

Truth-tracking via Approval Voting: Size Matters

Tahar Allouche,¹ Jérôme Lang,¹ Florian Yger¹

¹ LMSADE, CNRS, PSL, Université Paris-Dauphine

tahar.allouche@dauphine.eu, lang@lmsade.dauphine.fr, florian.yger@lmsade.dauphine.fr

Abstract

Epistemic social choice aims at unveiling a hidden ground truth given votes, which are interpreted as noisy signals about it. We consider here a simple setting where votes consist of approval ballots: each voter approves a set of alternatives which they believe can possibly be the ground truth. Based on the intuitive idea that more reliable votes contain fewer alternatives, we define several noise models that are approval voting variants of the Mallows model. The likelihood-maximizing alternative is then characterized as the winner of a weighted approval rule, where the weight of a ballot decreases with its cardinality. We have conducted an experiment on three image annotation datasets; they conclude that rules based on our noise model outperform standard approval voting; the best performance is obtained by a variant of the Condorcet noise model.

1 Introduction

Epistemic social choice deals with the problem of unveiling a hidden *ground truth* state from a set of some possible states, given the reports of some voters. Votes are seen as *noisy* reports on the ground truth. The distribution of these reports is modelled by a noise model, sometimes tuned by some parameter reflecting the competence (expertise, reliability) of these voters.

The space of frameworks for epistemic social choice varies along two dimensions: the nature of the ground truth and the format of the reports (ballots expressed by voters). Depending on the framework chosen, the ground truth may be a single alternative, a set of alternative or a ranking over alternatives. *We assume the simplest form of ground truth: it is a single alternative* (the unique correct answer). Still depending on the framework, ballots may also contain a single alternative, a set of alternatives, or a ranking over alternatives. *We assume that they are subsets of alternatives, that is, approval ballots.* Requiring voters to give only one answer (that is, a single alternative) is often too constraining because voters may be uncertain and believe that several alternatives may possibly be the ground truth. This is the path followed by (Procaccia and Shah 2015; Shah, Zhou, and Peres 2015; Caragiannis and Micha 2017).

Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

While some classical anonymous rules have been shown to be optimal under some assumptions, the aggregation rule may, when possible, assign different weights to the voters according to their expertise. Whilst this is doable if we have additional information about voter expertise or when we keep a record of their answers to different questions, estimating the individual expertise gets complicated when we have no prior information about voters and when the sole information we have are votes on a single issue. This leads to the *single-question wisdom of the crowd problem* for which the seminal work (Prelec, Seung, and McCoy 2017) proposes a novel solution, namely selecting the alternative which is *surprisingly popular*. Although it proved to be an efficient way to get around the problem of estimating the voters' reliabilities, its major drawback is that it requires the elicitation of further information: each voter has to report her answer and her beliefs about the answers of the remaining voters.

Now we suggest that there is an alternative approach that does not require this surplus of information and that simply relies on the truthfulness of voters. (Shah, Zhou, and Peres 2015) have defined a proper mechanism to incentivize the participants to select an alternative if and only if they believe it can be the winning one. An intuitive idea might be to consider that *smaller ballots, i.e. answers that contain less alternatives, are more reliable*: a voter who knows the true answer (or, more generally, who believes to know it) will probably select only one alternative and a voter who selects all alternatives has probably no idea whatsoever of the correct answer. For instance, if voters hear a speech and are asked to detect the language in which it is spoken, we may give more weight to a voter approving *Arabic* and *Hebrew* than to one approving *Arabic*, *Hebrew*, *Persian* and *Turkish*.

Based on this intuition, more weight must be assigned to smaller ballots. Rules that work this way, which we call *size-decreasing approval rules*, are part of the family of *size approval rules* (Alcalde-Unzu and Vorsatz 2009). Our goal is to motivate the use of such rules from an epistemic social choice point of view. To this purpose, we will study a family of noise models which are approval voting variants of the Mallows model, and prove that in many cases the optimal rule is *size-decreasing*.

The paper is structured as follows. In Section 2 we discuss related work. In Section 3 we define the framework and the family of noise models we consider. Section 4 charac-

terizes all anonymous noises whose associated optimal rule is size-decreasing. In Section 5, we consider a more general noise where voters have different noise parameters, prove that under some mild assumption, *the expected number of selected alternatives grows when the voter is less reliable*, and then give an explicit expression for the expected size of the ballot as a function of the reliability parameter of a voter for a Condorcet-like noise model. Section 6 focuses on real datasets on which first we test the hypothesis that smaller ballots are more reliable then we apply different size-decreasing rules associated to various noise models to assess their performances. Section 7 concludes.

2 Related Work

Epistemic social choice It studies how a ground truth can be recovered from noisy votes, viewing voting rules as maximum likelihood estimators. It dates back from (Condorcet 1785) and has lead to a lot of developments in the last 30 years. Condorcet’s celebrated *jury theorem* considers n independent, equally reliable voters and two alternatives that are *a priori* equally likely, and states that if every voter votes for the correct alternative with probability $p > \frac{1}{2}$, then the majority rule outputs the correct decision with a probability that increases with the number of voters and tends to 1 when the latter grows to infinity. See (Nitzan and Paroush 2017) and (Dietrich 2008) for proofs and discussion.

The framework was later generalized to more than two alternatives (Young 1988), to voters with different competences (Shapley and Grofman 1984; Drissi-Bakhkhat and Truchon 2004), to a nonuniform prior over alternatives (Ben-Yashar and Nitzan 1997; Ben-Yashar and Paroush 2001), to various noise models (Conitzer and Sandholm 2005; Conitzer, Rognlie, and Xia 2009), to correlated votes (Pivato 2013, 2017), to multi-issue domains (Xia, Conitzer, and Lang 2010) and to multi-winner voting (Caragiannis et al. 2020). Meir et al. (2019) define a method to aggregate votes weighted according to their average proximity to the other votes as an estimation of their reliability. A review of the field can be found in (Elkind and Slinko 2016).

Epistemic voting with approval ballots has scarcely been considered. (Procaccia and Shah 2015) study noise models for which approval voting is optimal given k -approval votes, in the sense that the objectively best alternative gets elected, the ground truth being a ranking over all alternatives. (Caragiannis and Micha 2017) prove that the number of samples needed to recover the ground truth ranking over alternatives with high enough probability from approval ballots is exponential if ballots are required to approve k candidates, but polynomial if the size of the ballots is randomized. (Everaere, Konieczny, and Marquis 2010) studies approval voting within the broader framework of belief merging.

Crowdsourcing (Kruger et al. 2014; Qing et al. 2014) give a social choice-theoretic study of collective annotation tasks. (Shah and Zhou 2020) design mechanisms for incentive-compatible elicitation with approval ballots in crowdsourcing applications. (Prelec, Seung, and McCoy 2017) introduce the *surprisingly popular* approach to solve the single-question problem. The approach was successfully general-

ized to the case where the ground truth is a ranking (Hosseini et al. 2021).

Beyond social choice, collective annotation has also been studied in the machine learning community. (Dawid and Skene 1979) used an expectation-maximization (EM) approach for retrieving true binary labels. This approach has been improved along with other methods namely in (Raykar et al. 2010; Welinder et al. 2010; Ronald and Combes 2017; Tao et al. 2018).

3 Framework

Consider a set N of n voters and a set of $m \geq 2$ alternatives $\mathcal{X} = \{a_1, \dots, a_m\}$. The (hidden) ground truth consists of a single alternative a^* . Voters cast approval ballots $A_i \subseteq \mathcal{X}$ consisting of their noisy estimates of the ground truth. Voters who approve no alternative or all alternatives do not bring any meaningful information, therefore without loss of generality, we assume that for all i , $A_i \neq \emptyset$ and $A_i \neq \mathcal{X}$.

All along this paper, we will model the distribution of these approval ballots by approval voting variants of the *Mallows* noise model. The Mallows distribution was originally defined on *rankings*: we adapt it to subsets of alternatives, keeping the idea that the probability of a subset decreases as its distance from a central point increases, the dispersion being modelled by a parameter ϕ .

In general, we will call an approval Mallows noise model any model where voters’ ballots are independent (we keep this hypothesis all along the paper) and there exist n parameters $\phi_i \in (0, +\infty)$ and a function $d : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \mapsto \mathbb{R}$ such that and for any voter $i \in N$:

$$P_{\phi_i, d}(A_i | a^* = a) = \frac{1}{\beta_i} \phi_i^{d(a^*, A_i)}, \forall a \in \mathcal{X}$$

where β_i is the corresponding normalization factor. If $\phi_i = \phi$ for all $i \in N$, we say the model is *anonymous*.

In the remaining of the paper we will only focus on *neutral* noise models. The neutrality of a noise is defined as its invariance by any permutation of the alternatives:

$$\forall \pi \in \sigma(\mathcal{X}), P_{\phi, d}(A | a^* = a) = P_{\phi, d}(\pi(A) | a^* = \pi(a))$$

We can immediately see that a Mallows noise is neutral if and only if its associated function d is neutral (invariant by a permutation of the alternatives).

A noise model is neutral if $d(a, A)$ depends only on $|a \cap A|$ (that is, 1 if $a \in A$ and 0 if $a \notin A$)¹ and $|A|$:

Proposition 1. *A noise model associated to a function d is neutral if and only if there exists a unique function $\psi_d : \{0, 1\} \times \{0, \dots, m\} \setminus (1, 0) \mapsto \mathbb{R}$, such that $d(a, A) = \psi_d(|a \cap A|, |A|)$.*²

Proof. If $d(a, A) = \psi_d(|a \cap A|, |A|)$ then since ψ_d is neutral, d is neutral. Conversely, assume d is neutral. We claim that for any two pairs (a, A) and (b, B) such that $(|a \cap A|, |A|) = (|b \cap B|, |B|)$ we have $d(a, A) = d(b, B)$. Assume first $a \in A$ (and therefore, $b \in B$). Consider a

¹We omit the curly brackets and write $a \cap A$ for $\{a\} \cap A$.

² $(1, 0)$ is excluded from the domain of ψ_d simply because $(|A| = 0 \text{ and } |a \cap A| = 1)$ is impossible.

permutation π such that $\pi(a) = b$ and $\pi(A) = B$. Then $d(b, B) = d(\pi(a), \pi(A)) = d(a, A)$. The argument for the case $a \notin A$ (and $b \notin B$) is similar. Thus, $d(a, A)$ depends only on $|a \cap A|$ and $|A|$, which means that there is a function $\psi_d : \{0, 1\} \times \{0, \dots, m\} \setminus (1, 0) \mapsto \mathbb{R}$ such that $d(a, A) = \psi_d(|a \cap A|, |A|)$. Uniqueness is immediate. \square

Example 1. For the Hamming distance we have that:

$$d_H(a, A) = |\bar{a} \cap A| + |a \cap \bar{A}| = 1 - 2|a \cap A| + |A|$$

so $d_H(a, A) = \psi_{d_H}(|a \cap A|, |A|)$, where $\psi_{d_H}(t, j) = 1 - 2t + j$. $d_H(a, A)$ takes its minimal value 0 for $A = \{a\}$ and its maximal value m for $A = \mathcal{X} \setminus \{a\}$.

For the Jaccard distance (Jaccard 1901) we have

$$d_J(a, A) = \frac{|\bar{a} \cap A| + |a \cap \bar{A}|}{|\bar{a} \cap A| + |a \cap \bar{A}| + |a \cap A|}$$

so $d_J(a, A) = \psi_{d_J}(|a \cap A|, |A|)$ where $\psi_{d_J}(t, j) = \frac{t}{1-t+j}$. $d_J(a, A)$ takes (again) its minimal value 0 for $A = \{a\}$ and its maximal value 1 for $A = \mathcal{X} \setminus \{a\}$.

4 Anonymous Noise and Size-decreasing Approval Rules

In this section, we suppose that voters share a common (unknown) noise parameter $\phi \in (0, 1)$ and that there exists some function $d : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \mapsto \mathbb{R}$ and its associated function ψ_d such that, for any $a \in \mathcal{X}$:

$$P_{\phi, d}(A_i | a^* = a) = \frac{1}{\beta} \phi^{d(a^*, A_i)} = \frac{1}{\beta} \phi^{\psi_d(|a^* \cap A_i|, |A_i|)}$$

After formally defining the notion of size-decreasing rules, we state the main result of this section which characterizes all the Mallows anonymous noises (that is, all functions d) whose associated maximum likelihood rule is size-decreasing. We will see that this is the case for some natural functions d , that we will test later on in the experiments.

Definition 1 (Size Approval Rule). Consider a function

$$v : \begin{array}{ccc} \mathcal{P}(\mathcal{X})^n & \longrightarrow & \mathcal{X} \\ (A_1, \dots, A_n) & \mapsto & v(A_1, \dots, A_n) \end{array}$$

that, for each approval profile $A = (A_1, \dots, A_n)$, assigns a winning alternative $v(A_1, \dots, A_n) \in \mathcal{X}$. We say that v is a size approval rule if there exists a vector $w = (w_0, \dots, w_m) \in \mathbb{R}^{m+1}$ such that:

$$v(A_1, \dots, A_n) = \arg \max_{a \in \mathcal{X}} app_w(a)$$

where app_w is the weighted approval score defined by:

$$app_w(a) = \sum_{i:a \in A_i} w_{|A_i|}$$

A size approval rule v is size-decreasing if its associated vector $w = (w_0, \dots, w_m) \in \mathbb{R}^{m+1}$ is such that $w_j > w_{j+1}$ for all $1 \leq j \leq m-1$.

Example 2. The size approval rule associated to the vector of weights given by $w_j = n^{m-j}$ is size-decreasing in the most extreme sense, as it is lexicographic: it outputs the alternative which appears most often in singleton ballots, in case of ties it considers ballots of size 2 and so on.

Definition 2 (Maximum Likelihood Estimation Rule). We define the function:

$$\zeta_d : \begin{array}{ccc} \mathcal{P}(\mathcal{X})^n & \longrightarrow & \mathcal{X} \\ (A_1, \dots, A_n) & \mapsto & \arg \max_{a \in \mathcal{X}} P_d(A_1, \dots, A_n | a^* = a) \end{array}$$

which given an approval profile outputs the maximum likelihood estimator of the ground truth alternative.

The next theorem aims to characterize the functions d for which the maximum likelihood estimation rule ζ_d is a size-decreasing approval rule.

Theorem 2. For $n \geq 3$, the maximum likelihood estimation rule ζ_d is a size-decreasing approval rule if and only if $\Delta \psi_d : j \mapsto \psi_d(0, j) - \psi_d(1, j)$ is decreasing.

Proof. First, for any approval profile $A = (A_1, \dots, A_n)$,

$$\begin{aligned} \zeta_d(A) &= \arg \max_{a \in \mathcal{X}} \prod_{i=1}^n \frac{1}{\beta} \phi^{d(a, A_i)} = \arg \min_{a \in \mathcal{X}} \sum_{i=1}^n d(a, A_i) \\ &= \arg \min_{a \in \mathcal{X}} \sum_{i=1}^n \psi_d(|a \cap A_i|, |A_i|) \\ &= \arg \min_{a \in \mathcal{X}} \sum_{i:a \in A_i} \psi_d(1, |A_i|) + \sum_{i:a \notin A_i} \psi_d(0, |A_i|) \\ &= \arg \min_{a \in \mathcal{X}} \underbrace{\sum_{i=1}^n \psi_d(0, |A_i|)}_{\text{constant}} - \underbrace{\sum_{i:a \in A_i} \psi_d(0, |A_i|) - \psi_d(1, |A_i|)}_{\Delta \psi_d(|A_i|)} \\ &= \arg \max_{a \in \mathcal{X}} \sum_{i:a \in A_i} \Delta \psi_d(|A_i|) \end{aligned}$$

\Leftarrow : If $\Delta \psi_d$ is decreasing then we can immediately prove that ζ_d is a size-decreasing approval rule with a weight vector w such that $w_j = \Delta \psi_d(j)$ for any $1 \leq j \leq m-1$.

\Rightarrow : Suppose that ζ_d is a size-decreasing approval rule. Thus, there exists a weight vector $w = (w_0, \dots, w_m) \in \mathbb{R}^{m+1}$ such that $w_j > w_{j+1}$ for $1 \leq j \leq m-2$ and for any approval profile $A = (A_1, \dots, A_n)$ we have:

$$\zeta_d(A) = \arg \max_{a \in \mathcal{X}} \sum_{i:a \in A_i} w_{|A_i|}$$

If $m = 3$: Let $\mathcal{X} = \{a, b, c\}$. To prove that $\Delta \psi_d(1) > \Delta \psi_d(2)$ consider the following approval profile:

$$\begin{cases} A_1 = \{a\} \\ A_i = \{a, b\} & , \forall i \in \{2, \dots, n-1\} \\ A_n = \{b, c\} \end{cases}$$

Which yields the following weighted approval scores:

$$\begin{cases} app_w(a) = w_1 + (n-2)w_2 \\ app_w(b) = (n-1)w_2 \\ app_w(c) = w_2 \end{cases}$$

Since $w_1 > w_2$ we have $\zeta_d(A) = a$, which implies that $\arg \max_{a \in \mathcal{X}} \sum_{i:a \in A_i} \Delta \psi_d(|A_i|) = a$. In particular:

$$\sum_{i:a \in A_i} \Delta \psi_d(|A_i|) > \sum_{i:b \in A_i} \Delta \psi_d(|A_i|)$$

So

$$\Delta\psi_d(1) + (n-2)\Delta\psi_d(2) > (n-1)\Delta\psi_d(2)$$

which implies that $\Delta\psi_d(1) > \Delta\psi_d(2)$.

If $m > 3$: Let $\mathcal{X} = \{a, b, c, e_1, \dots, e_{m-3}\}$. To prove that $\Delta\psi_d(1) > \Delta\psi_d(2)$ we use the same approval profile as above. To prove that $\Delta\psi_d(j) > \Delta\psi_d(j+1)$ for $j \geq 2$, consider the following approval profile:

$$\begin{cases} A_1 = \{a, e_1, \dots, e_{j-1}\} \\ A_i = \{a, b\} \\ A_n = \{b, c, e_1, \dots, e_{j-1}\} \end{cases}, \forall i \in \{2, \dots, n-1\}$$

Which yields the following weighted approval scores:

$$\begin{cases} app_w(a) = w_j + (n-2)w_2 \\ app_w(b) = w_{j+1} + (n-2)w_2 \\ app_w(c) = w_{j+1} \\ app_w(e_l) = w_j + w_{j+1} \end{cases} \quad \forall l \in \{1, \dots, j-1\}$$

Since $w_2 > w_j > w_{j+1}$ we have $\zeta_d(A) = a$, which implies that $\arg \max_{e \in \mathcal{X}} \sum_{i: e \in A_i} \Delta\psi_d(|A_i|) = a$. In particular, we have

$$\sum_{i: a \in A_i} \Delta\psi_d(|A_i|) > \sum_{i: b \in A_i} \Delta\psi_d(|A_i|)$$

So:

$$\Delta\psi_d(j) + (n-2)\Delta\psi_d(2) > \Delta\psi_d(j+1) + (n-2)\Delta\psi_d(2)$$

which implies that:

$$\Delta\psi_d(j) > \Delta\psi_d(j+1)$$

□

Interpretation: Consider an anonymous noise $P_{\phi,d}$, where d is such that $\Delta\psi_d$ is decreasing. Now consider any alternative $a \in \mathcal{X}$, and for any $k \in [1, m-2]$, let $A_k, A_{k+1}, B_k, B_{k+1}$ be four sets such that $a \in A_k \cap A_{k+1}$ and $a \notin B_k \cup B_{k+1}$ and $|A_k| = |B_k| = k$ and $|A_{k+1}| = |B_{k+1}| = k+1$. We can easily check that since $\phi \in (0, 1)$ and $d(a, B_k) - d(a, A_k) < d(a, B_{k+1}) - d(a, A_{k+1})$, we have

$$\frac{P_{\phi,d}(B_k|a^* = a)}{P_{\phi,d}(A_k|a^* = a)} < \frac{P_{\phi,d}(B_{k+1}|a^* = a)}{P_{\phi,d}(A_{k+1}|a^* = a)}$$

which implies the following:

- If it is more likely that a voter casts a k -sized ballot not containing the ground truth than a k -sized ballot that contains it, then it is even more likely that she casts a $(k+1)$ -sized ballot not containing the ground truth than a $(k+1)$ -sized ballot that contains it.
- If it is more likely that a voter casts a $(k+1)$ -sized ballot containing the ground truth than a $(k+1)$ -sized ballot that does not, then it is even more likely that she casts a k -sized ballot containing the ground truth than a k -sized ballot that does not.

We now give some examples with some usual functions d . We will see that the maximum likelihood estimation rule associated to the Jaccard distance is size-decreasing with weights $w_{|A|} = \frac{1}{|A|}$, and that the maximum likelihood estimation rule associated to the Hamming distance is not size-decreasing.

Example 3. Consider the Jaccard distance given by:

$$d_J(a, A) = \psi_{d_J}(|a \cap A|, |A|) = 1 - \frac{|a \cap A|}{|A| - |a \cap A| + 1}$$

which gives:

$$\Delta\psi_{d_J}(j) = \psi_{d_J}(0, j) - \psi_{d_J}(1, j) = 1/j$$

By Theorem 2, we conclude that the maximum likelihood estimation rule ζ_{d_J} is a size-decreasing approval rule with weights $w_j = 1/j$.

Example 4. Consider the Hamming distance given by:

$$d_H(a, A) = |a \cap \bar{A}| + |\bar{a} \cap A| = 1 + |A| - 2|a \cap A|$$

Which gives us that:

$$\Delta\psi_{d_H}(j) = \psi_{d_H}(0, j) - \psi_{d_H}(1, j) = 2$$

Therefore, the maximum likelihood estimation rule ζ_{d_H} is a size approval rule with constant weights: it is the standard approval rule (SAV), that selects the alternative with the maximum number of approvals. It can be seen immediately that SAV is not size-decreasing; however, it is, so to say, size-non-increasing, and thus can be seen as the limit of size-decreasing rules.

As a consequence of Theorem 2, we can easily prove that, for an anonymous noise, the maximum likelihood estimation rule associated to a function d defined as a linear combination of the quantities $|a \cap A|$ and $|A|$ is not a size-decreasing rule (this is the case for the Hamming distance). More generally, this applies to any function d such that ψ_d can be additively separated into two terms $\psi_d(|a \cap A|, |A|) = f(|a \cap A|) + g(|A|)$. In the next section, we will consider this particular family of separable functions with a non-anonymous noise, where each voter has her own noise parameter ϕ_i .

5 Non-anonymous Separable Noise

5.1 The General Case

Consider a set of m alternatives $\mathcal{X} = \{a_1, \dots, a_m\}$ and a ground truth answer a^* . Consider also a neutral function $d : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \rightarrow \mathbb{R}$ with an associated function $\psi_d : \{0, 1\} \times \{0, \dots, m\} \setminus (1, 0) \mapsto \mathbb{R}$ which can be separated into two quantities:

$$\psi_d(|a \cap A|, |A|) = f(|a \cap A|) + g(|A|)$$

We define a non-anonymous Mallows noise model, where for each voter $i \in N$ there exists a parameter $\phi_i \in]0, +\infty[$ such that, for any $a \in \mathcal{X}$:

$$P_{\phi_i, d}(A_i|a^* = a) = \frac{1}{\beta_i} \phi_i^{f(|a^* \cap A_i|) + g(|A_i|)}$$

Notice that in this case, a bigger individual noise parameter ϕ_i models a less reliable voter (her distribution is less condensed around the ground truth). The aim of the next result is

³Recall that voters cannot cast empty or full approval ballots. Therefore we suppose that $P(\emptyset|a^* = a) = P(\mathcal{X}|a^* = a) = 0$.

to motivate the use of size-decreasing approval rules to aggregate approvals generated from such distributions. More precisely, the goal is to find sufficient conditions on f and g that makes the expected size of the voter's ballot $\mathbb{E}_{\phi,d}[|A_i|]$ grows as the voter become less reliable (*i.e.* as her noise parameter ϕ_i grows).

We will denote $f(1) = f_1, f(0) = f_0$ and $\Delta f = f_0 - f_1$ which would naturally be positive $\Delta f > 0$. We will also denote by $\Delta g_{k,t} = g(k) - g(t)$.

Theorem 3. *If for every $1 \leq t < k \leq m-1$ we have that:*

$$\Delta g_{k,t} \geq \frac{k-t}{2} \Delta f$$

Then:

$$\frac{\partial \mathbb{E}_{\phi,d}[|A_i|]}{\partial \phi} \geq 0$$

Proof. We will just give a sketch of the proof. The full version is included in the appendix. For any prior distribution on a^* , it suffices to prove that for any $a \in \mathcal{X}$ we have that:

$$\frac{\partial \mathbb{E}_{\phi,d}[|A_i||a^* = a]}{\partial \phi} \geq 0$$

since $\mathbb{E}_{\phi,d}[|A_i|] = \sum_{a \in \mathcal{X}} \mathbb{E}_{\phi,d}[|A_i||a^* = a]P(a^* = a)$.

Let $a \in \mathcal{X}$. First we can compute the normalization factor:

$$\beta = \sum_{t=1}^{m-1} \binom{m-1}{t-1} \phi^{f_1+g(t)} + \binom{m-1}{t} \phi^{f_0+g(t)} = \sum_{t=1}^{m-1} h_t(\phi)$$

where $h_t(\phi) = \binom{m-1}{t-1} \phi^{f_1+g(t)} + \binom{m-1}{t} \phi^{f_0+g(t)}$. So:

$$\mathbb{E}_{\phi,d}[|A_i||a^* = a] = \sum_{t=1}^{m-1} t P(|A_i| = t | a^* = a) = \frac{\sum_{t=1}^{m-1} t \times h_t(\phi)}{\sum_{t=1}^{m-1} h_t(\phi)}$$

Thus the derivative reads:

$$\begin{aligned} \frac{\partial \mathbb{E}_{\phi,d}[|A_i||a]}{\partial \phi} &= \frac{\sum_{t=1}^{m-1} t h'_t(\phi) \sum_{k=1}^{m-1} h_k(\phi) - \sum_{t=1}^{m-1} t h_t(\phi) \sum_{k=1}^{m-1} h'_k(\phi)}{\left(\sum_{t=1}^{m-1} h_t(\phi)\right)^2} \\ &\propto \sum_{t=1}^{m-2} \sum_{k=t+1}^{m-1} (k-t) [h'_k(\phi)h_t(\phi) - h'_t(\phi)h_k(\phi)] \end{aligned}$$

We can already notice that to guarantee that $\frac{\partial \mathbb{E}_{\phi,d}[|A_i||a^* = a]}{\partial \phi} \geq 0$ it suffices that $\Delta h_{k,t}(\phi) = [h'_k(\phi)h_t(\phi) - h'_t(\phi)h_k(\phi)] \geq 0$ for all $1 \leq t < k \leq m-1$.

We have that:

$$\begin{aligned} h'_k(\phi)h_t(\phi) &= \left[(g(k) + f_1) \binom{m-1}{k-1} \phi^{g(k)+f_1-1} \right. \\ &\quad \left. + (g(k) + f_0) \binom{m-1}{k} \phi^{g(k)+f_0-1} \right] \\ &\times \left[\binom{m-1}{t-1} \phi^{f_1+g(t)} + \binom{m-1}{t} \phi^{f_0+g(t)} \right] \end{aligned}$$

So we can check that:

$$\Delta h_{k,t}(\phi) \propto \left[r(\phi) + \frac{m-k}{k} + \frac{m-t}{t} \right] \Delta g_{k,t} - \left(\frac{m-t}{t} - \frac{m-k}{k} \right) \Delta f$$

where $r(\phi) = \phi^{-\Delta f} + \frac{m-k}{k} \frac{m-t}{t} \phi^{\Delta f}$. By studying the derivative of r we get that:

$$r(\phi) \geq 2 \sqrt{\frac{m-k}{k}} \sqrt{\frac{m-t}{t}}$$

So in order to have $\Delta h_{k,t}(\phi) \geq 0$ it suffice that:

$$\left[\sqrt{\frac{m-k}{k}} + \sqrt{\frac{m-t}{t}} \right]^2 \Delta g_{k,t} - \left(\frac{m-t}{t} - \frac{m-k}{k} \right) \Delta f \geq 0$$

which is equivalent to:

$$\left[\sqrt{\frac{m-k}{k}} + \sqrt{\frac{m-t}{t}} \right] \Delta g_{k,t} - \left[\sqrt{\frac{m-t}{t}} - \sqrt{\frac{m-k}{k}} \right] \Delta f \geq 0$$

We prove that $\frac{\sqrt{k(m-t)} - \sqrt{t(m-k)}}{\sqrt{k(m-t)} + \sqrt{t(m-k)}} \leq \frac{k-t}{2}$ to recover the sufficient condition. \square

Example 5. For $\alpha, \beta > 0$ define the distance:

$$\begin{aligned} d_{\alpha,\beta}(a, A) &= \alpha |\bar{a} \cap A| + \beta |a \cap \bar{A}| \\ &= -(\alpha + \beta) |a \cap A| + \beta + \alpha |A| \end{aligned}$$

which generalizes the Hamming distance in the same way the Tversky index (Tversky 1977) generalizes Jaccard's. $d_{\alpha,\beta}$ is associated to the separable function:

$$y_{d_{\alpha,\beta}}(x, k) = -(\alpha + \beta)x + \beta + \alpha k = f(x) + g(k)$$

where $f(x) = -(\alpha + \beta)x + \beta$ and $g(k) = \alpha k$. We have:

$$\Delta g_{k,t} = \frac{\alpha}{\alpha + \beta} (k-t) \Delta f$$

So for every $d_{\alpha,\beta}$ such that $\alpha \geq \beta$ we have that:

$$\frac{\partial \mathbb{E}_{\phi,d_{\alpha,\beta}}[|A_i|]}{\partial \phi} \geq 0$$

5.2 The Hamming Distance Case - Condorcet Noise Model

The prototypical example of a separable noise is the noise associated to the Hamming distance, which is equivalent to the Condorcet-like noise model. We will prove that for this specific noise, we can express the expected size of a voter's ballot $\mathbb{E}[|A_i|]$ as a linear function of her reliability parameter. This enables us to estimate this parameter directly from the actual size of the ballot, without any prior belief about the ground truth.

Formally, consider the Condorcet noise model where for each voter $i \in N$ there exists a noise parameter $p_i \in (0, 1)$ such that:

$P_{p_i}(a \in A_i | a = a^*) = P_{p_i}(a \notin A_i | a \neq a^*) = p_i, \forall a \in \mathcal{X}$ and where the belonging or not of different alternatives to the voter's ballot are independent events. Notice that the

model supposes equal error-rates for false positives and false negatives. In particular, voters who select many alternatives would *ipso facto* have a low p_i (since their ballots contain many false positives) which can even be below 0.5. In fact, constraining p_i to be in $(0.5, 1)$ worsened the performance of our method in the experiments. Moreover, we can easily prove that in this case, the noise model is a non-anonymous Mallows noise model with the Hamming distance and with $\phi_i = \frac{1-p_i}{p_i}$ (We can have $\phi_i \geq 1$ since p_i can be below 0.5):

$$P_{p_i}(A_i|a=a^*) = p_i^m \left(\frac{1-p_i}{p_i} \right)^{d_H(a^*, A_i)}, \forall a \in \mathcal{X}$$

We will show that in this particular case, we can give an explicit formula of the expected size of a voter's approval ballot as a linear function of her precision parameter p_i .

Theorem 4. For $m \geq 2$, we have that:

$$\mathbb{E}_p[|A_i|] = (m-1) - (m-2)p$$

Proof. Let $a \in \mathcal{X}$:

$$\begin{aligned} \mathbb{E}_p[|A_i||a^* = a] &= \mathbb{E}\left[\sum_{b \in \mathcal{X}} \mathbf{1}\{b \in A_i\}|a^* = a\right] \\ &= \sum_{b \in \mathcal{X}} P(b \in A_i | a^* = a) \\ &= P(a \in A_i | a^* = a) + \sum_{b \neq a} P(b \in A_i | a^* = a) \\ &= p + (m-1)(1-p) \\ &= (m-1) - (m-2)p \end{aligned}$$

Thus we have that:

$$\begin{aligned} \mathbb{E}_p[|A_i|] &= \sum_{a \in \mathcal{X}} \mathbb{E}_p[|A_i||a^* = a] P(a^* = a) \\ &= \sum_{a \in \mathcal{X}} [(m-1) - (m-2)p] P(a^* = a) \\ &= (m-1) - (m-2)p \end{aligned}$$

□

Theorem 4 gives us a simple approach to estimate p_i by maximum likelihood estimations given some observations of A_i without a need to know the ground truth a^* .

6 Experiments

We took the three image annotation datasets, originally collected in (Shah, Zhou, and Peres 2015) for incentive-design purposes⁴, and used them to test our hypothesis and to assess the accuracy of different aggregation rules of interest.

Each dataset consists of a set of approval profiles of a number of voters (participants) who had to select *all the alternatives that they thought were correct* in a number of instances (images), namely:

- Animal task: 16 images/questions and 6 alternatives.
- Texture task: 16 images/questions and 6 alternatives.
- Language task: 25 images/questions and 8 alternatives.

⁴ Accessible on the author's webpage: https://cs.cmu.edu/~nihars/data/data_approval.zip

From now on, a dataset denotes the set N of n voters, $\mathcal{X} = \{a_1, \dots, a_m\}$ the set of alternatives, L approval profiles $A^z = (A_1^z, \dots, A_n^z)$ each associated to an image z with ground truth alternative a_z^* .

6.1 Ballot Size and Reliability

To test the hypothesis that smaller ballots are more reliable, we introduce the *size-normalized accuracy* which is defined for each dataset and each $k \in [1, m-1]$ as:

$$\text{SNA}(k) = \frac{1}{k} \frac{|\{A_i^z, |A_i^z| = k, a_z^* \in A_i^z\}|}{|\{A_i^z, |A_i^z| = k\}|}$$

It can be interpreted as the probability of recovering the ground truth after drawing randomly (uniformly) an alternative from a ballot of size k . Notice that if smaller ballots were not more reliable, one would expect that, for instance, ballots of size 2 are twice more probable to contain the ground truth than ballots of size 1, so the chance of finding the ground truth after randomly picking an alternative from a 2-sized ballot is equal to the chance that a singleton ballot selects the ground truth. So we would expect that SNA is almost constant for all k .

However when we compute the SNA for the three datasets (see Figure 1) we can clearly see that it decreases for the bigger ballots, which confirms that the alternatives selected in smaller approval ballots are more likely to coincide with the ground truth.

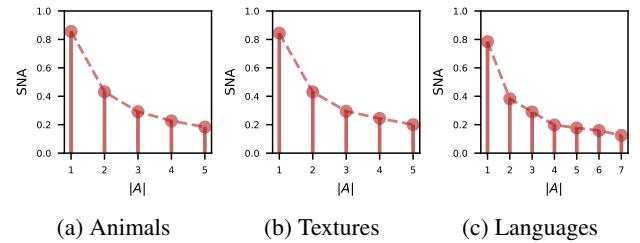


Figure 1: Size-normalized accuracy for three datasets

6.2 Aggregation

Since we are mostly interested in the single-question wisdom of the crowd problem, we will only consider aggregation rules that operate question-wise (voters' answers on different questions do not affect the output of the rule for a given question). We will use the following aggregation methods (we include more methods in the Appendix)⁵:

Condorcet: For each instance with approval profile $A^z = (A_1^z, \dots, A_n^z)$ and ground truth a_z^* , we suppose that each voter has a precision parameter p_i^z such that:

$$P_{p_i^z}(a \in A_i^z | a = a_z^*) = P_{p_i^z}(a \notin A_i^z | a \neq a_z^*) = p_i^z, \forall a \in \mathcal{X}$$

and where the belonging or not of different alternatives to the voter's ballot are independent events. We know that if

⁵The code can be found at https://github.com/taharallouche/Truth_Tracking-via-AV

these parameters were known, then the maximum likelihood estimation rule returns the alternative \hat{a}_z such that:

$$\hat{a}_z = \arg \max_{a \in \mathcal{X}} \sum_{i: a \in A_i^z} \ln \frac{p_i^z}{1 - p_i^z}$$

To estimate the parameters p_i^z , we will make use of the expression in Theorem 4 that states that:

$$\mathbb{E}_{p_i^z}[|A_i^z|] = (m-1) - (m-2)p_i^z$$

and set:

$$\hat{p}_i^z = \text{proj}_{[\varepsilon, 1-\varepsilon]} \frac{m-1-|A_i^z|}{m-2}$$

The projected quantity is simply the maximum likelihood estimation of p_i^z with a single sample (the actual observation of the voter’s ballot). We project it into a closed interval to avoid having $\hat{p}_i^z = 1$ (which yields an infinite weight to the voter) whenever $|A_i^z| = 1$, and $\hat{p}_i^z = 0$ whenever $|A_i^z| = m-1$. So the aggregation rule finally outputs:

$$\hat{a}_z = \arg \max_{a \in \mathcal{X}} \sum_{i: a \in A_i^z} \ln \frac{\hat{p}_i^z}{1 - \hat{p}_i^z}$$

which is size-decreasing.

Jaccard: Here we suppose that for each instance $A^z = (A_1^z, \dots, A_n^z)$, the noise model is as follows:

$$P_{\phi, d_J}(A_i^z | a_z^* = a) = \frac{1}{\beta} \phi^{d_J(a_z^*, A_i^z)}, \forall a \in \mathcal{X}$$

where $d_J(a, A) = 1 - \frac{|a \cap A|}{|A|+1-|a \cap A|}$.

We saw in Example 3 that the maximum likelihood estimation rule ζ_{d_J} is a size approval rule with weights $w_j = 1/j$.

Simple approval: We will compare all these rules to the benchmark SAV rule where for each instance $A^z = (A_1^z, \dots, A_n^z)$: $\hat{a}_z = \arg \max_{a \in \mathcal{X}} |\{A_i^z, a \in A_i^z\}|$

6.3 Results

For each task, we took 25 batches for each different number of voters, and applied the aforementioned rules. We measure the accuracy of each rule, outputting the estimates \hat{a}^z for each instance, defined as

$$\frac{1}{L} \sum_{z=1}^L \mathbb{1}\{a_z^* = \hat{a}_z\}$$

The results are shown in Figures 2a, 2b and 2c respectively for the Animals, Textures and Languages datasets.

Observations: First we notice that for all the three datasets, the aggregation rules associated to Jaccard anonymous noise show slightly better accuracy than the simple approval rule especially for small number of voters.

We can also see that the aggregation rule associated to the non-anonymous Condorcet noise show significant improvement in the accuracy compared to this rule for Animals and Languages (specially for relatively big numbers of voters). However it fails to outperform it for the Textures dataset, where it only shows similar accuracies to the standard rule as the number of voters grows. This can be the result of the poor estimation quality which uses only one sample.

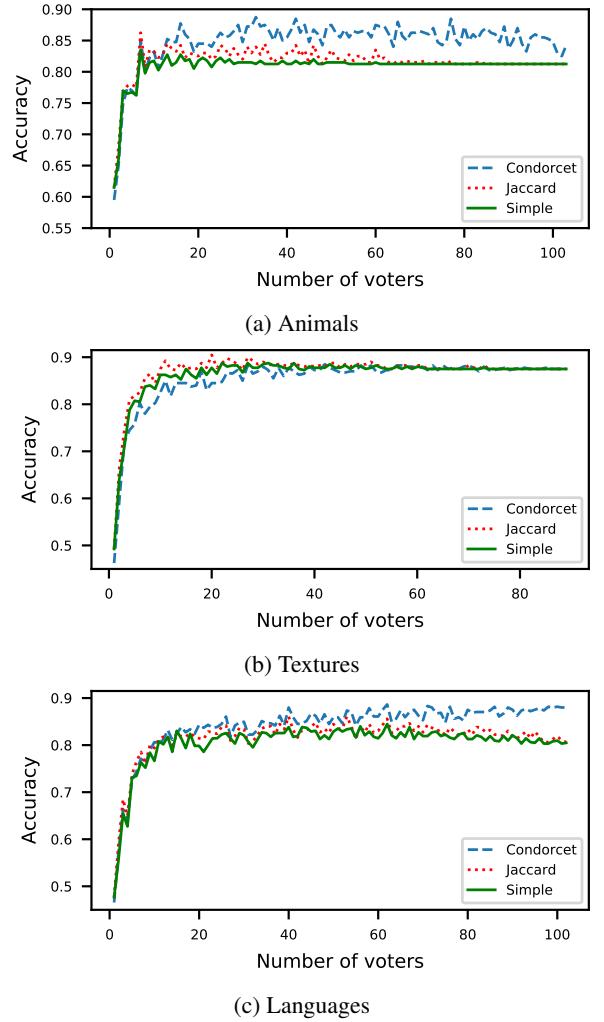


Figure 2: Accuracy of different rules on real-life datasets

7 Conclusion

We propose a novel approach for epistemic approval voting based on the intuition that more reliable votes contain fewer alternatives. First, we show that for different anonymous variants of Mallows-like noise models, the maximum likelihood rule is size-decreasing, *i.e.* it assigns more weight to smaller ballots. Then we consider non-anonymous noises and give a sufficient condition to have an expected size of the ballot which increases as a voter gets less reliable. In particular, we prove that for a Condorcet-like noise, the expected number of approved alternatives decreases linearly with the voter’s precision. Finally, we conduct experiments to test our hypothesis on real data and to assess the performances of different aggregation rules.

These methods may fail in two possible scenarios. First, if the voters do not respond truthfully. In this case, a voter can select a single alternative even though she totally ignores the correct answer. Second, if a large enough group of non-expert voters are mistakenly over-self-confident, whereas the experts are uncertain about their responses.

Acknowledgements

This work was funded in part by the French government under management of Agence Nationale de la Recherche as part of the “Investissements d’avenir” program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute).

References

- Alcalde-Unzu, J.; and Vorsatz, M. 2009. Size approval voting. *J. Econ. Theory*, 144(3): 1187–1210.
- Ben-Yashar, R.; and Paroush, J. 2001. Optimal decision rules for fixed-size committees in polychotomous choice situations. *Soc. Choice Welf.*, 18(4): 737–746.
- Ben-Yashar, R. C.; and Nitzan, S. I. 1997. The Optimal Decision Rule for Fixed-Size Committees in Dichotomous Choice Situations: The General Result. *International Economic Review*.
- Bonald, T.; and Combes, R. 2017. A minimax optimal algorithm for crowdsourcing. In *NIPS*.
- Caragiannis, I.; Kaklamanis, C.; Karanikolas, N.; and Krimpas, G. A. 2020. Evaluating Approval-Based Multiwinner Voting in Terms of Robustness to Noise. In *IJCAI*.
- Caragiannis, I.; and Micha, E. 2017. Learning a Ground Truth Ranking Using Noisy Approval Votes. In *IJCAI*.
- Condorcet. 1785. *Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix*.
- Conitzer, V.; Rognlie, M.; and Xia, L. 2009. Preference Functions that Score Rankings and Maximum Likelihood Estimation. In *IJCAI*.
- Conitzer, V.; and Sandholm, T. 2005. Common Voting Rules as Maximum Likelihood Estimators. In *UAI*.
- Dawid, A. P.; and Skene, A. M. 1979. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Journal of the Royal Statistical Society*.
- Dietrich, F. 2008. The Premises of Condorcet’s Jury Theorem Are Not Simultaneously Justified. *Episteme*.
- Drissi-Bakhkhat, M.; and Truchon, M. 2004. Maximum likelihood approach to vote aggregation with variable probabilities. *Social Choice and Welfare*.
- Elkind, E.; and Slinko, A. 2016. Rationalizations of Voting Rules. In *Handbook of Computational Social Choice*.
- Everaere, P.; Konieczny, S.; and Marquis, P. 2010. The Epistemic View of Belief Merging: Can We Track the Truth? In Coelho, H.; Studer, R.; and Wooldridge, M. J., eds., *ECAI 2010 - 19th European Conference on Artificial Intelligence, Lisbon, Portugal, August 16-20, 2010, Proceedings*, volume 215 of *Frontiers in Artificial Intelligence and Applications*, 621–626. IOS Press.
- Hosseini, H.; Mandal, D.; Shah, N.; and Shi, K. 2021. Surprisingly Popular Voting Recovers Rankings, Surprisingly! In Zhou, Z., ed., *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, 245–251.
- Jaccard, P. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37: 547–579.
- Kruger, J.; Endriss, U.; Fernández, R.; and Qing, C. 2014. Axiomatic analysis of aggregation methods for collective annotation. In *AAMAS*.
- Meir, R.; Amir, O.; Cohensius, G.; Ben-Porat, O.; and Xia, L. 2019. Truth Discovery via Proxy Voting. *arXiv:1905.00629*.
- Nitzan, S.; and Paroush, J. 2017. Collective decision making and jury theorems. *The Oxford Handbook of Law and Economics*, 1.
- Pivato, M. 2013. Voting rules as statistical estimators. *Social Choice and Welfare*.
- Pivato, M. 2017. Epistemic democracy with correlated voters. *Journal of Mathematical Economics*.
- Prelec, D.; Seung, H. S.; and McCoy, J. 2017. A solution to the single-question crowd wisdom problem. *Nature*, 541(7638): 532–535.
- Procaccia, A. D.; and Shah, N. 2015. Is Approval Voting Optimal Given Approval Votes? In *NIPS*.
- Qing, C.; Endriss, U.; Fernández, R.; and Kruger, J. 2014. Empirical Analysis of Aggregation Methods for Collective Annotation. In *COLING*.
- Raykar, V. C.; Yu, S.; Zhao, L. H.; Valadez, G. H.; Florin, C.; Bogoni, L.; and Moy, L. 2010. Learning from crowds. *JMLR*.
- Shah, N.; Zhou, D.; and Peres, Y. 2015. Approval voting and incentives in crowdsourcing. In *ICML*.
- Shah, N. B.; and Zhou, D. 2020. Approval Voting and Incentives in Crowdsourcing. *ACM Transactions on Economics and Computation*.
- Shapley, L.; and Grofman, B. 1984. Optimizing group judgmental accuracy in the presence of interdependencies. *Public Choice*.
- Tao, D.; Cheng, J.; Yu, Z.; Yue, K.; and Wang, L. 2018. Domain-weighted majority voting for crowdsourcing. *IEEE transactions on neural networks and learning systems*.
- Tversky, A. 1977. Features of similarity. *Psychological Review*, 84(4): 327–352.
- Welinder, P.; Branson, S.; Belongie, S. J.; and Perona, P. 2010. The Multidimensional Wisdom of Crowds. In *NIPS*.
- Xia, L.; Conitzer, V.; and Lang, J. 2010. Aggregating preferences in multi-issue domains by using maximum likelihood estimators. In *AAMAS*.
- Young, H. P. 1988. Condorcet’s theory of voting. *American Political science review*.