

Learning a Ground Truth Ranking Using Noisy Approval Votes

Ioannis Caragiannis

University of Patras
caragian@ceid.upatras.gr

Evi Micha

University of Patras
michap@ceid.upatras.gr

Abstract

We consider a voting scenario where agents have opinions that are estimates of an underlying common ground truth ranking of the available alternatives, and each agent is asked to approve a set with her most preferred alternatives. We assume that estimates are implicitly formed using the well-known Mallows model for generating random rankings. We show that k -approval voting — where all agents are asked to approve the same number k of alternatives and the outcome is obtained by sorting the alternatives in terms of their number of approvals — has exponential sample complexity for all values of k . This negative result suggests that an exponential (in terms of the number of alternatives m) number of agents is always necessary in order to recover the ground truth ranking with high probability. In contrast, by just asking each agent to approve a random number of alternatives, the sample complexity improves dramatically: it now depends only polynomially on m . Our results may have implications on the effectiveness of crowdsourcing applications that ask workers to provide their input by approving sets of available alternatives.

1 Introduction

How should preferences of different individuals be aggregated into a collective decision? This is the central question that social choice theory aims to answer. In a well-known model, there is a set of available alternatives and each individual preference is expressed as a ranking over these alternatives. Given such a profile of preferences, a *voting rule* outputs either a winning alternative or a ranking of all the alternatives as a collective decision.

Social choice theory has extensively studied the different voting rules that can be defined, following two main approaches. The first one, which is the most common, is *axiomatic*. It assumes that voting rules should satisfy some desirable social choice axioms and aims to characterize voting rules in terms of the axioms they satisfy. Well-known impossibility results — e.g. the results by Arrow [1953], Gibbard [1973], and Satterthwaite [1975] — indicate the challenges of this approach. The second approach assumes that there is

a *ground truth* (i.e., an underlying objective ranking over the alternatives) and views votes as *noisy estimates* of the ground truth. Then, a voting rule is better than another if it is more likely to output the ground truth; the best such rule is a *maximum likelihood estimator* (MLE) of the ground truth.

Even though it originates from Marquis de Condorcet — the founder of social choice theory — more than two centuries ago, the MLE approach became popular only very recently, due to the work of Young [1988], Conitzer and Sandholm [2005], and others. Interestingly, Marquis de Condorcet proposed a very simple process as a model of the mental process with which voters decide their individual rankings: each voter decides the relative ranks of every pair of alternatives a and b correctly (i.e., consistently to the relative ranks of a and b in the ground truth) with probability p and incorrectly with probability $1-p$. The process is repeated until these decisions define a ranking. This model was rediscovered by Mallows [1957] and today is known as the *Mallows noise model*.

When voting rules are viewed as estimators of a ground truth, an important measure of their effectiveness is their *sample complexity*. How many votes do we need in a profile so that the application of a voting rule recovers the ground truth with high probability? Caragiannis *et al.* [2016] proved that maximum likelihood estimators are the voting rules that have optimal sample complexity and presented positive and negative sample complexity results for Mallows. See also [Braverman and Mossel, 2008; Chierichetti *et al.*, 2014; Rubinfeld and Vardi, 2017] for results of similar flavour in slightly different (non-voting) contexts. Chierichetti and Kleinberg [2014] and Dey and Battacharyya [2015] present sample complexity results for voting rules that return a winning alternatives (instead of a complete ranking, which is our main focus here).

In this paper, we study the sample complexity of *approval voting* on profiles that are generated using the Mallows noise model. Approval voting is simultaneously a simplified format of voters' preferences and a voting rule. With an approval vote, a voter just approves some of the alternatives (and disapproves the rest). The approval voting rule takes a profile with approval votes as input, ranks the alternatives in monotone non-increasing order in terms of their number of approvals, and returns this ranking as output. Besides its simplicity, approval voting has many interesting properties; the books by Brams and Fishburn [2007] and Laslier and San-

ver [2010] are devoted to its study. It also seems to be ideal for several *crowdsourcing* tasks, as it provides a very simple interface through which workers can provide their input; see [Shah *et al.*, 2015; Lee *et al.*, 2014]. A concrete scenario could be the evaluation of alternative projects that a local government considers for implementation. Adapting the recent trend of *participatory democracy*, the local government can ask a “crowd” of citizens to approve a given number of preferred projects each, in order to come up with an ordering of them in terms of their approval score.

Here, our assumptions are that there is a ground truth ranking and each voter uses the Mallows model to form an internal opinion (i.e., a ranking) over the alternatives. Then, her approval vote consists of the top alternatives in her internal opinion. What is the sample complexity of approval voting in this setting? In other words, how many noisy approval votes do we need in order to recover (learn) the ground truth ranking with high probability?

As Procaccia and Shah [2015] show, approval voting is the maximum likelihood estimator for approval votes that have been produced using Mallows. Then, by the result of Caragiannis *et al.* [2016] mentioned above, we can conclude — after some adaptation of their proof — that approval voting has optimal sample complexity in the setting we consider. Is this a positive result? Not necessarily, as we will show in Section 3. In particular, we first consider profiles with approval votes that consist of k alternatives. We show that approval voting has *exponential* sample complexity for all possible values of k . This result extends a negative result from [Caragiannis *et al.*, 2016] for plurality. Is that all we can hope for with approval votes? Surprisingly, as we show in Section 4, in profiles where each voter approves a random number of alternatives, approval voting has only *polynomial* sample complexity. We use the term *randomized approval* voting to refer to this voting rule.

We have also conducted experiments with randomized approval (and with a deterministic variant of it) that verify our positive theoretical result; these are also presented in Section 4 together with a discussion on possible extensions of our proof. Actually, the Mallows sample complexity is experimentally observed to be even better than our theoretical bound. We conclude in Section 5, with a short discussion on open problems and an interesting (and somewhat surprising) observation about an alternative definition of the model of building Mallows approval votes.

2 Preliminaries

We begin with preliminary definitions. We consider settings with n voters (or *agents*) and m alternatives. We denote by $A = \{a_1, a_2, \dots, a_m\}$ the set of alternatives and by $N = \{1, \dots, n\}$ the set of agents. We use $\mathcal{L}(A)$ to denote the set of rankings of the alternatives in A . For a ranking $\sigma \in \mathcal{L}(A)$, we denote by $\sigma(a)$ the rank (or position, taking values between 1 to m) of alternative $a \in A$ in σ .

An approval vote is the set of alternatives approved by an agent; we use A_ℓ to denote the approval vote of agent ℓ . A k -approval vote has exactly k alternatives. We consider voting profiles in which each agent submits an approval vote. A vot-

ing profile is a collection of approval votes of the agents, i.e., $\{A_\ell | \ell \in N\}$. The approval score of an alternative $a \in A$ in a profile is simply the number of its appearances in approval votes, i.e., $|\{\ell \in N | a \in A_\ell\}|$.

The well-known approval voting rule computes a final ranking by sorting the alternatives in monotone non-increasing order in terms of their approval score. Ties (i.e., deciding the ranks of alternatives that have equal approval scores) can be broken either randomly or deterministically. The approval voting rule can be applied to profiles in which agents have approval votes of different sizes or to profiles in which all agents’ preferences are restricted to k -approval votes for some fixed k . As a convention, approval votes are always non-empty and exclude at least one alternative.

We assume that there is an unknown ground truth ranking $\sigma^* \in \mathcal{L}(A)$ of the alternatives. Each agent $\ell \in N$ has an opinion $\pi_\ell \in \mathcal{L}(A)$, which is a noisy estimate of σ^* . The noisy estimate of agent ℓ is computed using a noise model that takes as input the ground truth ranking σ^* and returns a random ranking $\sigma \in \mathcal{L}(A)$ with a probability that depends on σ and σ^* . So, when agent ℓ is asked to submit a k -approval vote, it returns the first k alternatives in π_ℓ ; we use top_k to denote the k top alternatives in ranking σ .

We assume that the agents decide their opinions by (implicitly) applying the Mallows noise model. This model uses a noise parameter $p \in [1/2, 1]$ (or, equivalently, noise parameter $\phi = \frac{1-p}{p}$) and produces the (random) ranking σ by deciding the relation of each pair of alternatives separately: for alternatives a and b from A with $\sigma^*(a) < \sigma^*(b)$, it is decided that a beats b (i.e., $a \succ b$) with probability p and that $b \succ a$ with probability $1 - p$. This is continued for each pair of alternatives; this process defines a round. If the relation \succ induces a ranking, the round is successful and this ranking is the noisy estimate returned by the Mallows process; otherwise, the whole process is repeated from scratch (with a new round that uses new randomness). We denote by $\text{Mallows}(\phi, \sigma^*)$ the random ranking returned by the Mallows process with noise parameter ϕ and ground truth σ^* .

It is well-known — e.g., see [Caragiannis *et al.*, 2016] — that the Mallows probabilities are defined as

$$\Pr[\text{Mallows}(\phi, \sigma^*) = \sigma] = \frac{\phi^{d(\sigma, \sigma^*)}}{\sum_{\sigma' \in \mathcal{L}(A)} \phi^{d(\sigma', \sigma^*)}},$$

where $d(\sigma_1, \sigma_2)$ is the Kendall-tau distance between rankings σ_1 and σ_2 defined as the number of pairwise relations between alternatives that differ in σ_1 and σ_2 , i.e.,

$$d(\sigma_1, \sigma_2) = \sum_{a, b \in A: \sigma_1(a) < \sigma_1(b)} \mathbb{I}\{\sigma_2(a) > \sigma_2(b)\}.$$

3 The Mallows Sample Complexity of k -Approval Voting

We begin with a negative result for k -approval voting, extending a negative result of Caragiannis *et al.* [2016] for plurality. Namely, we will show that, for all values of k , k -approval voting has exponential sample complexity for almost all values of noise parameters (Theorem 1).

Let us give some intuition for the result, for the two extreme values of the noise parameter. As the noise parameter approaches $1/2$, the votes are selected almost uniformly at random from $\mathcal{L}(A)$. Then, it should be clear that a huge number of votes is required in order to detect any bias towards the ground truth. In contrast, when the votes have no noise at all, all agents submit the same set of alternatives and there is no way to recover the ground truth ranking.

We begin with the following technical lemma; note that the assumption about σ^* is without loss of generality.

Lemma 1. *Let σ^* be a ground truth ranking that has alternatives a_i ranked i -th and let $S = \text{top}_k(\text{Mallows}(\phi, \sigma^*))$ be a k -approval set that is produced using Mallows with noise parameter ϕ . Then,*

$$\Pr[\{a_{m-1}, a_m\} \cap S \neq \emptyset] \leq 2\phi^{m-1-k} \quad (1)$$

$$\Pr[\{a_1, a_2\} \cap S = \emptyset] \leq 2\phi^{k-1}. \quad (2)$$

Proof. Due to lack of space, we prove only (1) here. First, observe that a round of the Mallows process is successful if the pairwise relations among the alternatives in $A \setminus \{a_{m-1}, a_m\}$ induce a (partial) ranking σ and there are different integers $i, j \in [m]$ so that the pairwise relations involving alternatives a_{m-1} and a_m are as follows:

- either $1 \leq i < j \leq m$ and alternative a_{m-1} is beaten by the top $i-1$ alternatives in σ , beats the rest including a_m , and alternative a_m is beaten by the top $j-2$ alternatives in σ (and by a_{m-1}) and beats the rest,
- or $1 \leq j < i \leq m$ and alternative a_{m-1} is beaten by the top $i-2$ alternatives in σ and by alternative a_m , beats the rest, and alternative a_m is beaten by the top $j-1$ alternatives in σ and beats the rest including a_{m-1} .

In both cases, the integers i and j denote the ranks of a_{m-1} and a_m in the ranking produced.

So, assuming that the round of the Mallows process has decided the pairwise relations among the alternatives in $A \setminus \{a_{m-1}, a_m\}$ so that they induce a partial ranking σ , the probability that the remaining pairwise relations that involve a_{m-1} and a_m are decided so that a full ranking is produced with alternatives a_{m-1} and a_m ranked i -th and j -th, respectively, is $p^{i+j-2}(1-p)^{2m-i-j-1}$ if $i < j$ and $p^{i+j-3}(1-p)^{2m-i-j}$ if $i > j$. To see why, observe that the probability that alternative a_{m-1} (or a_m) is decided to be beaten by any of $m-2$ alternatives in σ is p while the probability that a_{m-1} is decided to beat a_m is p as well. Clearly, these probabilities do not depend on the ranking σ .

Hence, the probability that a round of the Mallows process, that has decided the pairwise relations among the alternatives in $A \setminus \{a_{m-1}, a_m\}$ so that they induce a partial ranking, decides the remaining pairwise relations that involve a_{m-1} and a_m so that a full ranking is induced and neither a_{m-1} nor a_m occupies some of the t top position (i.e., $i, j \geq t+1$) is

$$\begin{aligned} P(t) &= \sum_{i=t+1}^{m-1} \sum_{j=i+1}^m p^{i+j-2}(1-p)^{2m-i-j-1} \\ &\quad + \sum_{i=t+2}^m \sum_{j=t+1}^{i-1} p^{i+j-3}(1-p)^{2m-i-j} \end{aligned}$$

$$= \frac{(1-p)^{2m-1}}{p^3} \sum_{i=t+1}^{m-1} \sum_{j=i+1}^m \left(\frac{p}{1-p} \right)^{i+j}. \quad (3)$$

The first (respectively, the second) sum in the first equality above runs over all pairs of positions i, j such that $t+1 \leq i < j \leq m$ (respectively, $t+1 \leq j < i \leq m$) and sums up the probabilities that alternatives a_{m-1} and a_m occupy these positions at the end of the round. The parameter t can take any value from 0 to $m-2$.

In order to simplify notation, set $\lambda = \frac{p}{1-p}$ (i.e., $\lambda = 1/\phi$) and define $f(t) = \sum_{i=t+1}^{m-1} \sum_{j=i+1}^m \lambda^{i+j}$ for any non-negative integer t . Then, (3) becomes

$$P(t) = \frac{(1-p)^{2m-1}}{p^3} f(t). \quad (4)$$

By simple calculations, we have

$$\begin{aligned} f(t) &= \sum_{i=t+1}^{m-1} \lambda^i \sum_{j=i+1}^m \lambda^j = \sum_{i=t+1}^{m-1} \lambda^i \frac{\lambda^{m+1} - \lambda^{i+1}}{\lambda - 1} \\ &= \frac{\lambda^{m+1}}{\lambda - 1} \sum_{i=t+1}^{m-1} \lambda^i - \frac{\lambda}{\lambda - 1} \sum_{i=t+1}^{m-1} \lambda^{2i} \\ &= \frac{\lambda^{m+1}(\lambda^m - \lambda^{t+1})}{(\lambda - 1)^2} - \frac{\lambda(\lambda^{2m} - \lambda^{2(t+1)})}{(\lambda - 1)(\lambda^2 - 1)} \\ &= \frac{\lambda^3(\lambda^m - \lambda^t)(\lambda^{m-1} - \lambda^t)}{(\lambda - 1)^2(\lambda + 1)}. \end{aligned} \quad (5)$$

Using (4) and (5), we have that the probability that the Mallows process returns a ranking with some alternative among a_{m-1} and a_m in the top k positions is equal to

$$\begin{aligned} 1 - P(k)/P(0) &= 1 - \frac{(\lambda^m - \lambda^k)(\lambda^{m-1} - \lambda^k)}{(\lambda^m - 1)(\lambda^{m-1} - 1)} \\ &= \frac{(\lambda^m + \lambda^{m-1} - 1 - \lambda^k)(\lambda^k - 1)}{(\lambda^m - 1)(\lambda^{m-1} - 1)} \\ &\leq \frac{2(\lambda^k - 1)}{\lambda^{m-1} - 1} \leq 2\lambda^{k+1-m} = 2\phi^{m-1-k}. \end{aligned}$$

The two inequalities follow since $\lambda \geq 1$ and $k \leq m-1$. This completes the proof of inequality (1). \square

In order to distinguish between two alternatives and learn their relative positions in the ground truth ranking using approval voting, we need at least one approval vote in which the one alternative is approved and the other is not. Therefore, the useful (for our purposes) information in the statement of Lemma 1 is that the probability that a Mallows k -approval vote can distinguish between the top two (if k is large) or the bottom two (if k is small) alternatives in the ground truth ranking σ^* is exponentially small. Then, it follows that an exponential number of Mallows approval votes is necessary just in order to distinguish between the top or the bottom two alternatives of σ^* .

Theorem 1. *For every integer $k \in [m-1]$ and $\epsilon \in [0, 1/4]$, at least $\Omega(\phi^{-m/2})$ Mallows k -approval votes are necessary in order to obtain the ground truth ranking among m alternatives with probability at least $1 - \epsilon$.*

Proof. Assuming that the number of agents is strictly smaller than $\frac{1}{4}\phi^{1-m/2}$, we will show that the probability that we can correctly decide the position of two particular alternatives in the ground truth ranking is less than $3/4 \leq 1 - \epsilon$. We distinguish between two cases:

When $k \leq m/2$, using inequality (1) from Lemma 1 and the union bound, we have that the probability that some k -approval set in the profile contains either alternative a_{m-1} or alternative a_m (or both) is less than $\frac{1}{4}\phi^{1-m/2} \times 2\phi^{m-1-k} \leq 1/2$; the inequality follows since $k \leq m/2$. Hence, the probability that a_{m-1} gets strictly more approvals than a_m or their tie is resolved in favour of a_{m-1} is less than $3/4$.

For $k > m/2$, we use inequality (2) from Lemma 1 and the union bound to obtain that the probability that some k -approval set in the profile does not contain alternatives a_1 and a_2 is less than $\frac{1}{4}\phi^{1-m/2} \times 2\phi^{k-1} \leq 1/2$ as well. So, we obtain that the probability that a_1 gets strictly more approvals than a_2 or their tie is resolved in favour of a_1 is less than $3/4$.

Hence, in order to get the ground truth with probability at least $1 - \epsilon$, at least $\frac{1}{4}\phi^{1-m/2}$ agents are necessary. \square

4 Randomized Approval Voting Has Only Polynomial Mallows Sample Complexity

The negative result in Section 3 is certainly disappointing. Still, there is a surprisingly simple way to recover the ground truth using only polynomially many Mallows approval votes. The main idea is to ask agents to approve different numbers of alternatives. In particular, let us define the randomized approval voting rule, which, for every agent, picks an integer r uniformly at random from $[m-1]$ and asks the agent to approve r alternatives. Then, it ranks the alternatives in decreasing order of approvals, breaking ties uniformly at random.

Before proving our sample complexity bound for randomized approval, we give some intuition. Let us identify the agents that are asked to approve r alternatives as the ones belonging to the r -th group. Now observe that if we have sufficiently many (say, $\Theta(m^2)$) agents in total, then, almost surely, all the $m-1$ possible groups of agents will be non-empty. If there is no noise at all, each agent in the r -th group will correctly approve the top r alternatives in the ground truth ranking and, as we have agent groups and sets of approved alternatives of all possible sizes, sorting the alternatives in terms of their number of approvals will reveal the ground truth.

This intuition is valid in the presence of noise as well. If we have sufficiently many agents in the first group, they will most probably give the highest approval score to the top alternative in the ground truth. Similarly, a sufficiently large number of agents in the second group will distinguish the two top alternatives in the ground truth. Combined with the information from the first group of agents, we will have the correct ranking for the two top alternatives. The intuition proceeds this way and indicates that, most probably, the whole information that is needed in order to recover the ground truth ranking is in the approved sets of alternatives.

We are now ready to formally prove our positive result. An interesting feature in the statement of Theorem 2 is that it involves agents with different noise levels.

Theorem 2. *For every $\epsilon > 0$, when the Mallows agents have possibly different noise parameters with average value $p^* = \frac{1}{n} \sum_{\ell \in N} p_\ell$ and their number is*

$$n \geq \frac{2(m-1)^2}{(2p^*-1)^2} \ln(m/\epsilon),$$

randomized approval returns the ground truth ranking of m alternatives with probability at least $1 - \epsilon$.

Proof. Without loss of generality, we assume that the ground truth ranking σ^* has alternative a_i of A at position i . For alternative $a_i \in A$, let q_i be the random variable that denotes the number of approvals a_i gets from all agents under the randomized approval voting rule. We will show that the probability that $q_i \leq q_{i+1}$ for some $i \in [m-1]$ is very small, proving the theorem.

Let $\ell \in N$ be an agent and define D_ℓ depending on whether ℓ approves alternatives a_i and a_{i+1} as follows:

$$D_\ell = \begin{cases} 1 & \text{if agent } j \text{ approves } a_i \text{ but not } a_{i+1} \\ 0 & \text{if agent } j \text{ approves or disapproves} \\ & \text{both } a_i \text{ and } a_{i+1} \\ -1 & \text{if agent } j \text{ approves } a_{i+1} \text{ but not } a_i \end{cases}$$

Hence,

$$q_i - q_{i+1} = \sum_{\ell \in N} D_\ell. \quad (6)$$

We would like to show that $\Pr[q_i - q_{i+1} \leq 0] \leq \epsilon/m$. The theorem will then follow by summing over all $i \in [m-1]$.

Observe that the quantity $q_i - q_{i+1}$ is the sum of n independent random variables, one per agent, taking values in $\{-1, 0, 1\}$. We will use the following well-known inequality:

Lemma 2. [Hoeffding, 1963] *Let X_1, X_2, \dots, X_t be independent random variables so that $\Pr[a_j \leq X_j \leq b_j] = 1$ for $j \in [t]$. Then, the expectation of the random variable $X = \sum_{j=1}^t X_j$ is $\mathbb{E}[X] = \sum_{j=1}^t \mathbb{E}[X_j]$ and, furthermore, for every $\nu \geq 0$,*

$$\Pr[X \leq \mathbb{E}[X] - \nu] \leq \exp\left(-\frac{2\nu^2}{\sum_{j=1}^t (b_j - a_j)^2}\right).$$

We apply Lemma 2 to the random variable $q_i - q_{i+1} = \sum_{\ell=1}^n D_\ell$ with $\nu = \mathbb{E}[q_i - q_{i+1}]$. Observe that $t = n$, and $a_\ell = -1, b_\ell = 1$ for $\ell \in [n]$. Hence, we have

$$\begin{aligned} \Pr[q_i - q_{i+1} \leq 0] &\leq \exp\left(-\frac{\mathbb{E}[q_i - q_{i+1}]^2}{2n}\right) \\ &= \exp\left(-\frac{(\sum_{\ell \in N} \mathbb{E}[D_\ell])^2}{2n}\right). \end{aligned} \quad (7)$$

It remains to bound $\mathbb{E}[D_\ell]$ from below. Let $\ell \in N$ be any agent. By the definition of D_ℓ , we have that

$$\mathbb{E}[D_\ell] = \Pr[D_\ell = 1] - \Pr[D_\ell = -1]. \quad (8)$$

Clearly, the event $D_\ell = 1$ happens when the Mallows process returns a ranking σ in which alternative a_i has better rank than a_{i+1} . Then, alternative a_i is approved and alternative

a_{i+1} is not approved when agent ℓ is asked to approve the r top alternatives with r lying between $\sigma(a_i)$ (the rank of a_i in σ) and $\sigma(a_{i+1}) - 1$. Since r is selected uniformly at random, we have that the probability that r lies in this particular range is $\frac{\sigma(a_{i+1}) - \sigma(a_i)}{m-1}$.

Similarly, the event $D_\ell = -1$ happens when the Mallows process returns a ranking σ in which alternative a_i has worse rank than a_{i+1} . Then, alternative a_{i+1} is approved and alternative a_i is not approved when agent ℓ is asked to approve the r top alternatives with r lying between $\sigma(a_{i+1})$ and $\sigma(a_i) - 1$. The probability that r lies in this particular range is $\frac{\sigma(a_i) - \sigma(a_{i+1})}{m-1}$.

Denote by $\tilde{\sigma}$ the ranking obtained by swapping alternatives a_i and a_{i+1} in σ and abbreviate $\Pr[\text{Mallows}(\phi_\ell, \sigma^*) = \sigma]$ by $\Pr[\sigma]$. By the discussion above, (8) yields

$$\begin{aligned}
 \mathbb{E}[D_\ell] &= \sum_{\substack{\sigma \in \mathcal{L}(A): \\ a_i \succ a_{i+1}}} \Pr[\sigma] \cdot \frac{\sigma(a_{i+1}) - \sigma(a_i)}{m-1} \\
 &\quad - \sum_{\substack{\sigma \in \mathcal{L}(A): \\ a_{i+1} \succ a_i}} \Pr[\sigma] \cdot \frac{\sigma(a_i) - \sigma(a_{i+1})}{m-1} \\
 &= \sum_{\substack{\sigma \in \mathcal{L}(A): \\ a_i \succ a_{i+1}}} \Pr[\sigma] \cdot \frac{\sigma(a_{i+1}) - \sigma(a_i)}{m-1} \\
 &\quad - \sum_{\substack{\sigma \in \mathcal{L}(A): \\ a_i \succ a_{i+1}}} \Pr[\tilde{\sigma}] \cdot \frac{\sigma(a_{i+1}) - \sigma(a_i)}{m-1} \\
 &= \sum_{\substack{\sigma \in \mathcal{L}(A): \\ a_i \succ a_{i+1}}} (1 - \phi_\ell) \cdot \Pr[\sigma] \cdot \frac{\sigma(a_{i+1}) - \sigma(a_i)}{m-1} \\
 &\geq \frac{1 - \phi_\ell}{m-1} \sum_{\substack{\sigma \in \mathcal{L}(A): \\ a_i \succ a_{i+1}}} \Pr[\sigma]. \tag{9}
 \end{aligned}$$

The first two equalities follow by the discussion above while the inequality uses the trivial fact that $\sigma(a_{i+1}) - \sigma(a_i) \geq 1$ when alternative a_i is ranked higher than a_{i+1} in σ . We need to justify the property $\Pr[\tilde{\sigma}] = \phi_\ell \cdot \Pr[\sigma]$ which is used to obtain the third equality in the derivation above. This follows by the definition of the Mallows probabilities by observing that $d(\tilde{\sigma}, \sigma^*) = d(\sigma, \sigma^*) + 1$. To see why, recall the definition of the Kendall-tau distance from Section 2 and notice that every pair of alternatives a_j and $a_{j'}$ with j, j' different than i and $i+1$ have the same contribution to the quantities $d(\tilde{\sigma}, \sigma^*)$ and $d(\sigma, \sigma^*)$. Also, for j different than i and $i+1$, the contribution of the pair of alternatives a_i and a_j to $d(\tilde{\sigma}, \sigma^*)$ (respectively, to $d(\sigma, \sigma^*)$) is equal to the contribution of the pair of alternatives a_{i+1} and a_j to $d(\sigma, \sigma^*)$ (respectively, to $d(\tilde{\sigma}, \sigma^*)$). It remains to consider the contribution of the pair of alternatives a_i and a_{i+1} . Clearly, this contributes (one) to $d(\tilde{\sigma}, \sigma^*)$ but not to $d(\sigma, \sigma^*)$.

Using again the fact $\Pr[\tilde{\sigma}] = \phi_\ell \cdot \Pr[\sigma]$, we also have

$$1 = \sum_{\substack{\sigma \in \mathcal{L}(A): \\ a_i \succ a_{i+1}}} (\Pr[\sigma] + \Pr[\tilde{\sigma}]) = (1 + \phi_\ell) \sum_{\substack{\sigma \in \mathcal{L}(A): \\ a_i \succ a_{i+1}}} \Pr[\sigma],$$

and (9) now yields

$$\mathbb{E}[D_\ell] \geq \frac{1 - \phi_\ell}{(1 + \phi_\ell)(m-1)} = \frac{2p_\ell - 1}{m-1}$$

and, hence,

$$\sum_{\ell \in N} \mathbb{E}[D_\ell] = \frac{n(2p^* - 1)}{m-1}.$$

Using (7) and the bound on the number of agents n , we get

$$\Pr[q_i - q_{i+1} \leq 0] \leq \exp\left(-\frac{n(2p^* - 1)^2}{2(m-1)^2}\right) \leq \epsilon/m,$$

as desired. \square

An advantage of randomization is that it guarantees that the r -th group of agents will have size very close to $n/(m-1)$ for every possible value of r in $[m-1]$.¹ Alternatively, this can be enforced deterministically, e.g., in a round-robin fashion; in this way, each group of agents will have size equal to either $\lceil n/(m-1) \rceil$ or $\lfloor n/(m-1) \rfloor$ for all values of k . Let us call this variant *balanced approval* voting. We do not see how the proof of Theorem 2 can be adapted to work in this case since our proof exploits the fact that a Mallows agent is asked to approve a random number of alternatives in order to avoid a detailed reasoning about the outcome of the Mallows distribution. Still, we strongly believe that balanced approval has polynomial sample complexity at least in the scenario of agents with identical Mallows parameters. Experimental results that are presented below fully justify this belief.

On the other hand, if agents have very different Mallows parameters, the bound of Theorem 2 should not be expected to hold for balanced approval voting. Indeed, it might happen that low-noise agents are the ones that are asked to approve either very few or too many alternatives. Then, all such an agent can do is to contribute towards distinguishing between the few alternatives at the top or the few alternatives in the bottom. At the same time, the agents who are asked to contribute the most (i.e., the ones who are asked to approve, say, half of the alternatives) might happen to be the high-noise ones. A possible solution for avoiding this correlation would be to keep the balancedness restriction but partition the agents into different groups randomly. Then, a statement similar to the one in Theorem 1 certainly holds but its proof would be more complicated, making use of martingale arguments (Azuma instead of Hoeffding bound; see e.g., the classical textbook on probabilistic analysis by Mitzenmacher and Upfal [2005]) in order to handle subtle dependencies between the random variables involved due to the balancedness restriction.

We conclude this section by presenting experimental results with randomized and balanced approval voting on 50 alternatives. We have used profiles with homogeneous Mallows agents with three different parameter values, namely $p = 0.95, 0.75$, and 0.6 . The results are depicted in Figure 1

¹Our intuition suggests that uniform selection of the number of alternatives to be approved is best possible. We do not have a formal proof of this though.

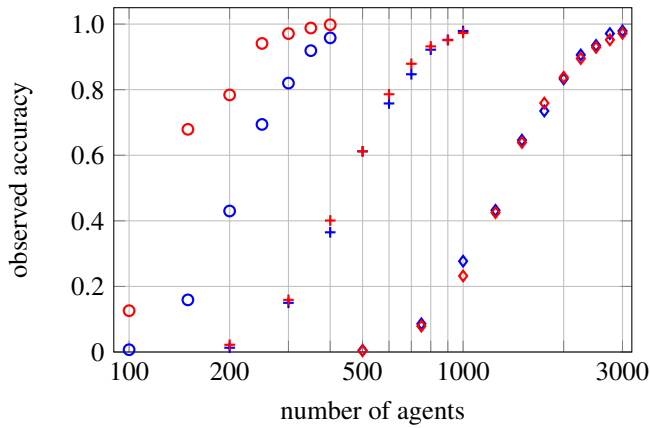


Figure 1: Observed accuracy of randomized approval (blue points) with Mallows agents. Number of agents range from 100 to 400 for agents with $p = 0.95$ (o), from 200 to 1000 for $p = 0.75$ (+), and from 500 to 3000 for $p = 0.6$ (\diamond). Observed accuracy is measured as the frequency of correctly recovering the ground truth among 1000 simulations. Ties are broken at random. The red points show that corresponding results for balanced approval voting.

and show a threshold phenomenon: as the number of agents increases in a short range of values, the accuracy jumps very steeply from 0 to almost 1. The results indicate that randomized approval is considerably better in practice compared to our theoretical bound. For example, in order to achieve accuracy of 90%, profiles with 250, 800, and 2250 agents with Mallows parameters 0.95, 0.75, and 0.6, respectively, are sufficient. The theoretical bounds are approximately 37 000, 120 000, and 750 000. Balanced approval voting is considerably more accurate for high values of the Mallows parameter ($p = 0.95$). This is due to the fact that the profiles considered in this scenario have very few agents compared to the number of alternatives and, in contrast to balanced approval, randomized approval leaves many agents groups empty. This advantage vanishes and there is essentially no statistical difference between the accuracy of balanced and randomized approval for higher noise levels (lower values of p).

5 Discussion

Let us attempt to define a seemingly different model that is inspired by Mallows but instead builds k -approval votes directly. In particular, consider the alternative random process $\text{altMallows}_k(\phi, \sigma^*)$ which takes as input the ground truth ranking σ^* among the m alternatives in A and the noise parameter ϕ (and, consequently, $p = \frac{1}{\phi+1}$) and works as follows. Like Mallows, it first decides the relation \succ between every pair of alternatives separately: for every pair of alternatives a and b with $\sigma^*(a) < \sigma^*(b)$, it is decided that $a \succ b$ with probability p and $b \succ a$ with probability $1 - p$. Now, if there is a set of k alternatives S so that each of them beats (according to \succ) all alternatives in $A \setminus S$, the current round is successful and the set S is returned. If no such set exists, the whole process is repeated from scratch with a new round.

Interestingly, this turns out to be *equivalent* to the model for building Mallows approval votes that we have used

throughout the paper. We will present the formal statement shortly (omitting the proof, due to lack of space). Before doing so, let us define the quantity

$$\delta(S, \sigma^*) = \sum_{a \in S} \sum_{b \in A \setminus S} \mathbb{I}\{\sigma^*(a) > \sigma^*(b)\}$$

for a non-empty set of alternatives $S \subseteq A$ of size at most $m - 1$. In a sense, $\delta(S, \sigma^*)$ is a “distance function” between a set of approved alternatives S and the ground truth ranking σ^* . For example, $\delta(S, \sigma^*)$ is equal to 0 if S consists of the top alternatives in σ^* while it takes its maximum value when S consists of alternatives at the bottom of σ^* .

Theorem 3. *Given a ground truth ranking σ^* of the alternatives in A , noise parameter ϕ , and integer $k \in [m - 1]$, the processes $\text{altMallows}_k(\phi, \sigma^*)$ and $\text{top}_k(\text{Mallows}(\phi, \sigma^*))$ return random subsets of A of size k according to the same probability distribution. In particular, the probability that the k -sized set $S \subseteq A$ is returned is proportional to $\phi^{\delta(S, \sigma^*)}$.*

We believe that this alternative view of Mallows approval votes could be useful in future theoretical investigations. Furthermore, it can be used for the implementation of the Mallows approval noise model without building a Mallows ranking first. Compared to the naive implementation of Mallows that mimics its definition (of deciding the pairwise relations between alternatives independently in rounds until these relations induce a ranking), the alternative implementation is clearly faster since it is less restrictive. Actually, it is considerably faster than several implementations in the literature such as the “multistage ranking method” of Fligner and Verducci [1988]. But Mallows has an extremely fast implementation using the “repeated insertion method” of Doignon *et al.* [2004] (see also [Lu and Boutilier, 2014]). In our experiments, we have used this implementation and obtain the approval votes after building Mallows rankings first. Whether the alternative process can yield an even improved implementation is an interesting open problem.

Besides Mallows, there are many other important noise models over rankings that can be combined with approval voting, e.g., see [Critchlow *et al.*, 1991; Lebanon and Lafferty, 2002; Lu and Boutilier, 2014]. What is the sample complexity of randomized approval voting when such models are used to form the internal agent opinions? Another direction for future research is to replace the requirement to learn the ground truth with high probability with the requirement to recover it *approximately*. This objective has been studied very recently in slightly different contexts in [de Weerd *et al.*, 2016] and [Caragiannis *et al.*, 2017]. Adapting the terminology in these papers to our case, the accuracy parameter $1 - \epsilon$ would refer to the expected fraction of pairwise relations of the ground truth ranking that are recovered correctly when applying approval voting. Finally, regardless of whether the objective is to learn the ground truth exactly or approximately, experimentation with real-world data could shed more light on the suitability of (randomized) approval voting in crowdsourcing environments.

References

- [Arrow, 1953] Kenneth J. Arrow. *Social Choice and Individual Values*. Cowles Foundation, 1953.
- [Brams and Fishburn, 2007] Steven J. Brams and Peter C. Fishburn. *Approval Voting*. Springer, 2007.
- [Braverman and Mossel, 2008] Mark Braverman and Elchanan Mossel. Noisy sorting without resampling. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 268–276, 2008.
- [Caragiannis *et al.*, 2016] Ioannis Caragiannis, Ariel D. Procaccia, and Nisarg Shah. When do noisy votes reveal the truth? *ACM Transactions on Economics and Computation*, 4(3):15, 2016.
- [Caragiannis *et al.*, 2017] Ioannis Caragiannis, Xenophon Chatzigeorgiou, George A. Krimpas, and Alexandros A. Voudouris. Optimizing positional scoring rules for rank aggregation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 2017.
- [Chierichetti and Kleinberg, 2014] Flavio Chierichetti and Jon M. Kleinberg. Voting with limited information and many alternatives. *SIAM Journal on Computing*, 43(5):1615–1653, 2014.
- [Chierichetti *et al.*, 2014] Flavio Chierichetti, Anirban Dasgupta, Ravi Kumar, and Silvio Lattanzi. On reconstructing a hidden permutation. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM*, pages 604–617, 2014.
- [Conitzer and Sandholm, 2005] Vincent Conitzer and Thomas Sandholm. Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 145–152, 2005.
- [Critchlow *et al.*, 1991] Douglas E. Critchlow, Michael A. Fligner, and Joseph S. Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, 35(3):294–318, 1991.
- [de Weerd *et al.*, 2016] Mathijs M. de Weerd, Emrico H. Gerding, and Sebastian Stein. Minimising the rank aggregation error. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 1375–1376, 2016.
- [Dey and Bhattacharyya, 2015] Palash Dey and Arnab Bhattacharyya. Sample complexity for winner prediction in elections. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1421–1430, 2015.
- [Doignon *et al.*, 2004] Jean-Paul Doignon, Aleksandar Pakeč, and Michel Regenwetter. The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika*, 69(1):33–54, 2004.
- [Fligner and Verducci, 1988] Michael A. Fligner and Joseph S. Verducci. Multi-stage ranking models. *Journal of the American Statistical Association*, 83:892–901, 1988.
- [Gibbard, 1973] Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41:587–601, 1973.
- [Hoeffding, 1963] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
- [Laslier and Sanver, 2010] Jean-Francois Laslier and M. Remzi Sanver, editors. *Handbook on Approval Voting*. Springer, 2010.
- [Lebanon and Lafferty, 2002] Guy Lebanon and John Lafferty. Cranking: Combining rankings using conditional probability models on permutations. In *Proceedings of the 19th International Conference on Machine Learning (ICML)*, pages 363–370, 2002.
- [Lee *et al.*, 2014] Jeehyung Lee, Wipapat Kladwang, Minjae Lee, Daniel Cantu, Martin Azizyan, Hanjoo Kim, Alex Limpaecher, Snehal Gaikwad, Sungroh Yoon, Adrien Treuille, Rhiju Das, and EteRNA participants. RNA design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences*, 111(6):2122–2127, 2014.
- [Lu and Boutilier, 2014] Tyler Lu and Craig Boutilier. Effective sampling and learning for Mallows models with pairwise-preference data. *Journal of Machine Learning Research*, 15:3963–4009, 2014.
- [Mallows, 1957] C. L. Mallows. Non-null ranking models. *Biometrika*, 44:114–130, 1957.
- [Mitzenmacher and Upfal, 2005] Michael Mitzenmacher and Eli Upfal. *Probability and Computing – Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.
- [Procaccia and Shah, 2015] Ariel D. Procaccia and Nisarg Shah. Is approval voting optimal given approval votes? In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems*, pages 1801–1809, 2015.
- [Rubinstein and Vardi, 2017] Aviad Rubinstein and Shai Vardi. Sorting from noisier samples. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 960–972, 2017.
- [Satterthwaite, 1975] Mark A. Satterthwaite. Strategy proofness and Arrow’s conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.
- [Shah *et al.*, 2015] Nihar B. Shah, Dengyong Zhou, and Yuval Peres. Approval voting and incentives in crowdsourcing. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 10–19, 2015.
- [Young, 1988] H. Peyton Young. Condorcet’s theory of voting. *American Political Science Review*, 82(4):1231–1244, 1988.