Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Voting Rules As Error-Correcting Codes

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

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

Yair Zick
Carnegie Mellon University
yairzick@cmu.edu

Abstract

We present the first model of optimal voting under adversarial noise. From this viewpoint, voting rules are seen as error-correcting codes: their goal is to correct errors in the input rankings and recover a ranking that is close to the ground truth. We derive worst-case bounds on the relation between the average accuracy of the input votes, and the accuracy of the output ranking. Empirical results from real data show that our approach produces significantly more accurate rankings than alternative approaches.

Introduction

Social choice theory develops and analyzes methods for aggregating the opinions of individuals into a collective decision. The prevalent approach is motivated by situations in which opinions are *subjective*, such as political elections, and focuses on the design of voting rules that satisfy normative properties (Arrow 1951).

An alternative approach, which was proposed by the marquis de Condorcet in the 18th Century, had confounded scholars for centuries (due to Condorcet's ambiguous writing) until it was finally elucidated by Young (1988). The underlying assumption is that the alternatives can be *objectively* compared according to their true quality. In particular, it is typically assumed that there is a ground truth *ranking* of the alternatives. Votes can be seen as noisy estimates of the ground truth, drawn from a specific *noise model*. For example, Condorcet proposed a noise model where — roughly speaking — each voter (hereinafter, *agent*) compares every pair of alternatives, and orders them correctly (according to the ground truth) with probability $p > 1/2$; today an equivalent model is attributed to Mallows (1957). Here, it is natural to employ a voting rule that always returns a ranking that is *most likely* to coincide with the ground truth, that is, the voting rule should be a *maximum likelihood estimator (MLE)*.

Although Condorcet could have hardly foreseen this, his MLE approach is eminently applicable to crowdsourcing and human computation systems, which often employ voting to aggregate noisy estimates; EteRNA (Lee et al. 2014) is a wonderful example, as explained by Procaccia et al. (2012). Consequently, the study of voting rules as MLEs has been gaining steam in the last decade (Conitzer and Sandholm 2005; Conitzer et al. 2009; Elkind et al. 2010; Xia et al. 2010; Xia and Conitzer 2011; Lu and Boutilier 2011; Procaccia et al. 2012; Azari Soufiani et al. 2012; 2013; 2014; Mao et al. 2013; Caragiannis et al. 2013; 2014).

Despite its conceptual appeal, a major shortcoming of the MLE approach is that the MLE voting rule is specific to a noise model, and that noise model — even if it exists for a given setting — may be difficult to pin down (Mao, Procaccia, and Chen 2013). Caragiannis et al. (2013; 2014) have addressed this problem by relaxing the MLE constraint: they only ask that the probability of the voting rule returning the ground truth go to one as the number of votes goes to infinity. This allows them to design voting rules that elicit the ground truth in a wide range of noise models; however, they may potentially require an infinite amount of information.

Our approach. In this paper, we propose a fundamentally different approach to aggregating noisy votes. We assume the noise to be *adversarial* instead of probabilistic, and wish to design voting rules that do well under worst-case assumptions. From this viewpoint, our approach is closely related to the extensive literature on *error-correcting codes*. One can think of the votes as a *repetition code*: each vote is a transmitted noisy version of a "message" (the ground truth). How many errors can be corrected using this "code"?

In more detail, let d be a distance metric on the space of rankings. As an example, the well-known *Kendall tau (KT)* distance between two rankings measures the number of pairs of alternatives on which the two rankings disagree. Suppose that we receive n votes over the set of alternatives $\{a, b, c, d\}$, for an even n, and we know that the average KT distance between the votes and the ground truth is at most $1/2$. Can we always recover the ground truth? No: in the worst-case, exactly $n/2$ agents swap the two highest-ranked alternatives and the rest report the ground truth. In this case, we observe two distinct rankings (each $n/2$ times) that only disagree on the order of the top two alternatives. Both rankings have an average distance of $1/2$ from the input votes, making it impossible to determine which of them is the ground truth.

Let us, therefore, cast a larger net. Inspired by *list decoding* of error-correcting codes (see, e.g., Guruswami 2005), our main research question is:

Fix a distance metric d. Suppose that we are given n noisy rankings, and that the average distance between these rankings and the ground truth is at most t. We wish to recover a ranking that is guaranteed to be at distance at most k from the ground truth. How small can k be, as a function of n and t?

Our results. We observe that for any metric d, one can always recover a ranking that is at distance at most $2t$ from the ground truth, i.e., $k \leq 2t$. Under an extremely mild assumption on the distance metric, we complement this result by proving a lower bound of (roughly) $k \geq t$. Next, we consider the four most popular distance metrics used in the social choice literature, and prove a tight lower bound of (roughly) $k \geq 2t$ for each metric. This lower bound is our main theoretical result; the construction makes unexpected use of Fermat's Polygonal Number Theorem.

The worst-case optimal voting rule in our framework is defined with respect to a *known* upper bound t on the average distance between the given rankings and the ground truth. However, we show that the voting rule which returns the ranking minimizing the total distance from the given rankings — which has strong theoretical support in the literature — serves as an approximation to our worst-case optimal rule, irrespective of the value of t. We leverage this observation to provide theoretical performance guarantees for our rule in cases where the error bound t given to the rule is an underestimate or overestimate of the tightest upper bound.

Finally, we test our worst-case optimal voting rules against many well-established voting rules, on two real-word datasets (Mao, Procaccia, and Chen 2013), and show that the worst-case optimal rules exhibit superior performance as long as the given error bound t is a reasonable overestimate of the tightest upper bound.

Related work. Our work is related to the extensive literature on error-correcting codes that use permutations (see, e.g., Barg and Mazumdar 2010, and the references therein), but differs in one crucial aspect. In designing error-correcting codes, the focus is on two choices: i) the *codewords*, a subset of rankings which represent the "possible ground truths", and ii) the *code*, which converts every codeword into the message to be sent. These choices are optimized to achieve the best tradeoff between the number of errors corrected and the rate of the code (efficiency), while allowing unique identification of the ground truth. In contrast, our setting has fixed choices: i) every ranking is a possible ground truth, and ii) in coding theory terms, our setting constrains us to the repetition code. Both restrictions (inevitable in our setting) lead to significant inefficiencies, as well as the impossibility of unique identification of the ground truth (as illustrated in the introduction). Our research question is reminiscent of coding theory settings where a bound on adversarial noise is given, and a code is chosen with the bound on the noise as an input to maximize efficiency (see, e.g., Haeupler 2014).

List decoding (see, e.g., Guruswami 2005) relaxes classic error correction by guaranteeing that the number of possible messages does not exceed a small quota; then, the decoder simply lists all possible messages. The motivation is that one can simply scan the list and find the correct message, as all other messages on the list are likely to be gibberish. In the voting context, one cannot simply disregard some potential ground truths as nonsensical; we therefore select a ranking that is close to every possible ground truth.

A bit further afield, Procaccia et al. (2007) study a probabilistic noisy voting setting, and quantify the *robustness* of voting rules to random errors. Their results focus on the probability that the outcome would change, under a random transposition of two adjacent alternatives in a single vote from a submitted profile, in the *worst-case* over profiles. Their work is different from ours in many ways, but perhaps most importantly, they are interested in how frequently common voting rules make mistakes, whereas we are interested in the guarantees of *optimal* voting rules that *avoid* mistakes.

Preliminaries

Let A be the set of alternatives, and $|A| = m$. Let $\mathcal{L}(A)$ be the set of rankings over A. A vote σ is a ranking in $\mathcal{L}(A)$, and a profile $\pi \in \mathcal{L}(A)^n$ is a collection of n rankings. A *voting rule* $f : \mathcal{L}(A)^n \to \mathcal{L}(A)$ maps every profile to a ranking.[1]

We assume that there exists an underlying ground truth ranking $\sigma^* \in \mathcal{L}(A)$ of the alternatives, and the votes are noisy estimates of σ^*. We use a distance metric d over $\mathcal{L}(A)$ to measure errors; the error of a vote σ with respect to σ^* is $d(\sigma, \sigma^*)$, and the average error of a profile π with respect to σ^* is $d(\pi, \sigma^*) = (1/n) \cdot \sum_{\sigma \in \pi} d(\sigma, \sigma^*)$. We consider four popular distance metrics over rankings in this paper.

- *The Kendall tau (KT) distance*, denoted d_{KT}, measures the number of pairs of alternatives over which two rankings disagree. Equivalently, it is also the minimum number of swaps of adjacent alternatives required to convert one ranking into another.
- *The (Spearman's) Footrule (FR) distance*, denoted d_{FR}, measures the total displacement of all alternatives between two rankings, i.e., the sum of the absolute differences between their positions in two rankings.
- *The Maximum Displacement (MD) distance*, denoted d_{MD}, measures the maximum of the displacements of all alternatives between two rankings.
- *The Cayley (CY) distance*, denoted d_{CY}, measures the minimum number of swaps (not necessarily of adjacent alternatives) required to convert one ranking into another.

All four metrics described above are *neutral*: A distance metric is called neutral if the distance between two rankings is independent of the labels of the alternatives; in other words, choosing a relabeling of the alternatives and applying it to two rankings keeps the distance between them invariant.

Worst-Case Optimal Rules

Suppose we are given a profile π of n noisy rankings that are estimates of an underlying true ranking σ^*. In the absence of any additional information, any ranking could potentially be the true ranking. However, because essentially all crowdsourcing methods draw their power from the often-observed

[1] They are known as *social welfare functions*, which differ from *social choice functions* that choose a single winning alternative.

fact that individual opinions are accurate on average, we can plausibly assume that while some agents may make many mistakes, the average error is fairly small. An upper bound on the average error may be inferred by observing the collected votes, or from historical data (but see the next section for the case where this bound is inaccurate).

Formally, suppose we are guaranteed that the average distance between the votes in π and the ground truth σ^* is *at most* t according to a metric d, i.e., $d(\pi, \sigma^*) \le t$. With this guarantee, the set of possible ground truths is given by the "ball" of radius t around π.

$$\mathcal{B}_t^d(\pi) = \{\sigma \in \mathcal{L}(A) \mid d(\pi, \sigma) \le t\}.$$

Note that we have $\sigma^* \in \mathcal{B}_t^d(\pi)$ given our assumption; hence, $\mathcal{B}_t^d(\pi) \neq \emptyset$. We wish to find a ranking that is as close to the ground truth as possible. Since our approach is worst case in nature, our goal is to find the ranking that minimizes the maximum distance from the possible ground truths in $\mathcal{B}_t^d(\pi)$. For a set of rankings $S \subseteq \mathcal{L}(A)$, let its *minimax ranking*, denoted $\textsc{MiniMax}^d(S)$, be defined as follows.[2]

$$\textsc{MiniMax}^d(S) = \underset{\sigma \in \mathcal{L}(A)}{\arg\min} \max_{\sigma' \in S} d(\sigma, \sigma').$$

Let the *minimax distance* of S, denoted $k^d(S)$, be the maximum distance of $\textsc{MiniMax}^d(S)$ from the rankings in S according to d. Thus, given a profile π and the guarantee that $d(\pi, \sigma^*) \le t$, the worst-case optimal voting rule OPT^d returns the minimax ranking of the set of possible ground truths $\mathcal{B}_t^d(\pi)$. That is, for all profiles $\pi \in \mathcal{L}(A)^n$ and $t > 0$,

$$\mathrm{OPT}^d(t, \pi) = \textsc{MiniMax}^d\left(\mathcal{B}_t^d(\pi)\right).$$

Furthermore, the output ranking is guaranteed to be at distance at most $k^d(\mathcal{B}_t^d(\pi))$ from the ground truth. We overload notation, and denote $k^d(t, \pi) = k^d(\mathcal{B}_t^d(\pi))$, and

$$k^d(t) = \max_{\pi \in \mathcal{L}(A)^n} k^d(t, \pi).$$

While k^d is explicitly a function of t, it is also implicitly a function of n. Hereinafter, we omit the superscript d whenever the metric is clear from context.

Let us illustrate our terminology with a simple example.

Example 1. Let $A = \{a, b, c\}$. We are given profile π consisting of 5 votes: $\pi = \{2 \times (a \succ b \succ c),\ a \succ c \succ b,\ b \succ a \succ c,\ c \succ a \succ b\}$.

The maximum distances between rankings in $\mathcal{L}(A)$ allowed by d_{KT}, d_{FR}, d_{MD}, and d_{CY} are 3, 4, 2, and 2, respectively; let us assume that the average error limit is half the maximum distance for all four metrics.[3]

Consider the Kendall tau distance with $t = 1.5$. The average distances of all 6 rankings from π are given below.

$d_{KT}(\pi, a \succ b \succ c) = 0.8$	$d_{KT}(\pi, a \succ c \succ b) = 1.0$
$d_{KT}(\pi, b \succ a \succ c) = 1.4$	$d_{KT}(\pi, b \succ c \succ a) = 2.0$
$d_{KT}(\pi, c \succ a \succ b) = 1.6$	$d_{KT}(\pi, c \succ b \succ a) = 2.2$

[2]We use $\textsc{MiniMax}^d(S)$ to denote a single ranking. Ties among multiple minimizers can be broken arbitrarily; our results are independent of the tie-breaking scheme.

[3]Scaling by the maximum distance is not a good way of comparing distance metrics; we do so for the sake of illustration only.

Thus, the set of possible ground truths is $\mathcal{B}_{1.5}^{d_{KT}}(\pi) = \{a \succ b \succ c,\ a \succ c \succ b,\ b \succ a \succ c\}$. This set has a unique minimax ranking $\mathrm{OPT}^{d_{KT}}(1.5, \pi) = a \succ b \succ c$, which gives $k^{d_{KT}}(1.5, \pi) = 1$. Table 1 lists the sets of possible ground truths and their minimax rankings[4] under different distance metrics.

Voting Rule	Possible Ground Truths $\mathcal{B}_t^d(\pi)$	Output Ranking
$\mathrm{OPT}^{d_{KT}}(1.5, \pi)$, $\mathrm{OPT}^{d_{CY}}(1, \pi)$	$\{a \succ b \succ c,\ a \succ c \succ b,\ b \succ a \succ c\}$	$a \succ b \succ c$
$\mathrm{OPT}^{d_{FR}}(2, \pi)$, $\mathrm{OPT}^{d_{MD}}(1, \pi)$	$\{a \succ b \succ c,\ a \succ c \succ b\}$	$\{a \succ b \succ c,\ a \succ c \succ b\}$

Table 1: Application of the optimal voting rules on π.

Note that even with identical (scaled) error bounds, different distance metrics lead to different sets of possible ground truths as well as different optimal rankings. This demonstrates that the choice of the distance metric is significant.

Upper Bound

Given a distance metric d, a profile π, and that $d(\pi, \sigma^*) \le t$, we can bound $k(t, \pi)$ using the *diameter* of the set of possible ground truths $\mathcal{B}_t(\pi)$. For a set of rankings $S \subseteq \mathcal{L}(A)$, denote its diameter by $\mathcal{D}(S) = \max_{\sigma, \sigma' \in S} d(\sigma, \sigma')$.

Lemma 1. $\frac{1}{2} \cdot \mathcal{D}(\mathcal{B}_t(\pi)) \le k(t, \pi) \le \mathcal{D}(\mathcal{B}_t(\pi)) \le 2t.$

Proof. Let $\widehat{\sigma} = \textsc{MiniMax}(\mathcal{B}_t(\pi))$. For rankings $\sigma, \sigma' \in \mathcal{B}_t(\pi)$, we have $d(\sigma, \widehat{\sigma}), d(\sigma', \widehat{\sigma}) \le k(t, \pi)$ by definition of $\widehat{\sigma}$. By the triangle inequality, $d(\sigma, \sigma') \le 2k(t, \pi)$ for all $\sigma, \sigma' \in \mathcal{B}_t(\pi)$. Thus, $\mathcal{D}(\mathcal{B}_t(\pi)) \le 2k(t, \pi)$.

Next, the maximum distance of $\sigma \in \mathcal{B}_t(\pi)$ from all rankings in $\mathcal{B}_t(\pi)$ is at most $\mathcal{D}(\mathcal{B}_t(\pi))$. Hence, the minimax distance $k(t, \pi) = k(\mathcal{B}_t(\pi))$ cannot be greater than $\mathcal{D}(\mathcal{B}_t(\pi))$.

Finally, let $\pi = \{\sigma_1, \ldots, \sigma_n\}$. For rankings $\sigma, \sigma' \in \mathcal{B}_t(\pi)$, the triangle inequality implies $d(\sigma, \sigma') \le d(\sigma, \sigma_i) + d(\sigma_i, \sigma')$ for every $i \in \{1, \ldots, n\}$. Averaging over these inequalities, we get $d(\sigma, \sigma') \le t + t = 2t$, for all $\sigma, \sigma' \in \mathcal{B}_t(\pi)$. Thus, we have $\mathcal{D}(\mathcal{B}_t(\pi)) \le 2t$, as required. ■

Lemma 1 implies that $k(t) = \max_{\pi \in \mathcal{L}(A)^n} k(t, \pi) \le 2t$ for all distance metrics and $t > 0$. In words:

Theorem 1. *Given n noisy rankings at an average distance of at most t from an unknown true ranking σ^* according to a distance metric d, it is always possible to find a ranking at distance at most $2t$ from σ^* according to d.*

Importantly, the bound of Theorem 1 is independent of the number of votes n. Most statistical models of social choice restrict profiles in two ways: i) the average error should be low because the probability of generating high-error votes is typically low, and ii) the errors should be distributed almost evenly (in different directions from the ground truth),

[4]Multiple rankings indicate a tie that can be broken arbitrarily.

which is why aggregating the votes works well. These assumptions are mainly helpful when n is large, that is, performance may be poor for small n (see, e.g., Caragiannis et al. 2013). In contrast, our model restricts profiles only by making the first assumption (explicitly), allowing voting rules to perform well as long as the votes are accurate on average, independently of the number of votes n.

We also remark that Theorem 1 admits a simple proof, but the bound is nontrivial: while the average error of the profile is at most t (hence, the profile contains a ranking with error at most t), it is generally impossible to pinpoint a single ranking within the profile that has low error (say, at most $2t$) with respect to the ground truth in the worst-case (i.e., with respect to every possible ground truth in $\mathcal{B}_t(\pi)$).

Lower Bounds

The upper bound of $2t$ (Theorem 1) is intuitively loose — we cannot expect it to be tight for every distance metric. However, we can prove a generic lower bound of (roughly speaking) t for a wide class of distance metrics.

Formally, let $d^{\downarrow}(r)$ denote the greatest feasible distance under distance metric d that is less than or equal to r. $k(t)$ is the minimax distance under some profile, and hence must be a feasible distance under d. Thus, we prove a lower bound of $d^{\downarrow}(t)$, which is the best possible bound up to t.

Theorem 2. *For a neutral distance metric d, $k(t) \geq d^{\downarrow}(t)$.*

Proof. For a ranking $\sigma \in \mathcal{L}(A)$ and $r \geq 0$, let $\mathcal{B}_r(\sigma)$ denote the set of rankings at distance at most r from σ. Neutrality of the distance metric d implies $|\mathcal{B}_r(\sigma)| = |\mathcal{B}_r(\sigma')|$ for all $\sigma, \sigma' \in \mathcal{L}(A)$ and $r \geq 0$. In particular, $d^{\downarrow}(t)$ being a feasible distance under d implies that for every $\sigma \in \mathcal{L}(A)$, there exists some ranking at distance exactly $d^{\downarrow}(t)$ from σ.

Fix $\sigma \in \mathcal{L}(A)$. Consider the profile π consisting of n instances of σ. It holds that $\mathcal{B}_t(\pi) = \mathcal{B}_t(\sigma)$. We want to show that the minimax distance $k(\mathcal{B}_t(\sigma)) \geq d^{\downarrow}(t)$. Suppose for contradiction that there exists some $\sigma' \in \mathcal{L}(A)$ such that all rankings in $\mathcal{B}_t(\sigma)$ are at distance at most t' from σ', i.e., $\mathcal{B}_t(\sigma) \subseteq \mathcal{B}_{t'}(\sigma')$, with $t' < d^{\downarrow}(t)$. Since there exists some ranking at distance $d^{\downarrow}(t) > t'$ from σ', we have $\mathcal{B}_t(\sigma) \subseteq \mathcal{B}_{t'}(\sigma') \subsetneq \mathcal{B}_t(\sigma')$, which is a contradiction because $|\mathcal{B}_t(\sigma)| = |\mathcal{B}_t(\sigma')|$. Therefore, $k(t) \geq k(t, \pi) \geq d^{\downarrow}(t)$. ■

The bound of Theorem 2 holds for all $n, m > 0$ and all $t \in [0, D]$, where D is the maximum possible distance under d. It can be checked easily that the bound is tight given the neutrality assumption, which is an extremely mild — and in fact, a highly desirable — assumption for distance metrics over rankings.

Note that since $k(t)$ is a valid distance under d, the upper bound of $2t$ from Theorem 1 actually implies a possibly better upper bound of $d^{\downarrow}(2t)$. Thus, Theorems 1 and 2 establish $d^{\downarrow}(t) \leq k(t) \leq d^{\downarrow}(2t)$ for a wide range of distance metrics d. For the four special distance metrics considered in this paper, we close this gap by establishing a tight lower bound of $d^{\downarrow}(2t)$, for a wide range of values of n and t.

Theorem 3. *If $d \in \{d_{KT}, d_{FR}, d_{MD}, d_{CY}\}$, and the maximum distance allowed by the metric is $D \in \Theta(m^{\alpha})$, then there exists $T \in \Theta(m^{\alpha})$ such that:*

1. *For all $t \leq T$ and even n, we have $k(t) \geq d^{\downarrow}(2t)$.*
2. *For all $L \geq 2$, $t \leq T$ with $\{2t\} \in (1/L, 1-1/L)$, and odd $n \geq \Theta(L \cdot D)$, we have $k(t) \geq d^{\downarrow}(2t)$. Here, $\{x\} = x - \lfloor x \rfloor$ denotes the fractional part of $x \in \mathbb{R}$.*

Theorem 3 is our main theoretical result. Here we only provide a proof sketch for the Kendall tau distance, which gives some of the key insights. The full proof for Kendall tau, as well as for other distance metrics, can be found in the full version of the paper.[5]

Proof sketch for Kendall tau. Let d be the Kendall tau distance; thus, $D = \binom{m}{2}$ and $\alpha = 2$. First, we prove the case of even n. For a ranking $\tau \in \mathcal{L}(A)$, let τ_{rev} be its reverse. Assume $t = (1/2) \cdot \binom{m}{2}$, and fix a ranking $\sigma \in \mathcal{L}(A)$. Every ranking must agree with exactly one of σ and σ_{rev} on a given pair of alternatives. Hence, every $\rho \in \mathcal{L}(A)$ satisfies $d(\rho, \sigma) + d(\rho, \sigma_{\text{rev}}) = \binom{m}{2}$. Consider the profile π consisting of $n/2$ instances of σ and $n/2$ instances of σ_{rev}. Then, the average distance of every ranking from rankings in π would be exactly t, i.e., $\mathcal{B}_t(\pi) = \mathcal{L}(A)$. It is easy to check that $k(\mathcal{L}(A)) = \binom{m}{2} = 2t = d^{\downarrow}(2t)$ because every ranking has its reverse ranking in $\mathcal{L}(A)$ at distance exactly $2t$.

Now, let us extend the proof to $t \leq (m/6)^2$. If $t < 0.5$, then $d^{\downarrow}_{KT}(2t) = 0$, which is a trivial lower bound. Hence, assume $t \geq 0.5$. Thus, $d^{\downarrow}(2t) = \lfloor 2t \rfloor$. We use Fermat's Polygonal Number Theorem (see, e.g., Heath 2012). A special case of this remarkable theorem states that every natural number can be expressed as the sum of at most three "triangular" numbers, i.e., numbers of the form $\binom{k}{2}$. Let $\lfloor 2t \rfloor = \sum_{i=1}^{3} \binom{m_i}{2}$. It is easy to check that $0 \leq m_i \leq 2\sqrt{t}$ for all $i \in \{1, 2, 3\}$. Hence, $\sum_{i=1}^{3} m_i \leq 6\sqrt{t} \leq m$.

Partition the set of alternatives A into four disjoint groups A_1, A_2, A_3, and A_4 such that $|A_i| = m_i$ for $i \in \{1, 2, 3\}$, and $|A_4| = m - \sum_{i=1}^{3} m_i$. Let σ^{A_4} be an arbitrary ranking of the alternatives in A_4; consider the partial order $\mathcal{P}_A = A_1 \succ A_2 \succ A_3 \succ \sigma^{A_4}$ over alternatives in A. Note that a ranking ρ is an extension of $\mathcal{P}_A$ iff it ranks all alternatives in A_i before any alternative in A_{i+1} for $i \in \{1, 2, 3\}$, and ranks alternatives in A_4 according to σ^{A_4}. Choose arbitrary $\sigma^{A_i} \in \mathcal{L}(A_i)$ for $i \in \{1, 2, 3\}$ and define

$$\sigma = \sigma^{A_1} \succ \sigma^{A_2} \succ \sigma^{A_3} \succ \sigma^{A_4},$$
$$\sigma' = \sigma^{A_1}_{\text{rev}} \succ \sigma^{A_2}_{\text{rev}} \succ \sigma^{A_3}_{\text{rev}} \succ \sigma^{A_4}.$$

Note that both σ and σ' are extensions of $\mathcal{P}_A$. Once again, take the profile π consisting of $n/2$ instances of σ and $n/2$ instances of σ'. It is easy to check that a ranking disagrees with exactly one of σ and σ' on every pair of alternatives that belong to the same group in $\{A_1, A_2, A_3\}$. Hence, every ranking $\rho \in \mathcal{L}(A)$ satisfies

$$d(\rho, \sigma) + d(\rho, \sigma') \geq \sum_{i=1}^{3} \binom{m_i}{2} = \lfloor 2t \rfloor. \tag{1}$$

[5]The full version is accessible from http://www.cs.cmu.edu/~arielpro/papers.html

Clearly an equality is achieved in Equation (1) if and only if ρ is an extension of $\mathcal{P}(A)$. Thus, every extension of $\mathcal{P}_A$ has an average distance of $\lfloor 2t \rfloor / 2 \leq t$ from π. Every ranking ρ that is not an extension of $\mathcal{P}_A$ achieves a strict inequality in Equation (1); thus, $d(\rho, \pi) \geq (\lfloor 2t \rfloor + 1)/2 > t$. Hence, $\mathcal{B}_t(\pi)$ is the set of extensions of $\mathcal{P}_A$.

Given a ranking $\rho \in \mathcal{L}(A)$, consider the ranking in $\mathcal{B}_t(\pi)$ that reverses the partial orders over A_1, A_2, and A_3 induced by ρ. The distance of this ranking from ρ would be at least $\sum_{i=1}^{3} \binom{m_i}{2} = \lfloor 2t \rfloor$, implying $k(\mathcal{B}_t(\pi)) \geq \lfloor 2t \rfloor$. (In fact, it can be checked that $k(\mathcal{B}_t(\pi)) = \mathcal{D}(\mathcal{B}_t(\pi)) = \lfloor 2t \rfloor$.)

While we assumed that n is even, the proof can be extended to odd values of n by having one more instance of σ than σ'. The key idea is that with large n, the distance from the additional vote would have little effect on the average distance of a ranking from the profile. Thus, $\mathcal{B}_t(\pi)$ would be preserved, and the proof would follow. ■

The impossibility result of Theorem 3 is weaker for odd values of n (in particular, covering more values of t requires larger n), which is reminiscent of the fact that repetition (error-correcting) codes achieve greater efficiency with an odd number of repetitions; this is not merely a coincidence. Indeed, an extra repetition allows differentiating between tied possibilities for the ground truth; likewise, an extra vote in the profile prevents us from constructing a symmetric profile that admits a diverse set of possible ground truths (see the full version of the paper for details).

Approximations for Unknown Average Error

In the previous sections we derived the optimal rules when the upper bound t on the average error is given to us. In practice, the given bound may be inaccurate. We know that using an estimate $\widehat{t}$ that is still an upper bound ($\widehat{t} \geq t$) yields a ranking at distance at most $2\widehat{t}$ from the ground truth in the worst case. What happens if it turns out that $\widehat{t} < t$? We show that the output ranking is still at distance at most $4t$ from the ground truth in the worst case.

Theorem 4. *For a distance metric d, a profile π consisting of n noisy rankings at an average distance of at most t from the true ranking σ^*, and $\widehat{t} < t$, $d(\text{OPT}^d(\widehat{t}, \pi), \sigma^*) \leq 4t$.*

To prove the theorem, we make a detour through minisum rules. For a distance metric d, let $\textsc{MiniSum}^d$, be the voting rule that always returns the ranking minimizing the sum of distances (equivalently, average distance) from the rankings in the given profile according to d. Two popular minisum rules are the Kemeny rule for the Kendall tau distance ($\textsc{MiniSum}^{d_{KT}}$) and the minisum rule for the footrule distance ($\textsc{MiniSum}^{d_{FR}}$), which approximates the Kemeny rule (Dwork et al. 2001). For a distance metric d (dropped from the superscripts), let $d(\pi, \sigma^*) \leq t$. We claim that the minisum ranking $\textsc{MiniSum}(\pi)$ is at distance at most $\min(2t, 2k(t, \pi))$ from σ^*. This is true because the minisum ranking and the true ranking are both in $\mathcal{B}_t(\pi)$, and Lemma 1 shows that its diameter is at most $\min(2t, 2k(t, \pi))$.

Returning to the theorem, if we provide an underestimate $\widehat{t}$ of the true worst-case average error t, then using Lemma 1,

$$d\left(\textsc{MiniMax}(\mathcal{B}_{\widehat{t}}(\pi)), \textsc{MiniSum}(\pi)\right) \leq 2\widehat{t} \leq 2t,$$
$$d\left(\textsc{MiniSum}(\pi), \sigma^*\right) \leq \mathcal{D}(\mathcal{B}_t(\pi)) \leq 2t.$$

By the triangle inequality, $d\left(\textsc{MiniMax}(\mathcal{B}_{\widehat{t}}(\pi)), \sigma^*\right) \leq 4t$.

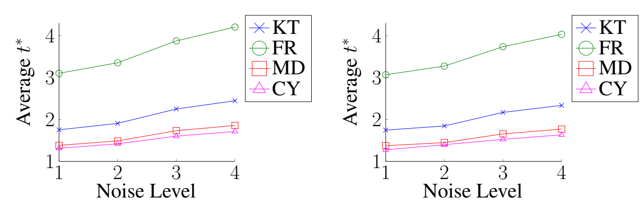
Experimental Results

We compare our worst-case optimal voting rules OPT^d against a plethora of voting rules used in the literature: plurality, Borda count, veto, the Kemeny rule, single transferable vote (STV), Copeland's rule, Bucklin's rule, the maximin rule, Slater's rule, Tideman's rule, and the modal ranking rule (for definitions see, e.g., Caragiannis et al. 2014).

Our performance measure is the distance of the output ranking from the *actual* ground truth. In contrast, for a given d, OPT^d is designed to optimize the worst-case distance to *any possible* ground truth. Hence, crucially, OPT^d *is not guaranteed to outperform other rules in our experiments*.

We use two real-world datasets containing ranked preferences in domains where ground truth rankings exist. Mao, Procaccia, and Chen (2013) collected these datasets — *dots* and *puzzle* — via Amazon Mechanical Turk. For dataset *dots* (resp., *puzzle*), human workers were asked to rank four images that contain a different number of dots (resp., different states of an 8-Puzzle) according to the number of dots (resp., the distances of the states from the goal state). Each dataset has four different noise levels (i.e., levels of task difficulty), represented using a single noise parameter: for dots (resp., puzzle), higher noise corresponds to ranking images with a smaller difference between their number of dots (resp., ranking states that are all farther away from the goal state). Each dataset has 40 profiles with approximately 20 votes each, for each of the 4 noise levels. Points in our graphs are averaged over the 40 profiles in a single noise level of a dataset.

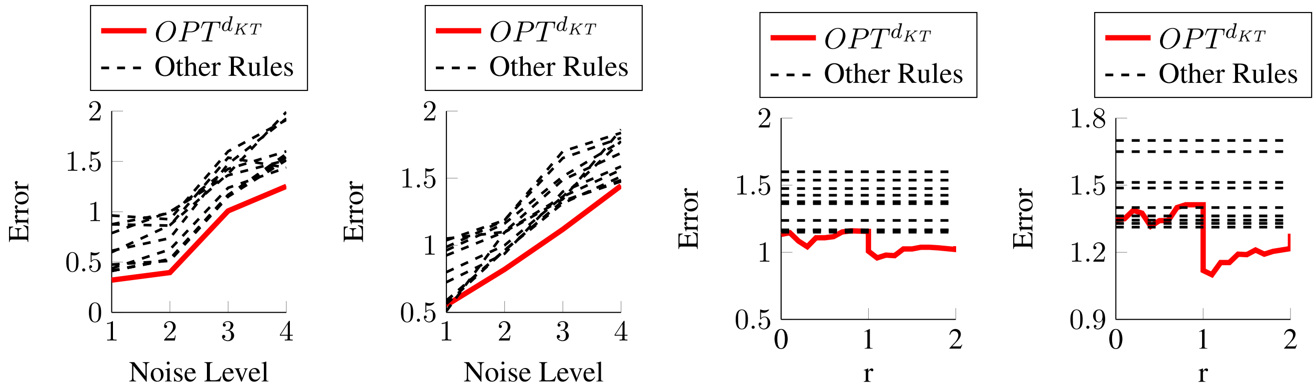
First, as a sanity check, we verified (Figure 1) that the noise parameter in the datasets positively correlates with our notion of noise — the average error in the profile, denoted t^* (averaged over all profiles in a noise level). Strikingly, the results from the two datasets are almost identical!



(a) Dots (b) Puzzle

Figure 1: Positive correlation of t^* with the noise parameter

Next, we compare OPT^d and $\textsc{MiniSum}^d$ against the voting rules listed above, with distance d as the measure of error. We use the average error in a profile as the bound t given to OPT^d, i.e., we compute $\text{OPT}^d(t^*, \pi)$ on profile π where $t^* = d(\pi, \sigma^*)$. While this is somewhat optimistic, note that t^* may not be the (optimal) value of t that achieves the lowest error. Also, the experiments below show that a reasonable estimate of t^* also suffices.



(a) Dots (b) Puzzle (c) Dots (d) Puzzle

Figure 2: Performance of different voting rules (Figures 2(a) and 2(b)), and of OPT with varying $\widehat{t}$ (Figures 2(c) and 2(d)).

Figures 2(a) and 2(b) show the results for the dots and puzzle datasets, respectively, under the Kendall tau distance. It can be seen that $\text{OPT}^{d_{KT}}$ (solid red line) significantly outperforms all other voting rules. The three other distance metrics considered in this paper generate similar results; the corresponding graphs appear in the full version.

Finally, we test OPT^d in the more demanding setting where only an estimate $\widehat{t}$ of t^* is provided. To synchronize the results across different profiles, we use $r = (\widehat{t} - \text{MAD})/(t^* - \text{MAD})$, where MAD is the *minimum average distance* of any ranking from the votes in a profile, that is, the average distance of the ranking returned by MINISUM^d from the input votes. For all profiles, $r = 0$ implies $\widehat{t} = \text{MAD}$ (the smallest value that admits a possible ground truth) and $r = 1$ implies $\widehat{t} = t^*$ (the true average error). In our experiments we use $r \in [0, 2]$; here, $\widehat{t}$ is an overestimate of t^* for $r \in (1, 2]$ (a valid upper bound on t^*), but an underestimate of t^* for $r \in [0, 1)$ (an invalid upper bound on t^*).

Figures 2(c) and 2(d) show the results for the dots and puzzle datasets, respectively, for a representative noise level (level 3 in previous experiments) and the Kendall tau distance. We can see that $\text{OPT}^{d_{KT}}$ (solid red line) outperforms all other voting rules as long as $\widehat{t}$ is a reasonable overestimate of t^* ($r \in [1, 2]$), but may or may not outperform them if $\widehat{t}$ is an underestimate of t^*. Again, other distance metrics generate similar results (see the full version for the details).

Discussion

Uniformly accurate votes. Motivated by crowdsourcing settings, we considered the case where the average error in the input votes is guaranteed to be low. Instead, suppose we know that every vote in the input profile π is at distance at most t from the ground truth σ^*, i.e., $\max_{\sigma \in \pi} d(\sigma, \sigma^*) \leq t$. If t is small, this is a stronger assumption because it means that there are no outliers, which is implausible in crowdsourcing settings but plausible if the input votes are expert opinions. In this setting, it is immediate that any vote in the given profile is at distance at most $d^{\downarrow}(t)$ from the ground truth. Moreover, the proof of Theorem 2 goes through, so this bound is tight *in the worst case*; however, returning a ranking from the profile is *not optimal for every profile*.

Randomization. We did not consider randomized rules, which may return a distribution over rankings. If we take the error of a randomized rule to be the expected distance of the returned ranking from the ground truth, it is easy to obtain an upper bound of t. Again, the proof of Theorem 2 can be extended to yield an almost matching lower bound of $d^{\downarrow}(t)$. While randomized rules provide better guarantees, they are often impractical: low error is only guaranteed when rankings are repeatedly selected from the output distribution of the randomized rule on the same profile; however, most social choice settings see only a single outcome realized.[6]

Complexity. A potential drawback of the proposed approach is computational complexity. For example, consider the Kendall tau distance. When t is small enough, only the Kemeny ranking would be a possible ground truth, and $\text{OPT}^{d_{KT}}$ or any finite approximation thereof must return the Kemeny ranking, if it is unique. The $\mathcal{NP}$-hardness of computing the Kemeny ranking (Bartholdi, Tovey, and Trick 1989) therefore suggests that computing or approximating $\text{OPT}^{d_{KT}}$ is $\mathcal{NP}$-hard.

That said, fixed-parameter tractable algorithms, integer programming solutions, and other heuristics are known to provide good performance for these types of computational problems in practice (see, e.g., Betzler et al. 2011; 2014). More importantly, the crowdsourcing settings that motivate our work inherently restrict the number of alternatives to a relatively small constant: a human would find it difficult to effectively rank more than, say, 10 alternatives. With a constant number of alternatives, we can simply enumerate all possible rankings in polynomial time, making each and every computational problem considered in this paper tractable. In fact, this is what we did in our experiments. Therefore, we do not view computational complexity as an insurmountable obstacle.

[6]Exceptions include cases where randomization is used for circumventing impossibilities (Procaccia 2010; Conitzer 2008; Boutilier et al. 2012).

Acknowledgements

Procaccia and Shah were partially supported by the NSF under grants IIS-1350598 and CCF-1215883.

References

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

Azari Soufiani, H.; Parkes, D. C.; and Xia, L. 2012. Random utility theory for social choice. In *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS)*, 126–134.

Azari Soufiani, H.; Parkes, D. C.; and Xia, L. 2013. Preference elicitation for general random utility models. In *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 596–605.

Azari Soufiani, H.; Parkes, D. C.; and Xia, L. 2014. Computing parametric ranking models via rank-breaking. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, 360–368.

Barg, A., and Mazumdar, A. 2010. Codes in permutations and error correction for rank modulation. *IEEE Transactions on Information Theory* 56(7):3158–3165.

Bartholdi, J.; Tovey, C. A.; and Trick, M. A. 1989. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare* 6:157–165.

Betzler, N.; Guo, J.; Komusiewicz, C.; and Niedermeier, R. 2011. Average parameterization and partial kernelization for computing medians. *Journal of Computer and System Sciences* 77(4):774–789.

Betzler, N.; Bredereck, R.; and Niedermeier, R. 2014. Theoretical and empirical evaluation of data reduction for exact Kemeny rank aggregation. *Autonomous Agents and Multi-Agent Systems* 28(5):721–748.

Boutilier, C.; Caragiannis, I.; Haber, S.; Lu, T.; Procaccia, A. D.; and Sheffet, O. 2012. Optimal social choice functions: A utilitarian view. In *Proceedings of the 13th ACM Conference on Economics and Computation (EC)*, 197–214.

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

Caragiannis, I.; Procaccia, A. D.; and Shah, N. 2014. Modal ranking: A uniquely robust voting rule. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*, 616–622.

Conitzer, V., and Sandholm, T. 2005. Communication complexity of common voting rules. In *Proceedings of the 6th ACM Conference on Economics and Computation (EC)*, 78–87.

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

Conitzer, V. 2008. Anonymity-proof voting rules. In *Proceedings of the 4th International Workshop on Internet and Network Economics (WINE)*, 295–306.

Dwork, C.; Kumar, R.; Naor, M.; and Sivakumar, D. 2001. Rank aggregation methods for the web. In *Proceedings of the 10th International World Wide Web Conference (WWW)*, 613–622.

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

Guruswami, V. 2005. *List Decoding of Error-Correcting Codes*. Springer.

Haeupler, B. 2014. Interactive channel capacity revisited. In *Proceedings of the 55th Symposium on Foundations of Computer Science (FOCS)*. Forthcoming.

Heath, T. L. 2012. *Diophantus of Alexandria: A study in the history of Greek algebra*. CUP Archive.

Lee, J.; Kladwang, W.; Lee, M.; Cantu, D.; Azizyan, M.; Kim, H.; Limpaecher, A.; Yoon, S.; Treuille, A.; and Das, R. 2014. RNA design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences*. Forthcoming.

Lu, T., and Boutilier, C. 2011. Robust approximation and incremental elicitation in voting protocols. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, 287–293.

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

Mao, A.; Procaccia, A. D.; and Chen, Y. 2013. Better human computation through principled voting. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI)*, 1142–1148.

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

Procaccia, A. D.; Rosenschein, J. S.; and Kaminka, G. A. 2007. On the robustness of preference aggregation in noisy environments. In *Proceedings of the 6th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 416–422.

Procaccia, A. D. 2010. Can approximation circumvent Gibbard-Satterthwaite? In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, 836–841.

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

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

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