provide a tight characterization by bounding not just the second moment, but any higher moment of Distortion.

The extreme case where $k = n$ is the deterministic setting, where it is known that the Copeland mechanism, or any mechanism based on choosing from the uncovered set [Miller, 1977], yields Distortion of 5 [Anshelevich *et al.*, 2018]. This bound was improved to 4.236 in [Munagala and Wang, 2019] via a weighted generalization of the uncovered set. However, both of these methods require eliciting full ordinal preferences from voters.

Communication and Sample Complexity

For a more thorough survey on the complexity of eliciting ordinal preferences to implement social choice rules, we refer the interested reader to [Brandt *et al.*, 2016]. [Conitzer and Sandholm, 2005] comprehensively characterizes the *communication complexity* (in terms of the number of bits communicated) of common deterministic voting rules. [Bouveret *et al.*, 2017] and [Caragiannis and Procaccia, 2010] design social choice mechanisms with low communication complexity when there are a small number of voters, but potentially a large number of alternatives.

[Dey and Bhattacharyya, 2015; Dey and Narahari, 2015] study the sample complexity of predicting the outcome of deterministic social choice rules. However, a "sample" in this work is the entire ordinal preference list for a single voter, whereas a sample for us is only the top alternative for a given voter. Even then, they show that predicting the outcome of rules with small Distortion (such as Copeland) requires a number of samples that grows with the total number of alternatives. We show that a smaller number of more limited samples suffice to bound higher moments of Distortion.

Recently, [Mandal *et al.*, 2019] studied a different notion of communication complexity in a non-metric implicit utilitarian model where voters can communicate bits of information about their *cardinal* preferences. In this case, the baseline is ordinal voting, and the other extreme is communicating the entire set of cardinal utilities. They show tight results for how Distortion trades off with the communication complexity in terms of bits of information communicated per voter. In our setting, voters only convey ordinal information and we study the *sample complexity* to bound not just the Distortion but also how well it concentrates.

2 Preliminaries

We have a set N of n voters and a set M of alternatives, from which we must choose a single outcome. For each agent $i \in N$ and alternative $a \in M$, there is some underlying dis-utility $d(i,a) \geq 0$. Let $p_i = \text{argmin}_{a \in M} d(i,a)$, that is, p_i is the favorite alternative for voter i. Ordinal preferences are specified by a total order σ_i consistent with these dis-utilities (i.e., an alternative is ranked above another only if it has lower dis-utility). A preference profile σ specifies the ordinal preferences of all agents, and we denote $\sigma \in \rho(d)$ to mean that σ_i is consistent with the dis-utilities for every i. A deterministic social choice rule is a function f that maps a preference profile σ to an alternative $a \in M$. A randomized social choice rule maps a preference profile σ to a distribution over M.

2.1 Metric Implicit Utilitarian Model

We measure the quality of an alternative $a \in S$ by its *social cost*, given by $SC(a,d) = \frac{1}{n}\sum_{i \in N} d(i,a)$. Where d is obvious from context, we will simply write $SC(a)$. Let $a^* \in M$ be the minimizer of social cost. The Distortion [Procaccia and Rosenschein, 2006] measures the worst case approximation to the optimal social cost of a given mechanism, in expectation for randomized mechanisms.

Definition 1. *The **Distortion** of a social choice rule f is*

$$\textit{Distortion}(f) = \sup_{d,\, \sigma \in \rho(d)} \frac{\mathbb{E}_{f(\sigma)}[SC(a,d)]}{SC(a^*,d)}.$$

We assume that $M \cup N$ is a set of points in a metric space. Specifically, we assume the disutility function d is the distance function over this metric space. This assumption models social choice scenarios where there is an objective notion of the distance between alternatives. The metric assumption is common in the implicit utilitarian literature [Anshelevich and Postl, 2017; Goel *et al.*, 2017; Fain *et al.*, 2017; Gross *et al.*, 2017; Cheng *et al.*, 2017; Anshelevich *et al.*, 2018; Feldman *et al.*, 2016; Fain *et al.*, 2019], and we consider an example from participatory budgeting in Section 5 where the metric assumption is plausible.

2.2 Sampling and Higher Moments of Distortion

We consider mechanisms that implement a randomized social choice rule by first eliciting favorite alternatives from a random sample of voters and then uses only these alternatives for the rest of the mechanism. The size of this random sample is the *sample complexity* of our mechanism. We are interested in mechanisms with constant sample complexity with respect to n and m. A mechanism with sample complexity s only requires voters to rank at most s alternatives, so constant sample complexity implies that the number of alternatives voters must rank is constant with respect to n and m.

We consider two models that differ in how voters participate after we elicit these alternatives. In the *full participation model* of Section 3 we allow all voters to rank the alternatives from the first step and we aggregate these votes to output the winner. While this requires two distinct rounds, it is close to how real Participatory Budgeting processes work, where proposals are constructed by a subset of the population in the first stage, and these are put to vote in the second stage. In the *limited participation model* of Section 4, only the sample of voters from the first step vote over the alternatives. Thus, mechanisms in the limited participation model do not require a second distinct round involving different voters. It is worth noting that while the sample complexity of our results are lower in the full participation model, the total communication complexity is higher because all voters participate in the second round.

In order to capture the notion of risk inherent in a randomized social choice mechanism, we consider higher statistical moments of Distortion. In order to fairly compare the bounds for different moments, we normalize by the k^{th} root.

Definition 2. *The **normalized $\mathbf{k^{th}}$ moment of Distortion** of a social choice rule f is*

$$\textit{Distortion}^k(f) = \sup_{d,\, \sigma \in \rho(d)} \frac{\left(\mathbb{E}_{f(\sigma)}\left[(SC(a,d))^k\right]\right)^{1/k}}{SC(a^*,d)}.$$

Note that by Jensen's inequality, if a mechanism f has $Distortion^k(f) \leq c$ then $Distortion^{k'}(f) \leq c$ for all $k' \leq k$. By contrast, lower moments do *not* imply anything about higher moments of Distortion.

2.3 Relationship Between Higher Moments and High Probability Guarantees

Upper bounds on higher moments of Distortion immediately provide high probability guarantees for approximating the optimal social cost via Markov's inequality (see Corollaries 1 and 2). However, one can reasonably ask whether the high probability bounds we achieve in this way are "tight."

More precisely, suppose we want to approximate the optimal social cost with high probability: i.e., for constant $c > 1$, find an alternative a such that $SC(a,d) \leq c \cdot SC(a^*,d)$ with probability at least $1-\delta$. How many samples (favorite alternatives of random voters) are necessary as a function of c and δ? The example in Theorem 1 shows that one needs at least $\frac{\log(1/\delta)}{\log(c+1)}$ samples in the full participation model. On the other hand, Corollary 2 shows that our PRC mechanism needs just $\frac{\log(1/\delta)}{\log(c/11)}$ samples (for $c > 11$). So our results are tight with respect to the dependence on the probability term δ, but the factor of 11 in Corollary 1 is a consequence of the analysis for Theorem 2 and may be improvable.

3 Full Participation Model

In this section, we consider mechanisms that first elicit alternatives by sampling a number of voters and querying them for their most preferred alternatives and then apply a social choice rule on the elicited alternatives with all voters. We begin with the lower bound on the number of samples needed to bound the k^{th} moment of Distortion.

Theorem 1. *Any mechanism f with sample complexity less than k has $Distortion^k(f) = \Omega(n^{1/k})$.*

Proof. Consider a metric space with two outcomes A and B separated by distance 1. The fraction of voters located at A is $\alpha > 1/2$ and at B is $1-\alpha$. Note that the average (per-voter) social cost of OPT is $1-\alpha$. If $k-1$ voters are sampled,

with probability $(1-\alpha)^{k-1}$, all of them lie at B, in which case any voting mechanism using these samples is run on only outcome B. Therefore, the social cost in this case is α. The k^{th} moment of Distortion is therefore at least:

$$\left((1-\alpha)^{k-1}\left(\frac{\alpha^k}{(1-\alpha)^k}\right)\right)^{1/k} = \frac{\alpha}{(1-\alpha)^{1/k}}$$

Choosing $\alpha = 1 - c/n$ for constant c so that all but c voters lie at A, the above expression is $\Omega(n^{1/k})$. □

3.1 The $\mathbf{PRC_s}$ Mechanism

On the constructive side, we consider a family of mechanisms that achieve constant normalized k^{th} moment of Distortion using the minimum possible number of samples. We call this family Partially Random Copeland rules.

Definition 3. *The **Partially Random Copeland** rule parameterized by positive integer s, denoted $\mathbf{PRC_s}$, proceeds as follows. First sample s voters $\tilde{N}$ drawn independently and uniformly at random from N with replacement. All voters in $\tilde{N}$ are queried for their favorite alternative, and the union of all such alternatives is denoted $\tilde{M}$. Finally, PRC_s returns the winning alternative under the Copeland social choice rule with voters N and alternatives $\tilde{M}$.*

In the rest of this section, we will show the following. Intuitively, Theorem 2 asserts that every additional sample in the elicitation step of PRC provides a constant approximation to the next higher moment of Distortion.

Theorem 2. *For any $n \geq 3$ voters, $Distortion^k(PRC_k) \leq 11 + \frac{8}{n-2}$, which approaches 11 as $n \to \infty$.*

Proof. We first present a useful lemma bounding the k^{th} moment of the minimum of $i.i.d.$ random variables. We provide a proof in the full version of the paper.[1]

Lemma 1. *Let $X_1, X_2, \ldots, X_k$ be drawn $i.i.d.$ from distribution X and let $\mu = \mathbb{E}[X]$. Then,*

$$\left(\mathbb{E}\left[\min(X_1, X_2, \ldots, X_k)^k\right]\right)^{1/k} \leq \mu$$

We now proceed to prove Theorem 2. Let $a^* = \text{argmin}_{a \in M} SC(a)$ denote the social optimum. Let $\mu = SC(a^*) = \frac{1}{n}\sum_{i \in N} d(i, a^*)$. Suppose we sample a set S of voters. For $i \in S$, let $X_i = d(i, a^*)$. Note that $\mathbb{E}[X_i] = \mu$, and the X_i are $i, i.d.$ random variables.

Let $m = \text{argmin}_{i \in S} X_i$ be the voter closest to a^*, and let a_m denote their favorite alternative. Note $d(m, a_m) \leq X_m$.

Let $\alpha = 1 + \frac{1}{n-2}$. Consider a ball centered at a^* of radius $\rho = 2\alpha\mu$ denoted B. By Markov's inequality, we know that a strict majority, at least $\frac{n}{2} + 1$, of all voters lie within the ball B, since the average distance of a voter to a^* is μ.

Given S, suppose $\mathbf{PRC_k}$ chooses alternative W, and suppose $d(W, a^*) = \beta\rho$. We will show an upper bound on β using the random variable X_m. Since a Copeland winner must be a member of the uncovered set [Miller, 1977], either a majority of voters prefer W to a_m, or a majority of voters must prefer W to an alternative W' such that a majority of voters also prefer W' to a_m. The first case is easier: if a majority of voters prefer W to a_m, then there exists a voter $j \in B$ that prefers W to a_m. This implies that

$$\beta\rho = d(a^*, W) \leq d(a^*, j) + d(j, a_m) \leq 2\rho + d(a^*, a_m).$$

Recall that $X_m = d(m, a^*)$ and $d(m, a_m) \leq X_m$, so

$$\beta\rho \leq 2\rho + 2X_m.$$

The second case yields a worse bound, so we continue the analysis in that case without loss of generality. Let $d(W', a^*) = \beta'\rho$. Since a majority of voters prefer W to W', there is at least one voter $j \in B$ that prefers W to W', that is, $d(j, W) < d(j, W')$. By triangle inequality,

$$d(j, W) \geq d(a^*, W) - d(j, a^*) \geq (\beta - 1)\rho$$
$$d(j, W') \leq d(a^*, W') + d(j, a^*) \leq (\beta' + 1)\rho$$

where the rightmost inequalities follow from the fact that $j \in B \implies d(j, a^*) \leq \rho$. Combining the above inequalities and assuming $\beta > 1$, we have $\beta \leq \beta' + 2$. Similarly, if a majority of voters prefer W' to a_m, there exists some $l \in B$ such that $d(l, W') < d(l, a_m)$. Again, by triangle inequality:

$$d(l, W') \geq d(a^*, W') - d(l, a^*) \geq (\beta' - 1)\rho$$
$$d(l, a_m) \leq d(l, a^*) + d(a^*, m) + d(m, a_m) \leq \rho + 2X_m$$

where we used that $i \in B$ and $d(m, X_m) \leq d(a^*, m) = X_m$. Combining the above inequalities, we have $\beta' < 2 + \frac{2X_m}{\rho}$. Since $\beta \leq \beta' + 2$, we have: $\beta < 4 + \frac{2X_m}{\rho}$

Thus, we know that for W to win Copeland,

$$d(W, a^*) = \beta\rho \leq 4\rho + 2X_m = 8\left(1 + \frac{1}{n-2}\right)\mu + 2X_m$$

By triangle inequality, and using $SC(a^*) = \mu$, we have:

$$SC(W) \leq d(W, a^*) + SC(a^*) \leq \left(9 + \frac{8}{n-2}\right)\mu + 2X_m$$

Setting $\gamma = \left(9 + \frac{8}{n-2}\right)\mu$, we have:

$$\mathbb{E}[SC(W)^k] \leq \mathbb{E}[(\gamma + 2X_m)^k] = \sum_{r=0}^{k}\binom{k}{r}\gamma^{k-r}\,\mathbb{E}[X_m^r]2^r$$

Since X_m is the minimum of k i.i.d. random variables with mean μ, applying Lemma 1, we have $\mathbb{E}[X_m^k] \leq \mu^k$. Applying Jensen's inequality, for $r \leq k$, we have

$$\mathbb{E}[X_m^r] = \mathbb{E}[(X_m^k)^{r/k}] \leq \mathbb{E}[X_m^k]^{r/k} = (\mu^k)^{r/k} = \mu^r.$$

Therefore, we have

$$\mathbb{E}[SC(W)^k] \leq \sum_{r=0}^{k}\binom{k}{r}\gamma^{k-r}(2\mu)^r = \left(11 + \frac{8}{n-2}\right)^k \mu^k$$

Therefore , we have $Distortion^k(PRC_k) \leq 11 + \frac{8}{n-2}$, completing the proof of Theorem 2.

□

As a simple consequence, using Markov's inequality, this yields a high probability bound on Distortion. In particular, every additional sample in the elicitation step of PRC provides a geometric improvement in the high probability bound.

Corollary 1. *As $n \to \infty$ and $c > 11$, the probability that PRC_k outputs an alternative with social cost more than c times that of the social optimum is at most $(11/c)^k$.*

[1] Available at https://arxiv.org/abs/2004.13153

4 Limited Participation Model

In this section, we consider mechanisms that sample some number of voters, query the voters for their most preferred alternatives, and then hold an election on just the sample of voters. We first show that limiting participation in this way necessarily increases the sample complexity.

Theorem 3. *Any anonymous limited participation randomized mechanism with sample complexity less than $2k-1$ has $Distortion^k(f) = \Omega(n^{1/k})$.*

Proof. Consider the same instance as Theorem 1. Suppose we sample $2k-2$ voters. Then the probability that we sample an equal number of voters located at A and B is

$$\binom{2k-2}{k-1}\alpha^{k-1}(1-\alpha)^{k-1} \geq (2\alpha)^{k-1}(1-\alpha)^{k-1} \geq (1-\alpha)^{k-1}$$

where we have assumed $\alpha \geq 1/2$. In this event, since there is no majority of voters in the sample that prefer either alternative, we assume that any anonymous mechanism outputs B with probability at least $1/2$, so that the social cost is at least $\frac{\alpha}{2}$. Therefore, the k^{th} moment of distortion is at least:

$$\left((1-\alpha)^{k-1}\left(\frac{(\alpha/2)^k}{(1-\alpha)^k}\right)\right)^{1/k} = \frac{\alpha}{2(1-\alpha)^{1/k}}$$

Choosing $\alpha = 1 - c/n$ for constant c so that all but c voters lie at A, the above expression is $\Omega(n^{1/k})$. □

4.1 The $\mathbf{FRC_s}$ Mechanism

Complementing the above impossibility, we show that sample complexity of $2k-1$ is also sufficient to achieve constant k^{th} moment of Distortion. In particular, we define another family of social choice rules called Fully Random Copeland.

Definition 4. *The **Fully Random Copeland** rule parameterized by positive integer s, denoted $\mathbf{FRC_s}$ proceeds as follows. It first samples s voters $\tilde{N}$ drawn independently and uniformly at random from N with replacement. All voters in $\tilde{N}$ are queried for their favorite alternative, and the union of all such alternatives is denoted $\tilde{M}$. Finally, FRC_s returns the winning alternative under the Copeland social choice rule with voters $\tilde{N}$ and alternatives $\tilde{M}$.*

We now show Theorem 4, which says that every additional two voters participating in FRC provide a constant approximation to the next higher moment of Distortion.

Theorem 4. $Distortion^k(FRC_{2k-1}) \leq 17$.

Proof. As in Section 3, we first present a result on bounding the k^{th} moment of a function of $i.i.d.$ random variables; this time the function is the median instead of the minimum. We provide a proof in the full version of the paper.[2]

Lemma 2. *Let $X_1, X_2, \ldots, X_{2k-1}$ be drawn $i.i.d.$ from distribution X and let $\mu = \mathbb{E}[X]$. Let Y denote the median of $X_1, X_2, \ldots, X_{2k-1}$. Then, $\left(\mathbb{E}[Y^k]\right)^{1/k} \leq 4\mu$.*

[2] Available at https://arxiv.org/abs/2004.13153

We will also need the following straightforward property of the Copeland Rule.

Lemma 3. *Suppose there are $2k-1$ alternatives and voters. Construct a tournament graph on the alternatives with a directed edge from alternative S to alternative T if at least k voters strictly prefer S to T. Then the Copeland rule always picks an alternative W with in-degree strictly less than k.*

We now proceed to prove Theorem 4. As in the proof of Theorem 2, let a^* denote the optimal alternative, and let $SC(a^*) = \mu = \frac{1}{n}\sum_{i\in N} d(i, a^*)$. Suppose we sample a subset of voters, S of size $2k-1$ For $i \in S$, let $X_i = d(i, a^*)$. Order these voters so that $X_1 \leq X_2 \leq \cdots \leq X_{2k-1}$ and let m be the voter that corresponds to the median of this sequence. Let $Y = d(m, a^*)$. Note from Lemma 2 that $\mathbb{E}[Y^k] \leq (4\mu)^k$.

Suppose the Copeland rule chooses an alternative W, and suppose $d(W, a^*) = \alpha Y$. We will find an upper bound for α. Consider the ball centered around a^* with radius Y; call this B. By definition, at least k agents in S lie within B. Note that for any $j \in B \cap S$, $d(j, a_j) \leq d(j, a^*) \leq Y$. Therefore, for $j, l \in B \cap S$, we have

$$d(j, a_l) \leq d(j, a^*) + d(l, a^*) + d(l, a_l) \leq Y + Y + Y = 3Y$$

Now, for $j \in B \cap S$, we have

$$d(j, W) \geq d(a^*, W) - d(j, a^*) \geq (\alpha - 1)Y$$

If $\alpha > 4$, then combining the above two observations, we have that for all $j, l \in B$, we have $d(j, a_l) \leq 3Y < d(j, W)$. This means that the set of at least k voters in $B \cap S$ strictly prefer all of the favorite alternatives $\{a_l, l \in B \cap S\}$ to W. From Lemma 3, this means that W cannot be the Copeland winner. Thus, for W to win in the Copeland rule, we must have $\alpha \leq 4$ so $d(W, a^*) \leq 4Y$. By triangle inequality,

$$SC(W) \leq SC(a^*) + d(a^*, W) \leq \mu + 4Y$$

Using Jensen's inequality in a fashion similar to the proof of Theorem 2, and using $\mathbb{E}[Y^k] \leq (4\mu)^k$ we have:

$$\mathbb{E}[SC(W)^k] \leq \mathbb{E}[(\mu + 4Y)^k] \leq \sum_{r=0}^{k}\binom{k}{r}\mu^k 16^r = 17^k\mu^k$$

so we have that $Distortion^k(FRC_{2k-1}) \leq 17$. This completes the proof of Theorem 4.

□

Again, as a consequence of Markov's inequality, we have the following high probability bound.

Corollary 2. *For $c \geq 17$, the probability that FRC_{2k-1} outputs an alternative with social cost more than c times that of the social optimum is at most $(17/c)^k$.*

5 Empirical Simulation

In this section, we augment our theoretical worst case analysis with a qualitative empirical demonstration of the concentration achieved by the PRC and FRC mechanisms on real world data. We use data from the Participatory Budgeting project; see [Goel *et al.*, 2015]. In this domain, there are a

number of public projects (such as new sidewalks, park renovations, etc.). Each project has a monetary cost, and we want to select a set of projects subject to not exceeding a total budget. In participatory budgeting, local community members vote directly over their preferred projects, and these votes are aggregated to decide which projects to fund.

We consider knapsack voting data [Goel *et al.*, 2015] where each voter reports the set of projects they most prefer subject to the total budget constraint. This makes knapsack voter data particularly useful for us: voters select their single favote alternative out of a very large space, the power set of projects. Because we also have information about the latent combinatorial space (specific projects selected and their costs), we can impose simplistic but natural notions of distance to allow us to simulate our mechanisms and study their performance with respect to the imposed distance.

Simulation. It is important to note that this is a simulation; actually running our mechanisms does not require specifying a notion of distance, and we do not know how these voters would have responded to ordinal queries in reality. We are treating an entire budget allocation as a single outcome and imputing preferences of voters over these outcomes. This reduces the problem to single winner election over a large space of alternatives in keeping with the theoretical model of this paper. Therefore, natural baseline mechanisms are single winner rules with small sample complexity, particularly Random Dictatorship which is the best-known mechanism with respect to the first moment of Distortion. Other mechanisms for participatory budgeting are tailored to specific models of voter preferences over the combinatorial space of projects, and do not, in general, provide constant Distortion guarantees for arbitrary metrics.

Setup. We consider two simple notions of distance: budget distance and Jaccard distance. Suppose there are p public projects numbered $1, \ldots, p$ with costs $c_1, \ldots, c_p$, and there is a total budget of B. A *feasible budget* is a set of projects P such that $\sum_{i \in P} c_p \leq B$. The *budget distance* between budgets P and Q is $1 - \frac{1}{B} \sum_{i \in P \cap Q} c_i$. The *Jaccard distance* between P and Q is $1 - \frac{|P \cap Q|}{|P \cup Q|}$. The social cost of a given budget is the average distance to the proposed budgets of the voters. We use knapsack voting data from the Participatory Budgeting election held in Cambridge, MA, USA in 2015. There were 945 voters, 23 projects (implying $2^{23} > 8$ million possible budgets), and a total budget constraint of $\$600,000$.

Results. In Figure 1, we present the box plots of the distributions of social cost of PRC and FRC alongside Random Dictatorship (RD) when simulating using budget distance and Jaccard distance respectively. The RD mechanism samples a single most preferred alternative uniformly at random and has Distortion at most 3 [Anshelevich and Postl, 2017], which is asymptotically the best known bound for any randomized social choice mechanism for arbitrary metrics. The examples qualitatively verify that the PRC and FRC mechanisms do provide substantial concentration in terms of the approximation to the optimal social cost. Furthermore, in practice we observe better average performance of PRC and FRC over that of RD, despite RD's theoretical optimality with respect to the first moment of Distortion. The results also show that FRC requires more samples to achieve similar performance as PRC. To summarize, even on real datasets, just a few additional samples provide substantially improved concentration.

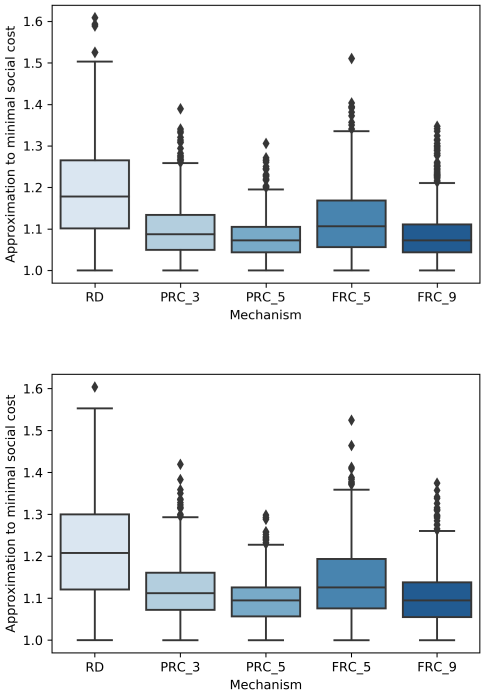

Figure 1: Distribution of approximation to optimal social cost for 1,000 runs of each mechanism on Cambridge 2015 knapsack voting data using budget distance (top) and Jaccard distance (bottom).

6 Future Directions

There are several avenues of future research. Our mechanisms first sample some alternatives and then put them to vote. Is there a one-shot mechanism that can bound higher moments of Distortion while only eliciting a constant (with respect to the number of alternatives) amount of information from each voter? Our intuition is that this should be impossible. Also, though our sample complexity bounds are tight, the exact constant in the Distortion bounds can likely be improved. This improvement may be nontrivial: We do not use the Distortion of Copeland as a black box, so results such as [Munagala and Wang, 2019] do not directly improve our bounds. As in [Mandal *et al.*, 2019], it would be interesting to analyze whether sample complexity can be decreased if voters can express limited amounts of information about cardinal preferences. In a related vein, could methods that make voters interact like [Fain *et al.*, 2017] help reduce the sample complexity of the process?

Acknowledgments

Kamesh Munagala was supported by NSF grants CCF-1408784, CCF-1637397, and IIS-1447554, ONR award N00014-19-1-2268, and awards from Adobe and Facebook.

References

[Anshelevich and Postl, 2017] Elliot Anshelevich and John Postl. Randomized social choice functions under metric preferences. *Journal of Artificial Intelligence Research*, 58(1):797–827, January 2017.

[Anshelevich *et al.*, 2018] Elliot Anshelevich, Onkar Bhardwaj, Edith Elkind, John Postl, and Piotr Skowron. Approximating optimal social choice under metric preferences. *Artificial Intelligence*, 264:27 – 51, 2018.

[Boutilier *et al.*, 2015] Craig Boutilier, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D Procaccia, and Or Sheffet. Optimal social choice functions: A utilitarian view. *Artificial Intelligence*, 227:190–213, 2015.

[Bouveret *et al.*, 2017] Sylvain Bouveret, Yann Chevaleyre, François Durand, and Jérôme Lang. Voting by sequential elimination with few voters. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 128–134, 2017.

[Brandt *et al.*, 2016] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia. *Handbook of Computational Social Choice*. Cambridge University Press, New York, New York, 2016.

[Caragiannis and Procaccia, 2010] Ioannis Caragiannis and Ariel D. Procaccia. Voting almost maximizes social welfare despite limited communication. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, AAAI'10, pages 743–748, 2010.

[Cheng *et al.*, 2017] Yu Cheng, Shaddin Dughmi, and David Kempe. Of the people: Voting is more effective with representative candidates. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, EC '17, 2017.

[Conitzer and Sandholm, 2005] Vincent Conitzer and Tuomas Sandholm. Communication complexity of common voting rules. In *Proceedings of the 6th ACM Conference on Electronic Commerce*, EC '05, pages 78–87, 2005.

[Dey and Bhattacharyya, 2015] Palash Dey and Arnab Bhattacharyya. Sample complexity for winner prediction in elections. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS'15, pages 1421–1430, Richland, SC, 2015. International Foundation for Autonomous Agents and Multiagent Systems.

[Dey and Narahari, 2015] Palash Dey and Y. Narahari. Estimating the margin of victory of an election using sampling. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 1120–1126, 2015.

[Fain *et al.*, 2017] Brandon Fain, Ashish Goel, Kamesh Munagala, and Sukolsak Sakshuwong. Sequential deliberation for social choice. In *Proceedings of the 13th International Conference on Web and Internet Economics*, WINE'17, pages 177–190, 2017.

[Fain *et al.*, 2019] Brandon Fain, Ashish Goel, Kamesh Munagala, and Nina Prabhu. Random dictators with a random referee: Constant sample complexity mechanisms for social choice. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, AAAI'19, 2019.

[Feldman *et al.*, 2016] Michal Feldman, Amos Fiat, and Iddan Golomb. On voting and facility location. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, EC '16, pages 269–286, 2016.

[Goel *et al.*, 2015] Ashish Goel, Anilesh K Krishnaswamy, Sukolsak Sakshuwong, and Tanja Aitamurto. Knapsack voting. *Collective Intelligence*, 2015.

[Goel *et al.*, 2017] Ashish Goel, Anilesh K. Krishnaswamy, and Kamesh Munagala. Metric distortion of social choice rules: Lower bounds and fairness properties. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, EC '17, pages 287–304, New York, NY, USA, 2017. ACM.

[Gross *et al.*, 2017] Stephen Gross, Elliot Anshelevich, and Lirong Xia. Vote until two of you agree: Mechanisms with small distortion and sample complexity. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, 2017.

[Mandal *et al.*, 2019] Debmalya Mandal, Ariel D Procaccia, Nisarg Shah, and David Woodruff. Efficient and thrifty voting by any means necessary. In *Thirty-third Conference on Neural Information Processing Systems*, pages 7178–7189, 2019.

[Miller, 1977] Nicholas R. Miller. Graph-theoretical approaches to the theory of voting. *American Journal of Political Science*, 21(4):769–803, 1977.

[Munagala and Wang, 2019] Kamesh Munagala and Kangning Wang. Improved metric distortion for deterministic social choice rules. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC'19, pages 245–262, 2019.

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