Proxy Voting for Revealing Ground Truth

Gal Cohensius and Reshef Meir
Technion—Israel Institute of Technology

ABSTRACT

We consider a social choice problem where only a small subset of voters actually votes. The outcome of a vote with low participation rate could be far from the outcome reached by a vote with full participation. A possible solution to a vote with low participation rate is allowing voting by proxies. Proxy voting is a scenario which enables the voters that do not vote to transfer their voting rights to another voter.

In some voting settings voters try to discover or agree upon some ground truth while each voter gets a noisy signal about that truth [1, 7]. From this viewpoint, different voting scenarios can be compared upon the expected distance of the aggregated outcome from the truth. By comparing voting with and without proxies, we try to define the conditions under which proxy voting helps to get closer to the truth. A specific model of proxy voting was suggested and studied in [4]. In this paper we apply this model to the case where a ground truth exists. We analyze datasets of social choice and multiple-choice questions and show that Proxy voting can be beneficial in order to find an outcome that is closer to the ground truth. When the participation rate is low enough, proxy voting is always beneficial. In some instances, proxy voting can even get closer to the truth than a vote with full participation. This is a bit surprising since proxy voting uses strictly less information than full participation vote.

1. INTRODUCTION

In the model of proxy voting suggested by [4], it is shown that in various domains, allowing proxy voting results in an outcome that is closer to the aggregated opinion of the entire population. This means that proxy voting improves the social outcome when the outcome reached by the whole population is assumed to be good. In contrast to [4], where it is assumed that the aggregated vote of the entire population is optimal, in this paper we consider voting profiles that are derived from some *ground truth*. Thus the criterion for successful voting mechanisms is finding an outcome that is close to this ground truth. This paper compares proxy voting to full participation vote and to partial participation vote, and categorizes the conditions for which proxy voting is beneficial i.e. it results in an outcome that is closer to the ground truth according to some natural metric.

For example, when votes are orders (permutations) over a set, a natural metric is the Kendall tau distance. We use this distance to assign proxies (i.e., an inactive voter will relegate her voting rights to the voter with nearest vote). Then we aggregate votes to a single order using some standard social welfare function (voting rule), where only active voters participate, and are weighted by their number of followers. Finally, we measure the Kendall tau distance from the aggregated outcome and the ground truth.

Appears at: 4th Workshop on Exploring Beyond the Worst Case in Computational Social Choice (EXPLORE 2017). Held as part of the Workshops at the 16th International Conference on Autonomous Agents and Multiagent Systems. May 8th-9th, 2017. Sao Paulo, Brazil.

1.1 Contribution

We apply the proxy voting model from [4] both to synthetic datasets generated from Mallow's distribution model with a given ground truth order, and to empirical datasets in two natural domains. The first is from a crowd sourcing experiment where subjects were requested to order four items according to their correct order [6], and the second is from Pisa standard tests take by Israeli students.

We show that for all voting rules we tried and for every sample size, it is beneficial to allow inactive voters to use proxies, in the sense that the aggregated vote becomes closer to the ground truth on expectation. For some datasets, proxy voting can be even better than a full participation vote.

We analyze the reasons for this improvement by looking at the distribution of proxies' weights, and suggest a preliminary theoretical result that explains why better proxies get higher weights.

2. PRELIMINARIES

We follow the model and the definitions of [4] as described below.

Domains.

$\mathcal{X}$ is the *space*, or set of possible votes, or voter types. Consider some finite set of alternatives $A = \{a_1, \ldots, a_l\}$. In this paper we will consider two Domains:

1. multiple discrete issues $\mathcal{X} = A^k$ with the Hamming distance
2. ordinal preferences $\mathcal{X} = \Pi(A)$ with the Kendall tau distance.

distances.

The Hamming distance between two agents' positions v_1, v_2 is the number of issues on which the agents disagree on. For example, the Hamming distance between $v_1 = (1, 5, 0, 0), v_2 = (2, 1, 0, 1)$ is 3 since the agents' positions are different in issues $j = 1, 2, 4$.

The Kendall tau distance between two ranking is the number of pairwise disagreements between them. Kendall tau distance is also called bubble-sort distance since it is equivalent to the number of swaps that the bubble sort algorithm would make to place one list in the same order as the other list. For example, the Kendall tau distance between (a,b,c,d) and (b,c,a,d) is 2, since two swaps are needed in order to get from the first ranking to the second. Note that Hamming distance between agents' positions equals the Kendall tau distance between their ranking.

Ground Truth and profiles.

The *Truth* $T \in \mathcal{X}$ is a particular point in space (either an m-size vector in domain (1), or an order over alternatives in domain (2)). The model does not assume a priori any dependence between the truth and the voting profiles.

S_N is the voting profile of the set of voters N of size n. The interesting cases are when society have some idea about the truth, that is to say, there is some dependency between S_N and the truth T.

Mechanisms.

A *mechanism* $\mathbf{g} : \mathcal{X}^n \to \mathcal{X}$ (also called a voting rule) is a function that maps any profile (set of positions) to a winning position.

For the binary issues we will focus on a simple *Majority mechanism* that aggregates each issue independently according to the majority of votes. That is, $(\mathbf{mj}(S))^{(j)} = 1$ if $|\{i : s_i^{(j)} = 1\}| > |\{i : s_i^{(j)} = 0\}|$ and 0 otherwise, where $s^{(j)}$ is the j'th entry of position vector s. For example, say that the number of voters n is 7, for any issue j the outcome of the $\mathbf{mj}$ mechanism will be '1' if at least 4 of the voters vote '1', else the outcome will be '0'. In all mechanisms we break ties uniformly at random.

For the ordinal preferences we use pairwise majority ($\mathbf{mj}$) in addition to four different voting rules: Kemeny, Borda, Plurality and Veto which are denoted respectively by $\mathbf{km}, \mathbf{bo}, \mathbf{pl}$ and $\mathbf{vt}$. In order to aggregate pairwise majority, ordinal votes are converted into issues by checking for each pair of alternatives (α, β) whether α is preferred over β. For example assume votes are a strict ranking of 4 alternatives, then a conversion into issues will result in $\binom{4}{2} = 6$ discrete binary issues. On each issue a majority voting rule is resolute.

All the voting rules (mechanisms) that we use naturally extend to weighted finite populations, by considering voting with w_i copies of voter i

Scenarios.

We label the 'everyone vote' scenario as E, the basic scenario as B and the proxy scenario as P. In scenario E, all voters votes and the result is $\mathbf{g}(S_N)$. In scenario B, only the subset of active voters $M \subseteq N$ votes, while inactive voters abstain. The result is $\mathbf{g}(S_M)$. In scenario P, active voters vote, while each unavailable voter grant her voting right to an active voter. Given a set M of active agents, the decisions of inactive voters are specified by a mapping $J_M : \mathcal{X} \to M$, where $J_M(x) \in M$ is the proxy of any voter located at $x \in \mathcal{X}$. Thus the results in scenario P is $\mathbf{g}(S_M, \mathbf{w}_M)$, where for each $j \in M$, $w_j = |\{i \in N : J(s_i) = j\}|$, i.e. the weight of proxy j is the number of inactive voters who select proxy j, plus himself.

Without further constraints, we will assume that the proxy of a voter at x is always its nearest active agent, i.e. the agent whose position (or vote) are most similar to x. Thus for every subset of active agents M, we get a partition of $\mathcal{X}$. We can compute according to the metric we choose the *weight* of each active agent j. This is done by summing the number of inactive agent that j is their closest active agent. Formally, $J_M(x) = \text{argmin}_{j \in M} \|x - s_j\|$ and $w_j = |\{i \in N : J(s_i) = j\}|$. For example, voter i' is at location $x_{i'}$ and is following the closest active voter to her, which is $J_M(x_{i'}) = \text{argmin}_{j \in M} \|x_{i'} - s_j\| = j'$, the weight of proxy j' is increased by one. If there are several proxies at the same distance, voter i selects one of them at random.

To recap, an *instance* is defined by a profile S_N and a truth T. Each instance produce an outcome according to the scenario $Q \in \{E, B, P\}$, mechanism $\mathbf{g} \in \{\mathbf{km}, \mathbf{bo}, \mathbf{pl}, \mathbf{vt}, \mathbf{mj}\}$, and the sample size $|M| = m$.

Evaluation.

We want to measure how close is $\mathbf{g}^Q(S_N)$ to the *truth* T. We define the *error* as the distance between $\mathbf{g}^Q(S_N)$ and the truth. Note that the Kendall tau distance is the Hamming distance over the induced binary vectors of pairwise preferences (where each pair of alternatives in A induces a binary issue). Thus the distance between any two votes $\mathbf{s}, \mathbf{s}' \in \mathcal{X}$ can be written as $\|\mathbf{s} - \mathbf{s}'\|$ (since these are binary vectors it does not matter which norm is used). In particular, the error of $\mathbf{g}$ on S_N in scenario Q is $\|\mathbf{g}^Q(S_N) - T\|$.

The *loss* of a mechanism $\mathbf{g}$ is calculated according to its mean square error (MSE)—the expected squared distance from the truth—over all samples of m available voters.

$$\mathcal{L}^Q(T, S_N, m) = \mathbb{E}_{M \sim U[N^m]} \left[\|\mathbf{g}^Q(S_N) - T\|^2 \right], \qquad (1)$$

where the mechanism $\mathbf{g}$ can be inferred from the context, and the expectation is over all subsets of m positions sampled uniformly without repetitions from S_N (sometimes omitted from the subscript).

3. SIMULATIONS

The experiments were designed to test two hypothesis:

1. $\mathcal{L}^P < \mathcal{L}^B$ for every voting rule. That is, whether for random samples of a given size m, proxy voting always yields an outcome which is closer to the truth than an outcome yield by unweighted vote with the same set of proxies.

2. there is some setting where $\mathcal{L}^P < \mathcal{L}^E$. That is, under certain parameters, taking a sample of active voters and use them as proxies will yield an outcome which is closer to the truth than the outcome reached by aggregating all votes.

3.1 Datasets

3.1.1 Generative model of votes

We generate synthetic profiles, by sampling rankings from Mallow's model. Mallow's distribution model is a distance-base ranking model, which is parametrized by a true order T and a dispersion parameter $\phi \in (0, 1]$. For any ranking $r \in \Pi(A)$, the Mallows model specifies:

$$\Pr(r) = \Pr(r|T, \phi) = \frac{1}{Z} \phi^{d(r,T)}$$

Where d is the Kendall tau distance and $Z = \sum_{r' \in \omega} \phi^{d(r',T)}$ is a normalization constant. When $\phi = 1$ the distribution is uniform over all permutations (very noisy), when $\phi \ll 1$ almost all the mass is concentrated at T (small amount of noise). Using synthetic datasets helps understanding the role of each parameter while fixing the others. Mallow's distribution model is one of the two most popular noise models in the machine learning community together with Plackett-Luce [5].

3.1.2 Natural experiments

We used two ranking datasets made by [6] using crowd-sourcing. One is referred to as the dots dataset. In that test voters were shown four pictures with dots and were asked to ranked the pictures by the number of dots from least to most. The number of

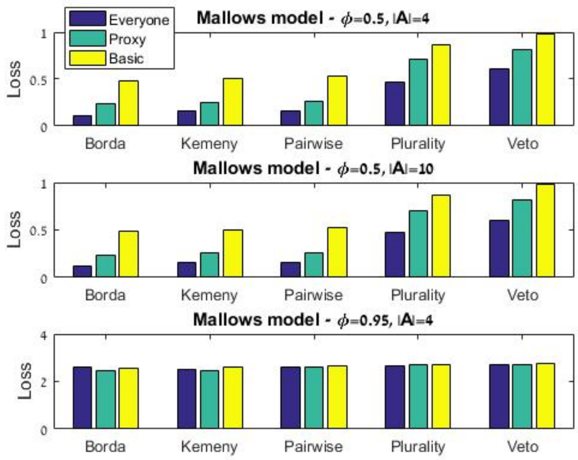

Figure 1: $\mathcal{L}^{\{E,B,P\}}$, under generative model of preference, using five voting rules. Note that for Mallow's model with low dispersion $\mathcal{L}^{E} < \mathcal{L}^{P} < \mathcal{L}^{B}$. For Mallow's model with high dispersion $\phi = 0.95$, (high mistakes probability) $\mathcal{L}^{P} < \mathcal{L}^{E}$.

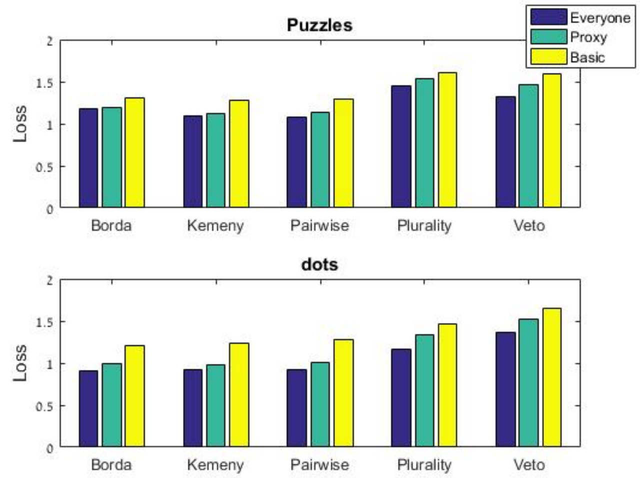

Figure 2: $\mathcal{L}^{\{E,B,P\}}$, for puzzle and dots datasets, using five voting rules. Proxy voting is doing better than random sample, i.e. $\mathcal{L}^{E} < \mathcal{L}^{P} < \mathcal{L}^{B}$.

dots in the pictures were $\{200, 200+i, 200+2i, 200+3i\}$ for $i = \{3, 5, 7, 9\}$. The data contain at each level of noise, i, 40 preference profiles. Each profile contain about 20 voters. This test has been suggested as a benchmark task for human computation in [2]. In the second dataset, refereed as the sliding puzzle, voters had four situations of a size 8 sliding puzzle, and they had to order those puzzles by the minimal number of moves left to solve them. The number of minimal moves were $\{d, d+3, d+6, d+9\}$, for $d = \{5, 7, 9, 11\}$. Again, 40 preference profiles, each containing 20 voters, were gathered. Those dataset were valuable for us since they contain a true order. A comprehensive explanation on how the data was gathered in the two experimental datasets can be found at [6]. Another dataset that we examined is the results of a test called the program for international student Assessment (PISA). This test evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. We where looking for a dataset of multi choice questions (without missing data), thus we included only the students who answered all of the first 18 multiple choice questions. This data match the first domain (multiple discrete issues). Our dataset contain $n = 571$ student (voters), answering $k = 18$ questions (issues), each question has 4 possible answers.

3.1.3 Method

The simulation start by creating a profile of votes, either by sampling from Mallow's model or by loading the empirical datasets. Then, a ranking profile of active voters was simulated for each scenario: The E scenario used the original profile with N voters. The B scenario sample uniformly at random, a given size $m < n$ of active voters while the other abstained. In P scenario M were active voters while each of the other voters $N \setminus M$ becomes a clone of her closest agent in M. Thirdly an aggregated order was calculated for each scenario using five popular voting rules: $\{\mathbf{km}, \mathbf{bo}, \mathbf{pl}, \mathbf{vt}, \mathbf{mj}\}$. Lastly, for each voting rule, we compared the Kendall tau distances (Hamming distance for Pairwise majority) from the truth T to the order obtained by each scenario.

3.2 Results

Our simulations on the generative model shows that $\mathcal{L}^{P}(\mathbf{g}) < \mathcal{L}^{B}(\mathbf{g})$ for all five voting rules $\mathbf{g}$, in all datasets, and for almost every sample size m (Figs. 1, 2 and 5). Same results are obtain analyzing the Pisa dataset (Fig. 9) with the Majority voting rule (In this dataset the domain is multiple discrete issues, thus the voting rule is majority.) This results supports our conjecture that proxy voting reveals ground truth better than a random sample, and often considerably better.

The big difference in the settings from [4] is that when there is some hidden ground truth, the outcome gathered from full participation vote (scenario E) is *not necessary optimal*. On some votes the best active voter get much closer than the aggregated decision of the entire population, thus there is hope that with the appropriate weights, proxy voting can do better than E. Indeed for some datasets proxy voting is even better than a full participation vote, e.g., Mallow's model with 4 alternatives, 20 voters and $\phi = 0.95$ (Fig. 1). This is also true for about half the crowd-sourcing datasets (Fig.3). This is an interesting phenomenon since proxy voting uses strictly less information than a full participation vote. In future work we will try to characterize the conditions for that to happen. For now we only found some rules of thumb:

- Amount of noise should be high enough in order to make the E scenario do pretty bad. If ordering is too easy, E will have almost no loss, thus no scenario can do better, this is the situation in [4] where the outcome reached by scenario E is the optimal one.

- There should be high variance in the individual performance of voters. If all voters have roughly the same accuracy then proxy voting does not help much.

3.3 Analysis

Denote by $R_i = \|\mathbf{s}_i - T\|$ the distance of voter i from the truth. The reason that proxy voting gets closer to the truth than a random sample lies in the weight distribution of the proxies. While in scenario B the weights are uniform by definition, at scenario P the weights are roughly decreasing in their ratio of mistakes R_i, that is

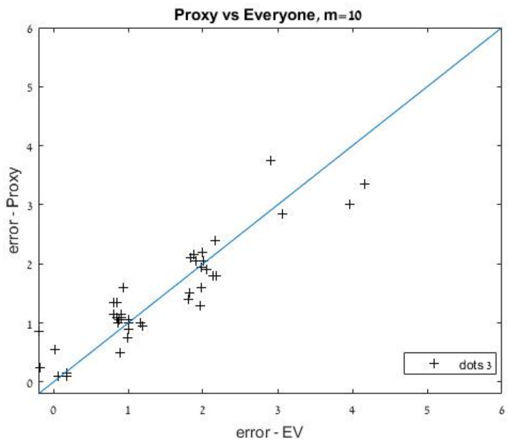

Figure 3: $\mathcal{L}^E$ vs. $\mathcal{L}^P$, examining the 40 dots datasets (noise level i=3), under Borda voting rule and $m = 10$ proxies. Markers under the 45 degree line are datasets where P is closer to the ground truth than E. A small random noise was added to $\mathcal{L}^E$ in order to better visualize close outcomes.

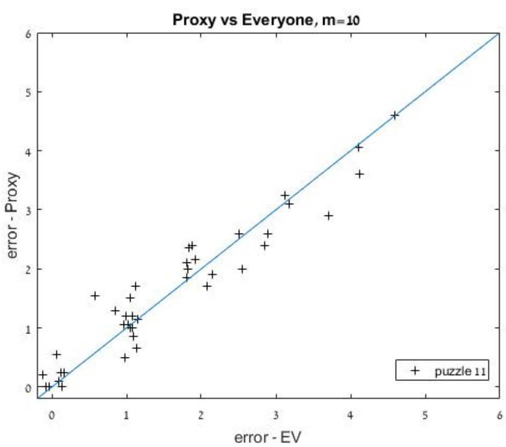

Figure 4: $\mathcal{L}^E$ vs. $\mathcal{L}^P$ shown for the 40 datasets of $d = 11$ puzzle.

to say, better proxies get more voting weight. When dispersion is low, the distribution of the weights is monotonic decreasing in R_i. When dispersion raises, some of the voting weight moves towered the worst proxy, resulting a single dip distribution, with peaks at the best and worst proxies, see Fig 7.

4. EXPLAINING PROXY WEIGHTS

In the previous section, we observed empirically that better proxies (i.e. ones closer to the ground truth T) tend to get more followers and thus higher weight. We are interested in a theoretical model that explains this. One such result was given in [4] for the limit case of $k \to \infty$ binary issues, where essentially *all* inactive voters select either the best or the worst proxy, according to which one is closer. However, in realistic scenarios (including our datasets), the number of issues is much smaller.

We model a simplified version of the problem, where there is one follower which is requested to choose a proxy from two active voters. A priori, we only know the distribution of votes, and we

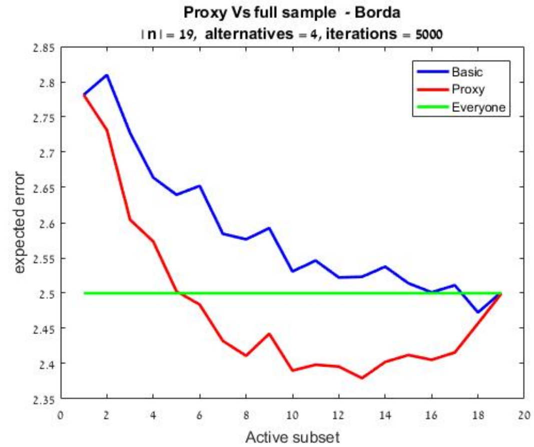

Figure 5: The expected error in scenarios $\{E, B, P\}$. Using Borda voting rule. Examining $d = 11$ puzzle dataset . $\mathcal{L}^P < \mathcal{L}^E$ for a subset of active voters M large enough.

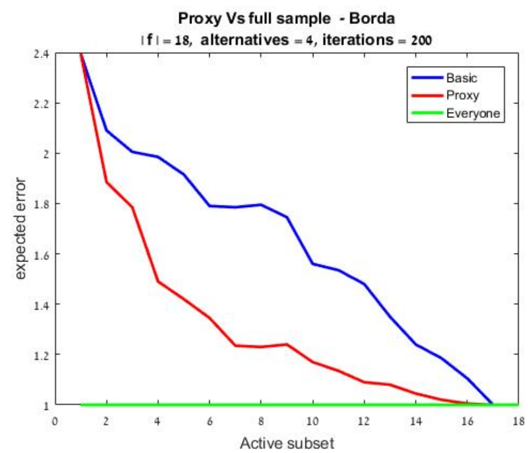

Figure 6: P is decreasing faster than B and better approximating E.

want to estimate the probability that the follower would choose the better proxy.

Following [4], we model each agent with a fixed error probability P_i. Consider two active agents with error probabilities $P_i < P_j < 0.5$, and an inactive agent with error probability $Z < 0.5$. W.l.o.g. $T = \mathbf{0}$. Thus $\mathbf{s}_i, \mathbf{s}_j$ and $\mathbf{z}$ are random binary vectors of length k, whose entries are '1' with respective probabilities of P_i, P_j, and Z.

Fix the values of the best proxy $P_i = P$, an inactive agent Z, and the number of issues k. Denote by $\epsilon = P_j - P_i > 0$ the difference in the quality of proxy j from the best proxy. We want to understand better how the probability of selecting the worse proxy P_j behaves as ϵ and k vary. Note that this probability is taken in expectation over all realizations of $\mathbf{s}_i, \mathbf{s}_j, \mathbf{z}$, as in each such realization the decision of the inactive voter is deterministic (up to tie-breaking).

[4] showed that z is more likely to be closer to $\mathbf{s}_i$, and that the probability of being closer to j drops exponentially with the number of issues k. Let $q_i = Pr(s_i^{(t)} \neq z^{(t)}) = P_i(1 - Z) + (1 -$

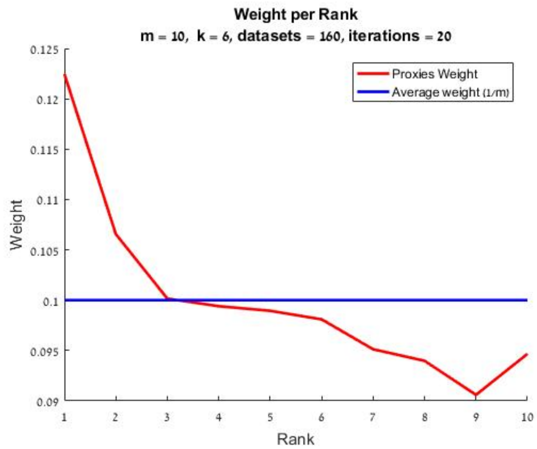

Figure 7: Weight per proxy, ordered by the ratio of wrong answers R_i in increasing order. The dots dataset.

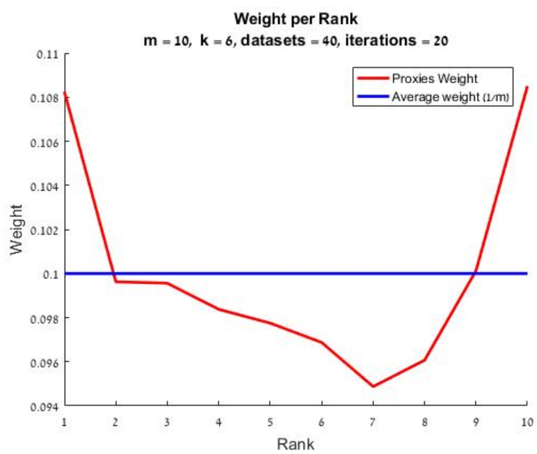

Figure 8: Weight per proxy, ordered by the ratio of wrong answers R_i in increasing order. Mallow's model with 4 alternatives, 20 voters, and dispersion parameter $\phi = 0.95$.

$P_i)Z = P_i + Z - 2P_iZ$. Indeed, they showed

$$Pr(\|z - s_i\| > \|z - s_j\|) \cong \Phi\left(\frac{\sqrt{k}(q_j - q_i)}{\sqrt{q_i(1-q_j) + q_j(1-q_i)}}\right),$$

where $\Phi(x) = Pr_{X\sim N(0,1)}(X > x)$, and the approximation is due to the Binomial-to-Normal approximation.

Note that $q_j - q_i = P_j + Z - 2P_jZ - (P_i + Z - 2P_iZ) = (P_j - P_i)(1 - 2Z) = \epsilon(1 - 2Z)$. Thus

$$Pr(\|z - s_i\| > \|z - s_j\|) \cong \Phi\left(\frac{\sqrt{k}(q_j - q_i)}{\sqrt{q_i(1-q_j) + q_j(1-q_i)}}\right)$$

$$= \Phi\left(\frac{\sqrt{k}(1-2Z)\epsilon}{\sqrt{\begin{array}{l}(P+Z-2PZ)(1-(P+\epsilon+Z-2(P+\epsilon)Z)) \\ \quad + (P+\epsilon+Z-2(P+\epsilon)Z)(1-(P+Z-2PZ))\end{array}}}\right)$$

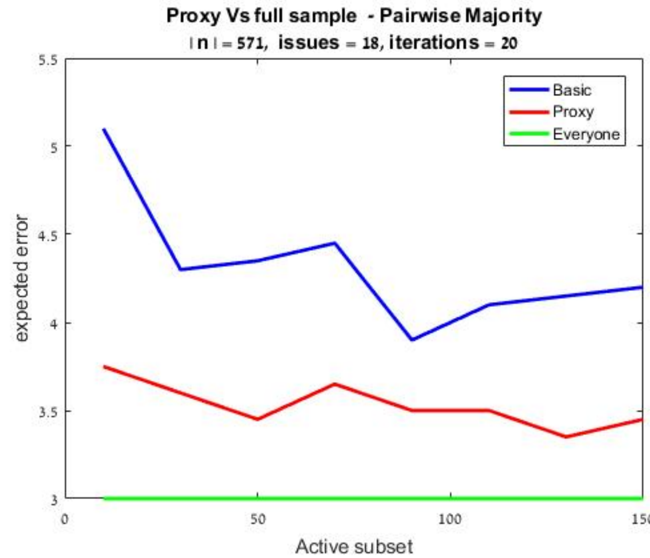

Figure 9: The expected error in scenarios $\{E, B, P\}$. Examining the Pisa dataset. The expected error of E is 3, meaning that the aggregated majority vote made 3 mistakes out of 18 questions. The expected error for both scenarios B, P is roughly decreasing in the number of active voters m but is substantially lower for P.

$$= \Phi\left(\frac{\sqrt{k}(1-2Z)\epsilon}{\sqrt{\begin{array}{l}(4Z^2 - 8PZ^2 + 8PZ - 4Z + 2P + 1)\epsilon \\ \quad + 2Z(1-Z) + 2P(1-P) - 8PZ(1-P)(1-Z)\end{array}}}\right)$$

$$= \Phi\left(\frac{C_1\epsilon}{\sqrt{C_2\epsilon + C_3}}\right) = \Phi\left(\Theta(\sqrt{\epsilon})\right) = exp(-\Theta(\epsilon)).$$

For some constants $C_1 > 0, C_2 > 0, C_3$

That is, the probability that j is selected decreases exponentially fast in the distance between the error rate of j and that of the best proxy. The drop is exponential when the distance is large enough and there are enough issues. Another observation is that if we fix $\epsilon < P < 0.5$ and Z approaches 0.5 (i.e. an ignorant inactive agent), then the term in brackets approaches 0. In other words, ignorant agents spread their weight roughly evenly over all active voters, whereas smart agents are substantially more likely to give their vote to a good active voter.

This supports the intuitive argument from [3] regarding the "Anna Karenina principle" (as good agents are indeed similar to one another), and thus at least partially explains the weight distribution of active agents. To see if this is a sufficient explanation, one needs to compare the actual weight distribution, and specifically $\frac{w_j}{w_1}$, to the expression above.

REFERENCES

[1] Ioannis Caragiannis, Ariel D Procaccia, and Nisarg Shah. When do noisy votes reveal the truth? In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pages 143–160. ACM, 2013.

[2] Lydia B Chilton, John J Horton, Robert C Miller, and Shiri Azenkot. Task search in a human computation market. In

Proceedings of the ACM SIGKDD workshop on human computation, pages 1–9. ACM, 2010.

[3] Gal Cohensius, Shie Manor, Reshef Meir, Eli Meirom, and Ariel Orda. Proxy voting for better outcomes. *arXiv preprint arXiv:1611.08308*, 2016.

[4] Gal Cohensius, Shie Manor, Reshef Meir, Eli Meirom, and Ariel Orda. Proxy voting for better outcomes. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2017)*, 2017.

[5] Tyler Lu and Craig Boutilier. Effective sampling and learning for mallows models with pairwise-preference data. *Journal of Machine Learning Research*, 15(1):3783–3829, 2014.

[6] Andrew Mao, Ariel D Procaccia, and Yiling Chen. Better human computation through principled voting. In *AAAI*. Citeseer, 2013.

[7] H Peyton Young. Condorcet's theory of voting. *American Political science review*, 82(04):1231–1244, 1988.