Aggregating Rankings from Heterogeneous Voters

Doron Kabla and Reshef Meir
Technion—Israel Institute of Technology

ABSTRACT

Background: Aggregating rankings arriving from different sources is a common task both in crowdsourcing (where sources are workers or experts) and in machine learning (where sources may be different algorithms). In crowdsourcing, it is common that the experts are heterogeneous in their expertise.
Objectives and Research Questions: We analyze the results of different voting rules under the assumption of a ground truth noise model. What are the most likely results? How close are voting rules to the ground truth?
Methods: This is a theoretical paper. We prove the optimality of some voting rules, characterize some optimal methods that are not voting rules. We also analyze the distance from the ground truth (number of pairwise errors) each voting rule makes asymptotically.
Results: Most known results on MLE generalize to heterogeneous experts (by using appropriate weights), with important caveats. We provide bounds on the asymptotic distance for known voting rules, and prove that some known voting rules are substantially better than others, even for a single voter.
Conclusions: Weighted aggregation is optimal when experts' expertise is known. Most common voting rules work better as the number of items increases, but some improve substantially faster than others.

CCS CONCEPTS

• **Applied computing** → *Voting / election technologies*; • **Information systems** → *Rank aggregation*; • **Mathematics of computing** → **Maximum likelihood estimation**.

KEYWORDS

Social Choice Theory, Rank Aggregation, Heterogeneous Agents

ACM Reference Format:
Doron Kabla and Reshef Meir, Technion—Israel Institute of Technology. 2026. Aggregating Rankings from Heterogeneous Voters. In *ACM Conference, Washington, DC, USA, July 2017*, IFAAMAS, 20 pages.

1 INTRODUCTION

Suppose that an expert (a person or an algorithm) is comparing two items A and B, where A is 'better' according to some objective (say, larger, faster, or with higher expected profit). The simplest way to capture the uncertainty involved in such a comparison, is to assign some 'success probability' p: with probability p the expert will correctly answer A, and with the complementary probability $1 - p$ they will answer B. This simple model is sometimes known as the 'one-coin' model. To decide which item is more likely to be the better one, we could use a Majority vote: This is the basis for the famous Condorcet Jury Theorem [3].

The Condorcet model. The picture gets more complicated when comparing multiple items, with the aim of producing a ranking. The earliest approach was proposed by Condorcet, who suggested the following generalization: given a set of items I, each pair is independently compared with success probability $p > 0.5$. In case of a failure, the expert will mistakenly estimate the wrong item to be better, possibly producing non-transitive comparisons.

The Kemeny-Young rule. The Kemeny voting rule selects the ranking that minimizes the total number of pairwise disagreements with experts' votes [13], also called Kentall-tau distance. A fundamental result by Young [20] shows that regardless of experts' accuracy p, the most likely true ranking is the one returned by the Kemeny voting rule.

In the same paper, Young also considered a different problem, which is to find the most likely *winner*. He showed that this may differ from the top of the most likely ranking, and studied conditions under which the result coincides with familiar voting rules: for experts with low accuracy, the most likely winner is found by the Borda rule; and by the Maximin rule for experts with high accuracy. The latter result about Maximin was only formally stated for three items, and in a more recent paper, Elkind and Shah [8] showed it is actually the Tideman rule (with appropriate tie-breaking) that returns the maximum likelihood winner in the high-accuracy regime.

Heterogeneous experts. Consider the two item case. If every expert is correct with a probability of p_i, then a simple likelihood calculation shows that the most likely item is selected by a weighted majority, where the weight of expert i is $q^*_i = q^*(p_i) := \log \frac{p_i}{1-p_i}$. This observation goes back to the Neyman-Pearson lemma [16], and was explained in the context of crowdsourcing in another important paper by Grofman et al. [12]. We therefore refer to these weights as 'Grofman weights'. These weights were further studied by Berend and Kontorovich [2], which provided sharp bounds on the error probability of weighted aggregation with optimal or near-optimal weights, later improved by [15].

1.1 Expected Distance from Ground-Truth

A related but different question regards bounds on the distance of voting outcome from the truth. This question comes naturally from the machine learning literature, which usually defines some loss function that should be optimized. Several distance measures are known for full rankings, with Kendall-tau being the most prominent. While this measure was used to construct voting rules, most famously the Kemeny-Young rule, the question of how good are voting rules as a function of their distance from the ground truth

---

ACM Conference, , July 2017, Washington, DC, USA. © 2026 Association for Computing Machinery. …$ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
…$15.00

ranking was rarely proposed. An important work in this regard is the study of voting under *adversarial noise* done by Proccacia et al. [17], which showed that there exists an algorithm that can find a ranking of at most twice the average distance of every vote from the truth, including for the Kendall-tau distance.

In the context of random noise, Caragiannis et al. [6] considered the sample complexity of general distance-based noise models, and proved that the distance for some classes of rules decreases exponentially in the number of experts. The dependency on the number of items was not explored. Another line of work is that of Azari Soufiani et al. [1] and Xia [19], which aim to characterize the optimal decision rules with respect to some noise models and loss functions (without providing concrete bounds).

Noisy Sorting. While not explored in the social choice literature (as far as we know), the problem of bounding the error was studied in the algorithms literature under the name *noisy sorting* [5, 11]. Specifically, their model of Noisy Sorting Without Resampling (shortened to NSWR) coincides with the Condorcet noise model with *a single voter.* The first work analyzing this model was done by Braverman and Mossel [5], and proposed a randomized polynomial time algorithm that with high probability finds the ranking with the minimal number of errors (the Kemeny ranking in our terminology).[1] In that paper it was also proved that the Kemeny ranking has an expected distance $O(m)$ from the truth, but the proof used for this claim is wrong, and we have not been able to verify the claim. Further work has produced different algorithms that have decreasing runtime, with the state of the art work by Geissman et al. [11] running in $O(m \log m)$ time, with $O(m)$ errors with high probability and maximum displacement of $O(\log m)$ with high probability. They also provided matching lower bounds. Another paper in this field is that of Kunisky et al. [14], which discusses the ability to distinguish the Condorcet noise model from the impartial culture, and also proves that the Borda method achieves $o(m^2)$ errors. It is curious to note that none of the above work uses any terminology of social choice despite working in an identical setting.

1.2 Our Contribution

In this paper we consider the Condorcet noise model for rankings, with heterogeneous experts whose accuracy is known. A natural conjecture is that Young's results should extend to the heterogeneous case, i.e. that Kemeny voting with appropriate weights would return the maximum likelihood ranking.

Our first contribution is proving this conjecture, showing that Kemeny voting rule with Grofman weights is indeed optimal in the MLE sense. While technically straight-forward, we believe this should still be formally shown.

We then consider the most likely winner problem, again extending Young's and followers work, and in particular the work of Elkind and Shah [8]. Here the generalization is somewhat more intricate and we show that the maximum likelihood disagrees with its respective voting rule, except in extreme cases which we demonstrate.

Our main contribution is providing bounds on the expected distance of voting rules from the ground truth. While the work on noisy sorting was searching for good solutions in this setting, the only results applicable to voting rules are the lower bounds, and the $o(m^2)$ result for Borda with a single voter[2]. We prove almost tight bounds for several known voting rules, which are asymptotically different. We also offer a method for calculating the optimal rule in this measure, which is not the Kemeny-Young rule but is related to it. Finally, we generalize the lower bound from [11] from a single to multiple heterogeneous voters.

2 PRELIMINARIES

We denote by I a fixed set of m items. Let $\mathcal{L}(I)$ denote all orders over I, and $\binom{I}{2} = \{\{x, y\} | x, y \in I\}$ be the set of all distinct pairs over I. For any order $L \in \mathcal{L}(I)$ and items $a, b \in I$ we denote $a \succ_L b$ if a is ranked above b in L.

Crowdsourcing. We consider the case where there exists some *ground truth ranking* $L^* \in \mathcal{L}(I)$, which is the ranking we are trying to approximate (we also call it *true ranking*). We consider a set N of n experts. Each expert sees and reports a noisy variation $\boldsymbol{s}_i$ of the *ground truth ranking* L^*. The report itself may not be a valid ranking. Rather, $\boldsymbol{s}_i$ is a vector with $\binom{m}{2}$ entries, specifying all pairwise relations among items.

Thus, an instance of crowdsourcing is a pair $\langle S = (\boldsymbol{s}_i)_{i \in N}, L^* \rangle$, where $\boldsymbol{s}_i$ is the report of expert i, and L^* is the true ranking. S is also called a *dataset.* For a subset of items $I' \subseteq I$, we denote by $S_{I'}$ the data obtained only from experts' rankings over I'. We also use the shorthand notation S_{xy} to denote $S_{\{x,y\}}$. Similarly, we denote $L_{I'}$ for some ranking L to be the ranking of L restricted to the items in I', and $L_{xy} = L_{\{x,y\}}$.

We use $a \succ_i b$ to denote that a is ranked above b in $\boldsymbol{s}_i$, and $a \succ b$ as a shorthand for $a \succ_{L^*} b$ (meaning that a is better than b according to the ground truth).

Finally, we denote by $\mathcal{S}(I)$ the set of all possible datasets over I. Our goal is to take $\mathcal{S}(I)$ as input and return a valid ranking $\hat{L} \in \mathcal{L}(I)$ which is most likely or closest to be the true ranking. Alternatively, we might only be interested in the most likely winner $\hat{a} \in I$.

The Condorcet noise model. For two items, the simplest assumption is that each expert ranks them correctly with probability $p > 0.5$.

For more than two items, Condorcet suggested that any pair is ranked by an i.i.d. coin-flip, even if this results in experts' reports that may be intransitive.

As explained in the introduction, two natural relaxations of the model are:

- Allowing a different accuracy p_i for every expert; and
- Allowing the accuracy to depend on how far the two items are in the true ranking.

In this paper we only consider the first extension, denoting by $\boldsymbol{p} = (p_i)_{i \in N}$ the vector of experts' accuracies. For an event E (some realized profile), we denote by $Pr_{\boldsymbol{p}}(E)$ the probability of that event under the weighted Condorcet noise model, with accuracies $\boldsymbol{p}$. We often omit $\boldsymbol{p}$ when clear from context.

---

[1] This does not contradict the worst-case computational hardness of Kemeny, since it is restricted to a particular input distribution induced by the noise model.

[2] The proofs for Kemeny are wrong.

Formally, the Condorcet model can be phrased by the following two equations:

$$Pr(S_{xy}|x \succ_L y) = Pr_{\boldsymbol{p}}(S_{xy}|x \succ_L y) := \prod_{i:x \succ_i y} p_i \prod_{i:y \succ_i x} (1 - p_i) \tag{1}$$

$$Pr(S|L) = Pr_{\boldsymbol{p}}(S|L) := \prod_{x,y \in I: x \neq y} Pr_{\boldsymbol{p}}(S_{xy}|L_{xy}), \tag{2}$$

where L_{xy} is either $x \succ_L y$ or $y \succ_L x$, according to which item is ranked higher in L.

2.1 Voting Rules

The term *voting rule* usually refers to a function that maps a set of rankings (a.k.a. a voting profile) to a single winning item or to a ranking over items. This is known as a social choice function (SCF) in the first case and a social welfare function (SWF) in the latter. Any SWF induces an SCF by picking the top item in the ranking, and every SCF induces a (randomized) SWF by iteratively selecting the next top item, breaking ties uniformly.

Some rules are based on scoring every item and/or ranking, then picking the one with the highest score.

A voting rule can be *weighted*. Intuitively, a expert with weight k should have the same effect as k unit-weight experts. We denote the weight of expert i by q_i. Thus a (weighted) SCF is a function $f : \mathcal{S}(I) \times \mathbb{R}^n \to I$, and a (weighted) SWF is a function $g : \mathcal{S}(I) \times \mathbb{R}^n \to \mathcal{L}(I)$.

Graph-based rules. Every profile (set of rankings) S induces a weighted directed graph where nodes are the items, and the weight of an edge $w(x, y)$ is the number of experts who rank x over y.

Note that this graph naturally extend to weighted votes, by setting the weight of an edge to the sum of weights of the respective experts (See Fig. 2),

$$w_{\boldsymbol{q}}(x, y) := \sum_{i:x \succ_i y} q_i.$$

It always holds that $w_{\boldsymbol{q}}(x, y) + w_{\boldsymbol{q}}(y, x) = \sum_{i \in N} q_i = W$ which is a constant, whereas $w_{\boldsymbol{q}}(x, y)$ depends on S. As usual, we omit the subscript from $w_{\boldsymbol{q}}$ when clear from context. We denote $a \succ_S b$ if $w_{\boldsymbol{q}}(a, b) > w_{\boldsymbol{q}}(b, a)$.

All the voting rules we consider in this paper are scoring-based rules defined on the above graph. Any Scoring-based SCF can be used to create an SWF, by ranking the items according to their scores.

The scores of some familiar graph-based voting rules are defined as follows (see [4]):

Borda The Borda score of an item is the sum of the weights of outgoing edges. Formally, for $a \in I$, weight vector $\boldsymbol{q} \in \mathbb{R}^n$, and profile S,

$$BS(a|\boldsymbol{q}, S) := \sum_{x \neq a} w_{\boldsymbol{q}}(a, x).$$

Copeland The Copeland score of an item is the number of outgoing edges (counting only defeated items).[3] Formally,

$$CS(a|\boldsymbol{q}, S) := \sum_{x \neq a} \mathbb{1}_{\{a \succ_S x\}}$$

experts	p_i	q_i^*	Ranking
1	2/3	1	$a \succ_i c \succ_i b$
2, 3	8/9	3	$a \succ_i c \succ_i b$
4, 5	2/3	1	$b \succ_i a \succ_i c$
6, 7, 8	4/5	2	$c \succ_i b \succ_i a$

S^1	a	b	c
a	–	7	9
b	8	–	2
c	6	13	–

Figure 1: Two representations of the example profile with $m = 3$ items and $n = 8$ heterogeneous experts. The q weights used here are proportional to the Grofman weights.

Maximin The Maximin score of an item is the weight of the minimal outgoing edge. Formally,

$$MS(a|\boldsymbol{q}, S) := \min_{x \neq a} w_{\boldsymbol{q}}(a, x).$$

Tideman The Tideman score of an item can be thought of as a compromise between Borda and Maximin, by summing only edges from defeating items, and selecting the item with *minimal* score. Formally,

$$TS(a|\boldsymbol{q}, S) := \sum_{x \neq a} \begin{cases} w_{\boldsymbol{q}}(x, a) - w_{\boldsymbol{q}}(a, x) & \text{, if } x \succ_S a \\ 0 & \text{, else} \end{cases}$$

Other graph-based scoring rules score rankings directly, picking the ranking with the lowest score. In particular, the Kemeny-Young rule selects the ranking that maximizes the Kemeny Score:

Kemeny The (weighted) Kemeny score of ranking L is the (weighted) sum of pairwise agreements between L and the profile S. Formally

$$KS(L|\boldsymbol{q}, S) := \sum_{x \succ_L y} w_{\boldsymbol{q}}(x, y)$$

Condorcet An item a s.t. $a \succ_S x$ for all $x \neq a$, is called the (weighted) Condorcet winner. Note that when it exists, it coincides with the Maximin, Tideman, Copeland and Kemeny winner.

In Fig. 1, item c is both the (weighted) Borda winner, the Maximin winner, and the Tideman winner; $c \succ b \succ a$ is the Kemeny ranking, and there is no Condorcet winner.

We say a graph based SCF f is *Item Sum-Scoring* if it has a scoring function $V_f : \mathbb{R} \to \mathbb{R}$ such that

$$f(\boldsymbol{q}, S) = \arg\max_a \sum_{x \neq a} V_f\left(w_{\boldsymbol{q}}(a, x)\right).$$

From the rules mentioned above, Borda, Copeland and Tideman are clearly item sum-scoring, whereas Maximin and Kemeny are not (Maximin since it minimizes the scores from head-to-head competitions, and Kemeny since it does not sum scores over items but over *pairs* of items).

Given a SWF g, we denote by $\hat{L}_g(S, \boldsymbol{q})$ the order returned by g. Note that this is a random variable due to tie-breaking. We denote by $D(g, \boldsymbol{p}, \boldsymbol{q}) := E_{S,L^*}[d_{KT}(\hat{L}_g(S, \boldsymbol{q}), L^*)]$ the expected Kendall-tau distance between the winning order and the ground truth, where L^* is the worst case order, and S is sampled using the Condorcet Noise model with parameters $\boldsymbol{p}$.

---

[3] This version counts draws as losses for both items. Other versions count draws differently. In this paper the specific variation does not change the results.

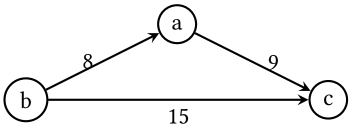


Figure 2: Graph representation of the same profile. For example $w^*(a, b) = q_1^* + q_2^* + q_3^* = 7$.

Constants. We will use $Q, Q', \ldots$ for constants that do not depend on the profile S. In particular, we denote $Q := \sum_{i \in N} \log(1 - p_i)$.

We use $Z, Z', \ldots$ for terms that may depend on S (and $\boldsymbol{p}$), but are constant w.r.t. the evaluated item or ranking (so Z is a shorthand for $Z(S)$).

3 MAXIMUM LIKELIHOOD

3.1 Most Likely Ranking

It is well known that in the homogeneous case, for both Condorcet noise model and Mallows model (a restriction of Condorcet model to transitive orders), the maximum likelihood estimator of L^* is obtained by applying the Kemeny voting rule on S, see [20].

A natural conjecture is that applying a *weighted version* of Kemeny with Grofman weights $\boldsymbol{q}_i^* = \boldsymbol{q}_i^*(\boldsymbol{p}) = \log(\frac{p_i}{1-p_i})$ is an MLE when success probabilities p_i are known. As it turns out, the conjecture is correct.

The first step is to compute the (log-)likelihood of a single edge in the graph. Calculation is the same as in [12], but shown here for completeness and as a demonstration of how our notation is used.

Lemma 1. *[12, 16] $\log Pr_{\boldsymbol{p}}(S_{xy} | x \succ y) = Q + w^*(x, y)$, where w^* is a shorthand for $w_{\boldsymbol{q}^*(\boldsymbol{p})}$.*

Theorem 2. *Suppose that the ground truth L^* is sampled from a uniform prior on $\mathcal{L}(I)$. Then for any rank L, dataset S, and heterogeneous accuracy vector $\boldsymbol{p}$, we have that*

$$\log Pr_{\boldsymbol{p}}(L^* = L | S) = KS(L | \boldsymbol{q}^*(\boldsymbol{p}), S) + Z,$$

where Z is a constant that does not depend on L.

An immediate corollary of the theorem is that the ranking with the highest weighted Kemeny score is also the most likely ranking, generalizing Young's result.

Proof of Theorem 2. We calculate the posterior of L by first calculating the likelihood:

$$Pr(S | L^* = L) = \prod_{x,y \in \binom{I}{2}} Pr(S_{xy} | L^* = L) = \prod_{x \succ_L y} Pr(S_{xy} | x \succ y)$$

Taking log of both sides, and plugging in Lemma 1,

$$\log Pr(S | L^* = L) = \sum_{x \succ_L y} \log Pr(S_{xy} | x \succ y)$$
$$= \sum_{x \succ_L y} (Q + w^*(x, y)) = \binom{m}{2} Q + KS(L | \boldsymbol{w}^*, S).$$

Now, using Bayes' rule, and setting $Z' := \frac{1}{m! Pr(S)}$,

$$\log Pr(L^* = L | S) = \log\left(Pr(S | L^* = L) \frac{Pr(L^* = L)}{Pr(S)}\right)$$
$$= \log Pr(S | L^* = L) + \log(\frac{1}{m! Pr(S)})$$
$$= \binom{m}{2} Q + KS(L | \boldsymbol{w}^*, S) + \log(Z') = KS(L | \boldsymbol{w}^*, S) + Z,$$

as required. □

3.2 Most Likely Winner

As Young explains in [20], the *most likely winner* is not necessarily the top item of the most likely order. Interestingly, Young's paper does not contain formal theorem statements or proofs. Rather, he provides a formula for the posterior winning probability, and a numerical example for three items with correctness probability p.

Previous papers [9, 18] have identified gaps in Young's reasoning and provided a corrected model (that we generalize below to the heterogeneous case). However, we also show in Appendix B that (only) in the three-item case, Young's formula provides the correct *relative* winning probabilities.

An Explicit Score for the Heterogeneous Case. Elkind and Shah [8] provide a way to calculate most likely winner in the homogeneous case, as the item *minimizing* a term they call $\kappa(a)$. We generalize their analysis to the heterogeneous case. Let $\kappa(a) := \prod_{b \in I / \{a\}} (1 + 2^{w^*(b,a) - w^*(a,b)})$.

Proposition 3. *$Pr(top(L^*) = a | S) = \frac{1}{\kappa(a)}$.*

Alternative scoring for items. In addition to the κ term, we need another representation of the item score. We continue by decomposing the likelihood of a potential true winner $a \in I$ in the general heterogeneous case. We define the following term:

$$P_{\boldsymbol{p}}^{ML}(a | S) := Pr_{\boldsymbol{p}}(S_{I \setminus \{a\}}) \prod_{x \neq a} Pr_{\boldsymbol{p}}(S_{ax} | a \succ x). \tag{3}$$

We argue that P^{ML} is proportional to the posterior probability that a is the true winner:

Theorem 4. *Suppose that the ground truth L^* is sampled from a uniform prior on $\mathcal{L}(I)$. Then for any item $a \in I$ and any heterogeneous accuracy vector $\boldsymbol{p}$, we have that*

$$Pr_{\boldsymbol{p}}(top(L^*) = a | S) = Z' \cdot P_{\boldsymbol{p}}^{ML}(a | S),$$

where Z' is a constant that does not depend on a.

The proof is deferred to the Appendix A.

Now, applying Lemma 1 and taking the log:

$$\log Pr(top(L^*) = a | S) = \log Pr(S_{I \setminus \{a\}}) + \sum_{x \neq a} (w^*(a, x) + Q) + \hat{Z}$$
$$= \hat{\hat{Z}} + BS(a | \boldsymbol{q}^*, S) + \log Pr(S_{I \setminus \{a\}}). \tag{4}$$

That is, we get that the log-likelihood of a being the true winner is the sum of two components (and a constant term $\hat{\hat{Z}}$ independent of a): the weighted Borda score of a with Grofman weights, and another term that only depends on the other alternatives.

Limit Analysis. As mentioned in the introduction, Young identified two important regimes: one where experts' accuracy is near guess level (p close to 0.5), and the other where accuracy is near-perfect (p close to 1).[4]

Elkind and Shah [8] have proven that the Borda winner is the most likely winner in the low accuracy regime. As for the high accuracy regime, they showed via a more detailed analysis that the most likely winner coincides with the Tideman rule. It is worth noting that in both cases, the MLE winner is a *refinement* of these rules; the most likely winner must be tied for first in these scores, but has its own, complicated tie-breaking rule.

In the following part we extend these results to the heterogeneous case, highlighting important caveats for these type of theorems in the heterogeneous setting.

3.3 Low Accuracy Regime

As in Young's observation, when all p_i are close to 0.5, the second term is a near-constant.

Lemma 5. *Denote* $\epsilon_i := |p_i - 0.5|$, *and let* $\epsilon := \max_i \epsilon_i$. *Then for any subset* $I' \subseteq I$, $Pr(S_{I'}) = \frac{1}{2}^{\binom{|I'|}{2}n-1} + O(\epsilon^2)$.

See Appendix C for the proof.

Observation 6. $q_i^* = \log \frac{0.5+\epsilon_i}{0.5-\epsilon_i} = \Theta(\epsilon_i) + O((\epsilon_i)^3)$.

I.e., the weight of each expert in the low accuracy regime is roughly linear in her accuracy. This follows by taking the first element of the Taylor expansion of $\log(x)$ for $x \cong 1$, see also [12].

Finally, we can use the log-likelihood decomposition from Eq. (4), together with the lemma and the observation above to show the following:

Proposition 7. *For any* a, b,

$$\begin{aligned}\log & Pr(top(L^*) = a|S) - \log Pr(top(L^*) = b|S) \\ &= BS(a|\boldsymbol{q}^*, S) - BS(b|\boldsymbol{q}^*, S) + O(\epsilon^2).\end{aligned}$$

Proof. From Eq. (4),

$$\begin{aligned}\log & Pr(top(L^*) = a|S) - \log Pr(top(L^*) = b|S) \\ &= BS(a|\boldsymbol{q}^*, S) + \log Pr(S_{I\setminus\{a\}}) \\ &- (BS(b|\boldsymbol{q}^*, S) + \log Pr(S_{I\setminus\{b\}})) \\ &= BS(a|\boldsymbol{q}^*, S) - BS(b|\boldsymbol{q}^*, S) + \log \frac{Pr(S_{I\setminus\{a\}})}{Pr(S_{I\setminus\{b\}})}\end{aligned}$$

Now, from Lemma 5, $\frac{Pr(S_{I\setminus\{a\}})}{Pr(S_{I\setminus\{b\}})} = \frac{\Theta(1)+O(\epsilon^2)}{\Theta(1)+O(\epsilon^2)} = 1 + O(\epsilon^2)$, so by Ob. 6,

$$\log \frac{Pr(S_{I\setminus\{a\}})}{Pr(S_{I\setminus\{b\}})} = \log(1 + O(\epsilon^2)) = O(\epsilon^2) + O(\epsilon^6) = O(\epsilon^2),$$

as required. □

Prop. 7 seems to suggest that in the low-accuracy regime, the (weighted) Borda winner is always the maximum likelihood winner in the limit. This is correct **if there is a Borda winner with *a* *constant margin*** over others, but we should be careful before concluding that this is always the case.

Example 8.

	p_i	L_i
expert 1	$1/2 + \epsilon$	$a \succ b \succ c$
expert 2	$1/2 + \epsilon$	$a \succ b \succ c$
expert 3	$1/2 + \epsilon + \epsilon^3$	$b \succ c \succ a$

This is the famous Borda draw, with a little exception: one of the experts is very slightly more accurate than the rest, and thus her weight is also larger by an additive factor of $\Theta(\epsilon^3)$, which is enough to break the tie in favor of b. Thus b is the Borda winner for any ϵ. Yet we argue that the most likely winner is item a, for sufficiently small ϵ. We show this in Appendix C by showing that $\log Pr(S_{I\setminus\{x\}})$ plays the major role in Eq. (4), when the Borda scores are nearly tied.

Still, in many cases the (weighted) Borda winner is the most likely winner. One case where the sufficient condition holds, is when $p_i = 0.5 + \alpha_i \cdot \epsilon$ for some vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$.

Indeed, it follows from Prop. 7 that if the Borda winner leads by a gap of at least $\omega(\epsilon^2)$, then as $\epsilon \to 0$ it will always become the most likely winner.

3.4 High Accuracy Regime

Elkind and Shah [8] showed that in the general homogeneous case the Tideman winner is the most likely one.

We next generalize this as well to the heterogeneous case, i.e. to show that the weighted Tideman (augmented with Grofman weights) with a sufficient margin, must be the most likely winner.

Recall that by Prop. 3, the ML winner is minimizing

$$\begin{aligned}\kappa(a) &= \prod_{b \in I/\{a\}} (1 + 2^{w^*(b,a) - w^*(a,b)}) \\ &= \sum_{J \subseteq I/\{a\}} 2^{\sum_{b \in J} w^*(b,a) - w^*(a,b)}\end{aligned}$$

We can now notice that the highest order term of this sum is for the J that contains all items that beat a, in which case the exponent is just the weighted Tideman score of player a. Indeed, using simple arguments, we can show that with a big enough gap the weighted Tideman winner is always the most likely winner:

Proposition 9. *For every item* a,

$$\log(\kappa(a)) \in [TS(a|\boldsymbol{q^*}, S), TS(a|\boldsymbol{q^*}, S) + m - 1]$$

Proof. $\kappa(a) = \sum_{J \subseteq I/\{a\}} 2^{\sum_{c \in J} w^*(c,a) - w^*(a,c)}$ and

$$2^{TS(a|\boldsymbol{q^*}, S)} \leq \sum_{J \subseteq I/\{a\}} 2^{\sum_{c \in J} w^*(c,a) - w^*(a,c)} \leq 2^{m-1} 2^{TS(a|\boldsymbol{q^*}, S)}$$

We conclude the proof by taking the log on both sides. □

This means that if a wins by a gap of at least $m - 1$ over all other items, we can tell that it is indeed the most likely winner with full certainty. This result is general: without assuming anything on the weights, if one of the items wins the Tideman rule by a big enough margin, it will always be the most likely winner.

One implication is the known result for homogeneous experts and unweighted Tideman, since the weighted Tideman score of

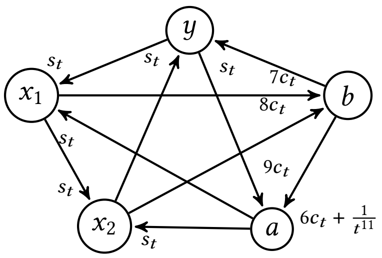
---

[4] Throughout this section we consider a constant dataset, and the Grofman weights.

experts with equal probabilities is equal to the unweighted Tideman score multiplied by their weight. As the weights approach infinity, the scores also approach infinity, and thus if any gap exists it will surpass any constant. We can also think of heterogeneous experts where p_i converge to 1 at different rates. In some cases the condition above is met and the weighted Tideman winner is the most-likely winner all along. For example, if $\forall i \in N,\ 1 - p_i = a_i \epsilon^{b_i}$ and $\epsilon \longrightarrow 0$, then the weighted Tideman winner is the ML winner for any ϵ, and $q_i = \log \frac{1 - a_i \epsilon^{b_i}}{a_i \epsilon^{b_i}} \cong -\log a_i + b_i \log \frac{1}{\epsilon}$, so in the limit weights are proportional to b_i, and the gap surpasses $m - 1$.

Again, we should be careful about claims of general convergence to the Tideman winner. Consider the following tournament-form game, with 10 experts and 5 items (the numbers indicate the weights of the victories, and are all close to the losing node, and also we denote $c_t = \log(t)$ and $s_t = 10c_t + \frac{1}{t^{11}}$ for easier display):

Example 10.



We first notice that this tournament is induced by 10 experts whose votes are *not* transitive orders (for example look at the triangle y, x_1, x_2), but since we are considering the Condorcet noise model this is a valid input S. Full table of the experts can be found in the Appendix I.

As in Example 8, there is a near-draw between a and b, that is decided in favor of b due to a single expert with slightly higher weight. Yet a is the most likely winner. Full proof of this counter-example can be found in the Appendix I.

4 HOW CLOSE IS THE TRUTH?

So far in this paper we have considered the most likely order or winner. Still, for ranking, it seems like this does not fully represent the quality of the results of the SWF. As mentioned in the introduction, in the machine learning literature we mainly consider minimizing the expected loss, and we offer a similar point of view here.

A natural way to measure the correctness of a ranking is the Kendall's-tau distance, which counts the number of switches needed to change from it to the ground truth. That is, we are interested in explicit bounds on $D(g, \boldsymbol{p}, \boldsymbol{q})$ for various voting rules g, and in particular its dependency on the number of items m.

A starting point for analyzing D is using the linearity of expectation:

$$\begin{aligned} D(g, \boldsymbol{p}, \boldsymbol{q}) &= E\left[d_{KT}(L^*, \hat{L}_g)\right] \\ &= E\left[\sum_{x,y \in \binom{I}{2}} \mathbb{1}_{\{x \succ_{L^*} y, y \succ_{\hat{L}_g} x\}}\right] \\ &= \sum_{x,y \in \binom{I}{2}} Pr\left(y \succ_{\hat{L}_g} x | x \succ_{L^*} y\right) \end{aligned}$$

Which offers an interesting perspective - which pairs of items are likely to be switched by rule g? Which are less likely? Consider a ranking that is selected uniformly at random. Since every switch occurs with probability $\frac{1}{2}$, we get expected distance of $\frac{1}{2}\binom{m}{2} = \Theta(m^2)$.

Another relatively simple statistic to check is the expected distance of the ground truth from the tournament S before resolution.[5] The switching probability for a pair of items is fully determined by the experts' expertise and the weights assigned to them, and does not change when comparing different pairs of items. Thus, if we denote $P(\boldsymbol{p}, \boldsymbol{q}) = Pr(y \succ_S x | x \succ_{L^*} y)$, we get that the expected distance is $P(\boldsymbol{p}, \boldsymbol{q})\binom{m}{2} = \Theta(m^2)$. Which means that with constant probability, the tournament gives a constant fraction of the number of errors of a random order.

While this result seems to mean there is no hope, it turns out that many voting rules actually have an asymptotically lower error rate.

4.1 Optimal rule and upper bounds

Optimal rule. The first natural question we have in this new framework is what is the optimal rule. We try to minimize the worst-case loss, but due to Theorem 1 in [19] we can instead minimize the expected posterior loss. We can easily find the posterior distribution over orders, using our Theorem 2:

Proposition 11. *The minimizer of* $\sum_{L' \in \mathcal{L}} d(L, L') e^{KS(L'|S, \boldsymbol{q}^*)}$ *minimizes the worst case expected error.*

A first surprising result about this rule is that it is not identical to the Kemeny-Young rule. In fact, it is not homogeneous, which means that duplicating the dataset can change the best order. It is, however, connected to some extent to the Kemeny-Young order. We prove that when duplicating the dataset enough times, the optimal rule becomes the Kemeny-Young rule. First denote $k \circ S$ to be k duplications of S (with the same items, and more experts). Now denote $k \circ \boldsymbol{q}$ similarly as k duplicates of the experts and their weights. Then we have:

Proposition 12. *For any dataset* S*, such that there is a single Kemeny winner, there exists a constant* $k(S)$ *such that* $\forall k > k(S)$, $\hat{L}_{KY}(S, \boldsymbol{q}^*) = \hat{L}_{KY}(k \circ S, k \circ \boldsymbol{q}^*) = \hat{L}_{OPT}(k \circ S, k \circ \boldsymbol{q}^*)$.

A proof for both propositions, as well as an example where the optimal rule is not equal to Kemeny-Young, are in Appendix E.

---

[5] This clearly might result in a non-transitive relation, but the Kendall-tau distance is still well-defined.

Between-items set. For every pair of items $x \succ_{L^*} y$, consider the following set of items: $B_{xy} = \{c \in I | x \succ_{L^*} c \succ_{L^*} y\}$ the set of items between these items. Notice that, if an item is ranked in L^* above both x, y, then it is equally likely to beat any of them in S (and likewise if ranked below both). But items B_{xy} are more likely to beat y and lose to x. More specifically, for any $c \notin B_{xy}$, for any vector of accuracies and weights $\boldsymbol{p}, \boldsymbol{q}$, the values $w_{\boldsymbol{q}}(x, c), w_{\boldsymbol{q}}(y, c)$ are equally distributed.

Item sum-scoring rules. Using this observation, we can start analyzing the item sum-scoring rules:

Notice that for a graph based item sum-scoring rule f and the Condorcet noise model, we have that

$$c \notin B_{xy} \Rightarrow E\left[V_f(w_{\boldsymbol{q}}(y,c)) - V_f(w_{\boldsymbol{q}}(x,c))\right] = 0;$$
$$c \in B_{xy} \Rightarrow E\left[V_f(w_{\boldsymbol{q}}(y,c)) - V_f(w_{\boldsymbol{q}}(x,c))\right] = \mu,$$

where μ is constant with respect to c.

Reversed Rule-Weight Pairs. A graph based sum-scoring rule, paired with a set of weights $\boldsymbol{q}$, are called *reversed* if $\mu \geq 0$. Since they give (on average) a (weakly) higher score for the worse item, these rules are (weakly) more likely to choose a ranking that is reversed to the ground-truth ranking. An example for a reversed rule is the random rule, or the Borda rule with weights that are negative to the Grofman weights. Using this definition, and our observation about item sum-scoring rules, we can get the following theorem:

THEOREM 13. *For any graph based item sum-scoring rule g, if g is not reversed,* $D(g, \boldsymbol{p}, \boldsymbol{q}) = O(m\sqrt{m \log(m)})$.

PROOF SKETCH. The theorem is based on bounding the probability that the rule will switch a pair of items based on the number of items between these items. This number represents the size of the "signal" the rule gets for this pair, where the rest of the items only make the estimation noisy. For signal-to-noise ratio of at least $\Theta(\sqrt{m \log(m)})$ we get that the probability is decreasing quickly enough to bound it effectively. All the too-close pairs of items contribute to the expected distance. The proof is deferred to Appendix E. □

Any of the above mentioned sum-scoring rules, assuming $\forall i, p_i > \frac{1}{2}$ with uniform weights or Grofman weights, is not reversed.

Kemeny-Young. Finally, we prove an upper bound for the Kemeny-Young rule[6] with optimal weights:

THEOREM 14. *If* $\exists i \in N$ *s.t.* $p_i \neq \frac{1}{2}$, $D(KY, \boldsymbol{p}, \boldsymbol{q}^*) = O(m \log m)$.

PROOF. The proof is based on the (wrong) proof by [5], extended to heterogeneous experts. Essentially, we prove that rankings with distance that is of order $m \log m$ are very unlikely to have the highest Kemeny score, and then use the union bound, similarly to [5], but adapted to the heterogeneous setting and without the error. We first prove that any order too far away is likely to have lower Kemeny score than the ground-truth order. Let $L \in \mathcal{L}(I)$. Define $A = \{(a,b) \in I^2 | a \succ_{L^*} b, b \succ_L a\}$ the set of discordant pairs, and notice that $|A| = d_{KT}(L, L^*)$. Recall $\forall x, y \in I, W = w^*(x,y) + w^*(y,x)$. Then

$$Pr(KS(L^*) \leq KS(L)) = Pr\left(\sum_{a,b \in A} w^*(a,b) - w^*(b,a) \leq 0\right)$$
$$= Pr\left(\sum_{a,b \in A} 2w^*(a,b) - W \leq 0\right)$$
$$= Pr\left(\sum_{a,b \in A} w^*(a,b) \leq \frac{1}{2} W d_{KT}(L, L^*)\right)$$

Consider the following problem: a set of experts of size $d_{KT}(L, L^*) \cdot n$ whom have the same accuracies as d_{KT} duplicates of our initial $\boldsymbol{p}$, compare a single pair of items, x, y. Assume that in truth, $x \succ y$. Then the exact event of the weighted majority making a mistake is $\sum_{x \succ_i y} w_i^* \leq \frac{1}{2} W d_{KT}(L, L^*)$. Now split the pairs to sets denoted $C_{a,b}$, one for each of the $a, b \in A$, so that every set contains exactly one duplicate (i.e. experts with the same accuracy) of the original experts. We denote for every set S of experts $w_S^*(x,y) := \sum_{i \in S: x \succ_i y} w_i^*$. We can conclude that the event of the mistake of these new experts is $\sum_{a,b \in A} w_{C_{a,b}}^*(x,y) \leq \frac{1}{2} W d_{KT}(L, L^*)$, and the left hand side has the same distribution of the left hand side the in inequality above (since $w_{C_{a,b}}^*(x,y)$ has the same distribution as $w^*(a,b)$, and they are all independent).

Essentially, we proved above the the event $\sum_{a,b \in A} w^*(a,b) \leq \frac{1}{2} W d_{KT}(L, L^*)$ has the same probability as a set of $d_{KT}(L, L^*)$ duplicates of the experts succeeding in a single comparison. We can now directly use results from the binary case by [2] to bound this probability. Denote for any set of experts S, $\Phi(S) := \sum_{i \in S} q_i^* (p_i - \frac{1}{2})$. The results for the binary case states that the probability of a mistake when using the weighted answers of set S is upper bounded by $e^{-\frac{1}{2}\Phi(S)}$. Notice that if $\exists i \in S, p_i \neq \frac{1}{2}, \Phi(S) > 0$. Also, we denote $k \circ S$ a duplication of the experts (i.e, k independent experts with the same accuracy as in S), and notice that $\Phi(k \circ S) = k\Phi(S)$. Finally, notice that $\Phi(N)$ is independent of m. Now we get

$$Pr(\hat{L}_{KY} = L) = Pr(L \in argmax_{L'} KS(L')) \leq Pr(KS(L) \geq KS(L^*))$$
$$\leq e^{-\frac{1}{2}\Phi(d_{KT}(L,L^*) \circ N)} = e^{-\frac{1}{2} d_{KT}(L,L^*) \Phi(N)}$$

We denote $d := \frac{4}{\Phi(N)} m \log m$ and $X := d_{KT}(\hat{L}_{KY}, L^*)$. Now we use the union bound:

$$Pr(X > d) = Pr(d_{KT}(\hat{L}_{KY}, L^*) > d) \leq \sum_{L: d_{KT}(L,L^*) > d} Pr(\hat{L}_{KY} = L)$$
$$\leq m! e^{-\frac{1}{2} d \Phi(N)} = m! e^{-2m \log m} \leq \frac{1}{m!}$$

using this can obtain:

---

[6] The proof used by [5] is wrong. It uses the union bound over all permutations of distance at least cm for a constant c, but substitutes a bound over the number of all permutations of distance at most cm. The correct number is significantly higher, which makes the proof unfixable. See appendix G for a full explanation.

$$\begin{aligned} D(KY,\boldsymbol{p},\boldsymbol{q}^*) &= E[X] \\ &= Pr(X \le d)E[X|X \le d] + Pr(X > d)E[X|X > d] \\ &\le d + \frac{m^2}{m!} = O\left(m \log m\right) \end{aligned}$$

As needed. □

4.2 Lower bounds

Maximin rule. The simplest lower bound comes for the maximin rule. Assume there are n experts, each with (possibly) different accuracy. Then the opposition for each item can be one of 2^n possible values (since the experts may have different accuracies). Using the pigeonhole principle, this means that there are at least $\frac{m}{2^n}$ items that have the same value of opposition, and thus maximin will randomly assign one above the other. This creates in expectation $\frac{1}{2}\binom{2^{-n}m}{2} = \Theta(m^2)$ errors, which is also the upper bound for any rule. Notice that this holds even when we have a perfect expert ($p = 1$)!

We now look again into the item sum-scoring rules:

Theorem 15. *For any graph based item sum-scoring rule g, where no expert has $p_i \in \{0,1\}$, $D(g,\boldsymbol{p},\boldsymbol{q}) = \Omega(m\sqrt{m})$.*

Note that the theorem holds even for a single expert. The proof can be found in Appendix F.

Proof sketch. The switch probabilities are decreasing when the items are further away from each other (in the sense that the B_{xy} is large). The switch probabilities do not approach zero when they are separated by $O(\sqrt{m})$ items, and decrease polynomially (which is needed for the upper bound) if the separation is at least $\Omega(\sqrt{m\log(m)})$. □

Theorems 13 and 15 apply in particular to Borda, Copeland and Tideman. The proof relies on a similar separation of pairs by their distance from one another. This also explains how, despite the tournament itself having a huge distance ($\Theta(m^2)$) from the ground-truth, the voting rules can actually close the gap significantly—most of the distance from the tournament to the ground truth comes from 'distant pairs' of items, and voting rules are more likely to correct these mistakes. These methods can also be applied when analyzing other distance measures.

Our last claim is a general lower bound for any voting rule, but a single expert.

Proposition 16. *For any rule g, where no expert has $p_i \in \{0,1\}$, $D(g,\boldsymbol{p},\boldsymbol{q}) = \Omega(m)$*

The proof of this Proposition relies on the optimal rule g being neutral, and proving that the probability of switching pairs of items that have no items between them (distance 1 from each other). It can be seen in Appendix F.

We can summarize our results in the following table:

	Mnmx	ISS	KY	OPT
UB	m^2	$m\sqrt{m\log(m)}$	$m\log m$	m (#)
LB	m^2	$m\sqrt{m}$	m	m

All results in the table are in asymptotic rate. Mnmx is maximin, ISS are the item sum-scoring rules, KY is Kemeny-Young, OPT is the optimal rule. The result marked with # is from [10].

4.3 Open Questions

The first open questions are the obvious gaps left in our results: What is the expected distance of the Kemeny-Young method, and is it like the optimal? Can the $\sqrt{\log(m)}$ gap be closed?

While we mainly discussed the asymptotic rate of increase in distance with respect to the number of items, a very natural question would be how do the experts affect expected distance. For this we can generalize some results from the binary case:

Observation 17. *If S is induced from L^* via the Condorcet noise model, the probability that there is at least one wrong pair in S is at most $\binom{m}{2}e^{-\frac{1}{2}\Phi}$, where $\Phi = \sum_{i\in N}(p_i - \frac{1}{2})w_i^*$.*

This trivially upper bounds the error probability of some voting rule based on S.[7] It would be interesting to further tighten this bounds for specific rules, and find some lower bounds as well. We conjecture that the bounds for the binary case are tight for the full ranking case. Another question due to Berend and Kontorovich is what happens to this dependence on the accuracy and number of experts when the rules don't have exact values of them (but instead use estimations, or uniform weights).

5 CONCLUSION

We set out to generalize known results on maximum likelihood voting rules to the heterogeneous experts case of the Condorcet noise model. This generalization is straight-forward for social-welfare functions (which return the most likely *ranking*), but requires some care when considering social-choice functions that return the most likely winner.

However (somewhat surprisingly), the Kemeny voting rule, which maximizes the likelihood of the true ranking, does not minimize the distance to it. It does, however, at least almost optimal asymptotically.

A family of some known rules have surprisingly good performance in this setting: they make asymptotically less mistakes than the experts themselves, by efficiently identifying distant pairs of items. We extended the lower bound for the multi-voter problem, and showed that the maximin rule has the worst possible asymptotic expected distance.

.

REFERENCES

[1] Hossein Azari Soufiani, David C Parkes, and Lirong Xia. 2014. A statistical decision-theoretic framework for social choice. *Advances in Neural Information Processing Systems* 27 (2014).
[2] Daniel Berend and Aryeh Kontorovich. 2015. A finite sample analysis of the Naive Bayes classifier. *J. Mach. Learn. Res.* 16, 1 (2015), 1519–1545.
[3] Philip J Boland. 1989. Majority systems and the Condorcet jury theorem. *Journal of the Royal Statistical Society Series D: The Statistician* 38, 3 (1989), 181–189.
[4] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. 2016. *Handbook of computational social choice.* Cambridge University Press.
[5] Mark Braverman and Elchanan Mossel. 2008. Noisy sorting without resampling. In *19th Annual ACM-SIAM Symposium on Discrete Algorithms.* 268–276.

---

[7] Caragiannis et al. [6] named the rules that always choose the pairwise winner graph if it exists as PM-c and explored their properties. They include the Kemeny-Young method, as well as the Copeland method.

[6] Ioannis Caragiannis, Ariel D Procaccia, and Nisarg Shah. 2016. When do noisy votes reveal the truth? *ACM Transactions on Economics and Computation (TEAC)* 4, 3 (2016), 1–30.

[7] Vincent Conitzer and Tuomas Sandholm. 2005. Common voting rules as maximum likelihood estimators. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*. 145–152.

[8] Edith Elkind and Nisarg Shah. 2014. Electing the Most Probable Without Eliminating the Irrational: Voting Over Intransitive Domains.. In *UAI*. 182–191.

[9] Edith Elkind and Arkadii Slinko. 2016. Rationalizations of voting rules. In *Handbook of Computational Social Choice*, Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (Eds.). Cambridge University Press.

[10] Barbara Geissmann, Stefano Leucci, Chih-Hung Liu, and Paolo Penna. 2017. Sorting with recurrent comparison errors. *arXiv preprint arXiv:1709.07249* (2017).

[11] Barbara Geissmann, Stefano Leucci, Chih-Hung Liu, and Paolo Penna. 2018. Optimal sorting with persistent comparison errors. *arXiv preprint arXiv:1804.07575* (2018).

[12] Bernard Grofman, Guillermo Owen, and Scott L Feld. 1983. Thirteen theorems in search of the truth. *Theory and Decision* 15, 3 (1983), 261–278.

[13] John G Kemeny. 1959. Mathematics without numbers. *Daedalus* 88, 4 (1959), 577–591.

[14] Dmitriy Kunisky, Daniel A Spielman, and Xifan Yu. 2024. Inference of rankings planted in random tournaments. *arXiv preprint arXiv:2407.16597* (2024).

[15] Edoardo Manino, Long Tran-Thanh, and Nicholas R Jennings. 2019. On the efficiency of data collection for multiple Naïve Bayes classifiers. *Artificial Intelligence* 275 (2019), 356–378.

[16] Jerzy Neyman and Egon Sharpe Pearson. 1933. IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231, 694-706 (1933), 289–337.

[17] Ariel D Procaccia, Nisarg Shah, and Yair Zick. 2016. Voting rules as error-correcting codes. *Artificial Intelligence* 231 (2016), 1–16.

[18] Lirong Xia. 2014. Deciphering young's interpretation of condorcet's model. mimeo.

[19] Lirong Xia. 2016. Bayesian Estimators As Voting Rules. In *Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence (UAI)*. AUAI Press, Corvallis, Oregon, USA, 785–794. https://dblp.org/rec/conf/uai/Xia16

[20] H Peyton Young. 1988. Condorcet's theory of voting. *American Political science review* 82, 4 (1988), 1231–1244.

[21] H Peyton Young and Arthur Levenglick. 1978. A consistent extension of Condorcet's election principle. *SIAM Journal on applied Mathematics* 35, 2 (1978), 285–300.

A PROOFS OF SECTION 3.2

Theorem 4. *Suppose that the ground truth L^* is sampled from a uniform prior on $\mathcal{L}(I)$. Then for any item $a \in I$ and any heterogeneous accuracy vector $\boldsymbol{p}$, we have that*

$$Pr_{\boldsymbol{p}}(top(L^*) = a|S) = Z' \cdot P_{\boldsymbol{p}}^{ML}(a|S),$$

where Z' is a constant that does not depend on a.

Proof. For any $a \in I$

$$\begin{aligned} Pr(top(L^*) = a|S) &= Pr(S|top(L^*) = a)\frac{Pr(top(L^*) = a}{Pr(S)} && \text{(Bayes rule)} \\ &= \frac{1/m}{Pr(S)} \sum_{L:top(L)=a} Pr(S|L)Pr(L|top(L) = a) \end{aligned}$$

For every profile S and order $L \in \mathcal{L}$ with $top(L) = a$ we have

$$Pr(S|L) = \left(\prod_{x \neq a} Pr(S_{ax}|L_{ax})\right) Pr(S_{I\setminus\{a\}}|L_{I\setminus\{a\}}) = \left(\prod_{x \neq a} Pr(S_{ax}|a \succ_L x)\right) Pr(S_{I\setminus\{a\}}|L_{I\setminus\{a\}})$$

Also,

$$\begin{aligned} &\sum_{L:top(L)=a} Pr(S_{I\setminus\{a\}}|L_{I\setminus\{a\}})Pr(L|top(L) = a) \\ &\qquad = \sum_{L' \in \mathcal{L}(I\setminus\{a\})} Pr(S_{I\setminus\{a\}}|L')Pr(L') = Pr(S_{I\setminus\{a\}}). \end{aligned}$$

Therefore we get

$$\begin{aligned} Pr(top(L^*) = a|S) &= \frac{1/m}{Pr(S)} \sum_{L:top(L)=a} Pr(S|L)Pr(L|top(L) = a) \\ &= \frac{1/m}{Pr(S)} \sum_{L:top(L)=a} \left(\prod_{x \neq a} Pr(S_{ax}|a \succ_L x)\right) Pr(S_{I\setminus\{a\}}|L_{I\setminus\{a\}})Pr(L|top(L) = a) \\ &= \frac{1/m}{Pr(S)} \left(\prod_{x \neq a} Pr(S_{ax}|a \succ_L x)\right) \sum_{L:top(L)=a} Pr(S_{I\setminus\{a\}}|L_{I\setminus\{a\}})Pr(L|top(L) = a) \\ &= \frac{1}{m \cdot Pr(S)} Pr(S_{I\setminus\{a\}}) \prod_{x \neq a} Pr(S_{ax}|a \succ x) && (5) \\ &= Z' \cdot Pr(S_{I\setminus\{a\}}) \prod_{x \neq a} Pr(S_{ax}|a \succ x) = Z' \cdot P_{\boldsymbol{p}}^{ML}(a|S), \end{aligned}$$

where Z' is a constant that does not depend on the alternative a. ◻

Proposition 3. $Pr(top(L^*) = a|S) = \frac{1}{\kappa(a)}$.

Proof.

$$\begin{aligned} Pr(top(L^*) = a|S) &= Pr(\forall b \in I/\{a\},\, a \succ_{L^*} b|S) \\ &= \prod_{b \in I/\{a\}} Pr(a \succ_{L^*} b|S) = \prod_{b \in I/\{a\}} Pr(a \succ_{L^*} b|S_{ab}) \\ &= \prod_{b \in I/\{a\}} \frac{Pr(a \succ_{L^*} b)}{Pr(S_{ab})} Pr(S_{ab}|a \succ_{L^*} b) \end{aligned}$$

Now, using the uniformly random prior we get

$$
\begin{aligned}
&= \prod_{b \in I/\{a\}} \frac{Pr(S_{ab}|a \succ_{L^*} b)}{Pr(S_{ab}|a \succ_{L^*} b) + Pr(S_{ab}|b \succ_{L^*} a)} \\
&= \prod_{b \in I/\{a\}} \frac{1}{1 + \frac{Pr(S_{ab}|a \succ_{L^*} b)}{Pr(S_{ab}|a \succ_{L^*} b)}} \\
&= \prod_{b \in I/\{a\}} \frac{1}{1 + \frac{\prod_{i:a \succ_i b} 1-p_i \prod_{i:b \succ_i a} p_i}{\prod_{i:a \succ_i b} p_i \prod_{i:b \succ_i a} 1-p_i}} \\
&= \frac{1}{\prod_{b \in I/\{a\}} (1 + \prod_{i:a \succ_i b} \frac{1-p_i}{p_i} \prod_{i:b \succ_i a} \frac{p_i}{1-p_i})}
\end{aligned}
$$

Using this we look at the denominator. The most likely winner is the player that **minimizes** this score. Recall that the Grofman weights are $2^{q_i} = \frac{p_i}{1-p_i}$, and using this we obtain

$$
\begin{aligned}
\frac{1}{\kappa(a)} &= \frac{1}{\prod_{b \in I/\{a\}} (1 + 2^{w^*(b,a) - w^*(a,b)})} \\
&= \frac{1}{\prod_{b \in I/\{a\}} (1 + \prod_{i:a \succ_i b} 2^{-q_i} \prod_{i:b \succ_i a} 2^{q_i})} \\
&= \frac{1}{\prod_{b \in I/\{a\}} (1 + \prod_{i:a \succ_i b} \frac{1-p_i}{p_i} \prod_{i:b \succ_i a} \frac{p_i}{1-p_i})} \\
&= Pr(top(L^*) = a|S),
\end{aligned}
$$

as required. ☐

B ANALYZING YOUNG'S CALCULATIONS

In pp. 1237-1238, Young makes the following calculations, for a particular profile of $n = 100$ homogeneous experts and three items (translated to our notation):

$$
\begin{aligned}
Pr(x \succ_{L^*} y|vote) &= Pr(vote|x \succ_{L^*} y) \frac{Pr(x \succ_{L^*} y)}{Pr(vote)} \\
&= \frac{p^{w(x,y)}(1-p)^{w(y,x)}}{p^{w(x,y)}(1-p)^{w(y,x)} + (1-p)^{w(x,y)} p^{w(y,x)}}.
\end{aligned} \tag{Y1}
$$

Next, Young states that

$$
Pr(top(L^*) = x|vote) = Pr(x \succ y|vote) Pr(x \succ z|vote). \tag{Y2}
$$

Young does not clarify what 'vote' means. If 'vote' stands for the full voting profile S then both Eq. (Y1) and Eq. (Y2) hold only approximately, due to independence between terms. See discussion in [9].[8]

However Eq. (Y1) does hold if 'vote' only considers the part of the profile that compares x and y, i.e., S_{xy}.

Following Young's two equations (Y1) and (Y2), denote: $P^Y(a|vote) := \prod_{x \neq a} Pr(a \succ x|S_{ax})$, as Young's estimation of the winning probability of alternative a.

By the discussion above, we know that $Pr(a \succ x|S_{ax})$ is only an approximation of $Pr(a \succ x|S)$; and that $\prod_{x \neq a} Pr(a \succ x|S)$ is only an approximation of $Pr(top(L^*) = a|S)$. So a-priori it is not clear under what conditions, if at all, $Pr(top(L^*) = a|S) = P^Y(a|vote)$. In fact, P^Y may not even be a valid probability distribution.

We thus define $P^{NY}(a|vote) := \frac{P^Y(a|vote)}{\sum_x P^Y(x|vote)}$ as the *Normalized Young* probability estimation.

Normalized Young works for three alternatives. Strangely enough, the following holds:

Proposition 18. *For three alternatives and heterogeneous experts,* $\frac{P^{ML}(a|S)}{P^{ML}(b|S)} = \frac{P^Y(a|vote)}{P^Y(b|vote)}$ *for every two items* a, b.

---

[8] See also discussion in [18], which suggests an alternative noise model where Young's calculations are justified.

Proof.

$$\begin{aligned}\frac{P^Y(a|vote)}{P^Y(b|vote)} &= \frac{Pr(ab|S_{ab})Pr(ac|S_{ac})}{Pr(ba|S_{ab})Pr(bc|S_{bc})}\\ &= \frac{Pr(S_{ab}|ab)Pr(S_{ac}|ac)}{Pr(S_{ab}|ba)Pr(S_{bc}|bc)}\frac{Pr(S_{ab})Pr(S_{bc})}{Pr(S_{ab})Pr(S_{ac})}\\ &= \frac{Pr(S_{ab}|ab)Pr(S_{ac}|ac)Pr(S_{bc})}{Pr(S_{ab}|ba)Pr(S_{bc}|bc)Pr(S_{ac})}\\ &= \frac{P^{ML}(a|S)}{P^{ML}(b|S)}, && \text{(by Eq. (3))}\end{aligned}$$

as required. □

Corollary 19. *For three alternatives and heterogeneous experts,*

$$Pr(top(L^*) = a|S) = P^{NY}(a|vote)$$

for every item $a \in \{x, y, x\}$.

Proof. Denote $q_a := Pr(top(L^*) = a|S), p_a := P^{NY}(a|vote)$, and note that

$$\frac{p_a}{p_b} = \frac{P^{NY}(a|vote)}{P^{NY}(b|vote)} = \frac{P^Y(a|vote)}{P^Y(b|vote)}.$$

By Prop. 18 and Theorem 4, we get the following three equations:

$$\frac{p_x}{p_y} = \frac{Z' \cdot q_x}{Z' \cdot q_y} = \frac{q_x}{q_y}; \quad \frac{p_x}{p_z} = \frac{Z' \cdot q_x}{Z' \cdot q_z} = \frac{q_x}{q_z}; \quad \frac{p_y}{p_z} = \frac{Z' \cdot q_y}{Z' \cdot q_z} = \frac{q_y}{q_z};$$

We also have the fourth equation by normalization: $p_x + p_y + p_z = 1 = q_x + q_y + q_z$. Therefore the system is overdetermined, and the unique solution is $p_a = q_a$ for all a. □

Therefore, for three alternatives, despite using two unjustified approximations in the process, Young's formula $P^Y(a|vote)$ indeed finds the most likely winner! Unfortunately, this does no longer hold for four alternatives or more, even for homogeneous experts.

C PROOFS OF SECTION 3.2

C.1 Low Accuracy

Lemma 5. *Denote $\epsilon_i := |p_i - 0.5|$, and let $\epsilon := \max_i \epsilon_i$. Then for any subset $I' \subseteq I$, $Pr(S_{I'}) = \left(\frac{1}{2}\right)^{\binom{|I'|}{2}n-1} + O(\epsilon^2)$.*

The proof is similar to the two-alternative case, except we now have $|I'|!$ orders, and we pair together opposite rankings.

Proof. For any ranking $L \in \mathcal{L}(I')$ let $\overline{L}$ be the opposite ranking. Let $a, b \in I'$ be two arbitrary items. Clearly if $a \succ_L b$ then $b \succ_{\overline{L}} a$. Denote by $\mathcal{L}_{ab} \subseteq \mathcal{L}(I')$ all rankings where $a \succ_L b$.

$$\begin{aligned}
Pr(S_{I'}) &= \frac{1}{|I'|!} \sum_{L \in \mathcal{L}(I')} Pr(S_{I'}|L) \\
&= \frac{1}{|I'|!} \left(\sum_{L \in \mathcal{L}_{ab}} Pr(S_{I'}|L) + \sum_{L \in \mathcal{L}_{ba}} Pr(S_{I'}|L) \right) \\
&= \frac{1}{|I'|!} \sum_{L \in \mathcal{L}_{ab}} (Pr(S_{I'}|L) + Pr(S_{I'}|\overline{L})) \\
&= \frac{1}{|I'|!} \sum_{L \in \mathcal{L}_{ab}} \left(\prod_{x,y \in \binom{I'}{2}} Pr(S_{xy}|L) + \prod_{x,y \in \binom{I'}{2}} Pr(S_{xy}|\overline{L}) \right) \\
&= \frac{1}{|I'|!} \sum_{L \in \mathcal{L}_{ab}} \left(\prod_{x,y \in \binom{I'}{2}} \prod_{i: x \succ_i y} p_i \prod_{i: y \succ_i x} (1 - p_i) \ + \prod_{x,y \in \binom{I'}{2}} \prod_{i: x \succ_i y} (1 - p_i) \prod_{i: y \succ_i x} p_i \right) \\
&= \frac{1}{|I'|!} \sum_{L \in \mathcal{L}_{ab}} \left(\prod_{x,y \in \binom{I'}{2}} \prod_{i: x \succ_i y} (\frac{1}{2} + \epsilon_i) \prod_{i: y \succ_i x} (\frac{1}{2} - \epsilon_i) \ + \prod_{x,y \in \binom{I'}{2}} \prod_{i: x \succ_i y} (\frac{1}{2} - \epsilon_i) \prod_{i: y \succ_i x} (\frac{1}{2} + \epsilon_i) \right)
\end{aligned}$$

Now in each product, ϵ_i appears as a first-order term only when multiplied by $k - 1$ other constant $(\frac{1}{2})$ terms, where $k := \binom{|I'|}{2} n$. Thus

$$\prod_{x,y \in \binom{I'}{2}} \prod_{i: x \succ_i y} (\frac{1}{2} + \epsilon_i) \prod_{i: y \succ_i x} (\frac{1}{2} - \epsilon_i) \tag{6}$$

$$= (\frac{1}{2})^k + \sum_{x,y \in \binom{I'}{2}} \sum_{i: x \succ_i y} \left(\frac{1}{2}\right)^{k-1} \epsilon_i + \sum_{x,y \in \binom{I'}{2}} \sum_{i: y \succ_i x} \left(\frac{1}{2}\right)^{k-1} (-\epsilon_i) + O(\epsilon^2) \tag{7}$$

Every ϵ_i term that appears with a positive sign in $Pr(S_{I'}|L)$, appears with a negative sign in $Pr(S_{I'}|L)$ (see Eq. (7)), and vice versa. Thus all the first-order terms in $Pr(S_{I'})$ cancel out, and we get

$$Pr(S_{I'}) = \left(\frac{1}{2}\right)^{k-1} + O(\epsilon^2),$$

as required. □

Proposition 20. *There is an example with 3 experts and 3 items, where the competence of all experts is bounded by* $0.5 + \epsilon$*, the Borda winner is b for any* ϵ*, and the maximum likelihood winner is a for sufficiently low* ϵ*.*

Proof. Consider Example 8.

$$\begin{aligned}
Pr(S_{I \setminus \{a\}}) &= \frac{1}{2}(1/2 + \epsilon)^2 (1/2 + \epsilon + \epsilon^3) \\
+\frac{1}{2}(1/2 - \epsilon)^2 (1/2 - \epsilon - \epsilon^3) &= \frac{1}{8} + \frac{3}{2}\epsilon^2 + O(\epsilon^3) \\
Pr(S_{I \setminus \{b\}}) &= \frac{1}{2}(1/2 + \epsilon)^2 (1/2 - \epsilon - \epsilon^3) \\
+\frac{1}{2}(1/2 - \epsilon)^2 (1/2 + \epsilon + \epsilon^3) &= \frac{1}{8} + O(\epsilon^{2.5})
\end{aligned}$$

The asymptotic behavior of taking the log over both sided is the same as without the log (because of the constant term inside of it). Since the difference in the Borda score term has order $O(\epsilon^{2.5})$ while the difference in the other term has order $\Omega(\epsilon^2)$, as we take $\epsilon \longrightarrow 0$ we will eventually have that item a is always the most likely winner, while item b is still the Borda winner. This contradicts the idea that the Borda winner is always the most likely winner for low enough accuracies.

□

D COMPARISON TO CONITZER AND SANDHOLM

In a fundamental paper, Conitzer and Sandholm [7] characterize which voting rules (both SCFs and SWFs) can be maximum-likelihood estimators for some i.i.d vote distribution.

Their Lemma 1 states that a necessary condition on rule f for such a distribution to exist, is that the rule obeys *reinforcement* [21]. That is, that if $f(S^1) = a$ and $f(S^2) = a$, then $f(S^1 + S^2) = a$ as well. The lemma is stated both for SCF and SWF.

Since Maximin violates the condition (as they show in their Theorem 6), the authors conclude it is not a maximum-likelihood estimator for any i.i.d. vote distribution. This result seems to contrast with Young's informal statement about Maximin being the solution for the Condorcet noise model with p close to 1 (which we know to hold for $m = 3$ items).

An initial possible explanation may be that Theorem 6 in [7] only provides a violation example for $m = 4$ items. However this cannot solve the discrepancy since it is also possible to construct a violating Maximin example for $m = 3$ (see Appendix H).

The actual reason for this discrepancy is that the noise models considered in [7] for SCFs are only models that have a *ground-truth winner.* Since the Condorcet noise model has a *ground-truth ranking* over items, results from [7] do not apply.

E PROOFS OF SUBSECTION 4.1

Proposition 11. *The minimizer of* $\sum_{L' \in \mathcal{L}} d(L, L') e^{KS(L')}$ *minimizes the expected error.*

Proof. We want to minimize the worst case expected error:

$$\hat{L}_{OPT} = Sup_{L^*} min_{\hat{L}} \left(E[d(\hat{L}, L^*)|S] \right)$$

The first tool is from [19], theorem 1, which will allow us to replace the worst-case analysis with a Bayesian posterior error, by using the uniform prior. Translating our papers' notations to the notations in that paper, we get that the following framework: $\mathcal{F} = \left(\mathcal{M}^{\boldsymbol{p}}_{Co}, \mathcal{L}, d_{KT} \right)$ where given some vector $\boldsymbol{p}$, $\mathcal{M}^{\boldsymbol{p}}_{Co}$ is the heterogeneous noise model described in this paper. It is simple to verify that if we use the same permutations $\sigma = \sigma_\Theta = \sigma_D$ for the noise model, decision space, and Kendall-tau distance, we get exactly the neutrality condition on the theorem. Also, since we use the entire permutation space as our parameters, the parameter connectivity condition is also trivially satisfied.

The result of using this theorem is that we can use the uniform prior, and instead of minimizing worst-case expected distance, we can minimize posterior expected distance, which is simpler:

$$E[d_{KT}(\hat{L}, L^*)|S] = \sum_{L^*} P(L^*|S) d_{KT}(\hat{L}, L^*)$$

The second and last step is to use our theorem 2 and exponentiating both sides, which gives us

$$\sum_{L^*} P(L^*|S) d_{KT}(\hat{L}, L^*) = \sum_{L^*} d_{KT}(\hat{L}, L^*) e^{KS(L^*)+Z}$$

Which has the same minimizer as

$$\sum_{L^*} d_{KT}(\hat{L}, L^*) e^{KS(L^*)}$$

As needed.

□

Optimal rule is not Kemeny-Young. Using the equation from proposition 11, we can show that there are some cases where the Kemeny-Young rule and the optimal rule disagree. The example below requires 3 items and 5 experts. We could not find smaller examples, and could not prove if they exist or not. The full calculation for proving these are indeed the winners are omitted, but brute-force computer calculation can verify these calculations without problem. We also note that in this example all experts have full rankings as output, and the same accuracy:

expert number			
2	A	B	C
2	C	A	B
1	B	C	A

In this example there are two winning Kemeny orders: ABC and CAB. The optimal order is ACB.

Proposition 12. *For any dataset* S*, such that there is a single Kemeny winner, there exists a constant* $k(S)$ *such that* $\forall k > k(S)$, $\hat{L}_{KY}(S, \boldsymbol{q}^*) = \hat{L}_{KY}(k \circ S, k \circ \boldsymbol{q}^*) = \hat{L}_{OPT}(k \circ S, k \circ \boldsymbol{q}^*)$.

Proof. The first equality comes is exactly the homogeneity of the Kemeny-Young rule. For the second equality, let L be any order and consider the Kemeny score of L under the duplicated dataset:

$$KS(L|k \circ S, k \circ \boldsymbol{q}^*) = \sum_{x \succ_L y} w^*_{k \circ S}(x,y) = \sum_{x \succ_L y} k w^*_S(x,y) = k \sum_{x \succ_L y} w^*_S(x,y) = k \cdot KS(L|S, \boldsymbol{q}^*)$$

Now using proposition 11, we can see that:

$$\hat{L}_{OPT}(k \circ S) = min_{\hat{L}} \sum_{L \in \mathcal{L}} d_{KT}(L, \hat{L}) e^{k \cdot KS(L|S,\boldsymbol{q}^*)}$$

Since the d_{KT} is a distance measure, it only equals zero when both elements are equal. Denote the highest and second highest Kemeny scores as KS_1, KS_2, respectively, and let $L' \in \mathcal{L}$ different from $\hat{L}_{KY}$. We now have that:

$$\sum_{L \in \mathcal{L}} d_{KT}(L, \hat{L}_{KY}) e^{k \cdot KS(L|S,\boldsymbol{q}^*)} \leq \sum_{\substack{L \in \mathcal{L} \\ L \neq \hat{L}_{KY}}} \binom{m}{2} e^{k \cdot KS_2} = (m! - 1) \binom{m}{2} e^{k \cdot KS_2} \leq e^{k \cdot KS_1} \leq \sum_{L \in \mathcal{L}} d_{KT}(L, L') e^{k \cdot KS(L|S,\boldsymbol{q}^*)}$$

For any $k \geq \frac{\log(\binom{m}{2}(m!-1))}{KS_1 - KS_2}$, so for any k large enough the optimal rule is the Kemeny-Young rule, as needed.

□

Theorem 13. *For any graph based item sum-scoring rule g, if g is not reversed, $D(g, \boldsymbol{p}, \boldsymbol{q}) = O(m\sqrt{m \log(m)})$.*

Proof. Let $x \succ_{L^*} y$, $c \in B_{xy}$, and recall the definition of $\mu = E\left[V_g(w_{\boldsymbol{q}}(x,c)) - V_g(w_{\boldsymbol{q}}(x,c))\right]$. Using the fact that g is a graph based sum-scoring rule, we get that

$$\begin{aligned}
Pr\left(y \succ_{\hat{L}_g} x | x \succ_{L^*} y\right) &= Pr\left(\sum_{a \neq y} V_g(w_{\boldsymbol{q}}(y,a)) \geq \sum_{b \neq x} V_g(w_{\boldsymbol{q}}(x,b)) | x \succ_{L^*} y\right) \\
&= Pr\left(V_g(w_{\boldsymbol{q}}(y,x)) - V_g(w_{\boldsymbol{q}}(x,y)) + \sum_{a \neq x,y} V_g(w_{\boldsymbol{q}}(y,a)) - \sum_{b \neq x,y} V_g(w_{\boldsymbol{q}}(x,b)) - (d-1)\mu - \mu \geq -(d-1)\mu - \mu\right) \\
&\leq Pr\left(\sum_{c \in B_{xy}} \left[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c)) - \mu\right] + \left[V_g(w_{\boldsymbol{q}}(y,x)) - V_g(w_{\boldsymbol{q}}(x,y)) - \mu\right]\right. \\
&\quad \left. + \sum_{c \notin B_{xy}, c \neq x,y} \left[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c))\right] \geq -d\mu\right)
\end{aligned}$$

Now using Hoeffding's inequality for bounded random variables, and noticing that the random variables are indeed independent, bounded by $K = \max_{w^*(x,y)} V_g(w^*(x,y))$ which does not depend on m, and with mean zero (since we removed the mean), and with $m-1$ random variables, we get

$$Pr(y \succ_{\hat{L}_g} x | x \succ_{L^*} y) \leq e^{-2\frac{d^2\mu^2}{K^2(m-1)}} = e^{-d^2\frac{2\mu^2}{K^2(m-1)}}$$

In order to get the polynomial decrease in m (specifically m^{-1}), we need to make sure $d \geq \sqrt{\frac{K^2(m-1)}{2\mu^2}\log(m)} = \Omega(\sqrt{m \log(m)})$, then we get

$$Pr(y \succ_{\hat{L}_g} x | x \succ_{L^*} y) \leq e^{-d^2\frac{2\mu^2}{K^2(m-1)}} \leq e^{-\log(m)} = m^{-1}$$

Using this we can finally get a bound on the expected distance. Denoting $I_d = \left\{x, y \in \binom{I}{2} | L^*(y) - L^*(x) = d\right\}$, notice $|I_d| = m - d \leq m$. Denote also $d^* := \sqrt{\frac{K^2(m-1)}{2\mu^2}\log(m)}$. This gives us:

$$
\begin{aligned}
D(g,\boldsymbol{p},\boldsymbol{q}) &= \sum_{x,y\in\binom{m}{2}} Pr(y \succ_{\hat{L}_g} x | x \succ_{L^*} y) = \\
&\sum_{d=1}^{d^*}\sum_{x,y\in I_d} Pr(y \succ_{\hat{L}_g} x | x \succ_{L^*} y) + \sum_{d=d^*}^{m}\sum_{x,y\in I_d} Pr(y \succ_{\hat{L}_g} x | x \succ_{L^*} y) \\
&\le \sum_{d=1}^{d^*} m + \sum_{d=d^*}^{m} m\cdot m^{-1} = O(m\sqrt{m\log(m)})
\end{aligned}
$$

As needed. □

F PROOFS OF SUBSECTION 4.2

Theorem 15.
For any graph based item sum-scoring rule g, $D(g,\boldsymbol{p},\boldsymbol{q}) = \Omega(m\sqrt{m})$.

Proof. In this proof we will use $C_0, C_1, \dots$ to denote absolute constants. We will first analyze the switch probabilities, which will lead to the bound on the distance. Specifically, we will prove that the switch probability of two items x, y which are of distance $d \le \sqrt{m}$ apart is not approaching zero. Note that

$$
\begin{aligned}
Pr(y \succ_{\hat{L}_g} x | x \succ_{L^*} y) &= Pr\left(\sum_{c\neq y} V_g(w_{\boldsymbol{q}}(y,c)) \ge \sum_{d\neq x} V_g(w_{\boldsymbol{q}}(x,d)) | x \succ_{L^*} y\right) \\
&= Pr\left(\sum_{c\neq x,y} V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c)) + V_g(w_{\boldsymbol{q}}(y,x)) - V_g(w_{\boldsymbol{q}}(x,y)) \ge 0 | x \succ_{L^*} y\right)
\end{aligned}
$$

Now, notice that we can split the sum over the items c into three types: $c \succ x,y; x \succ c \succ y; x,y \succ c$. If we treat the value of the sum as a random variable, it becomes the (not identically but indeed independent) sum of random variables. Under the conditions that $c \succ x,y; x,y \succ c$,the expected value of $E\left[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c))\right] = 0$, since $w_{\boldsymbol{q}}(y,c)$ and $w_{\boldsymbol{q}}(x,c)$ are identically distributed. We denote $\mu_c = E\left[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c))\right], \mu_0 = E\left[V_g(w_{\boldsymbol{q}}(y,x)) - V_g(w_{\boldsymbol{q}}(x,y))\right]$ the expected difference. The increments of this sum are bounded with respect to m (the weights $w_{\boldsymbol{q}}(a,b)$ have a set of $2^{|N|}$ possible values, so does the function, and these do not depend on m), denote the upper bound as K. If we focus our comparisons only on items x, y that are close together, in the sense that $d \le \sqrt{m} = O(\sqrt{m})$, we get the following inequality for $m \ge 4$ (omitting the conditional, which remain to the end of the proof):

$$
\begin{aligned}
Pr(y \succ_{\hat{L}_g} x) &= Pr\left(\sum_{c\neq x,y}\left[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c)) - \mu_c\right] + V_g(w_{\boldsymbol{q}}(y,x)) - V_g(w_{\boldsymbol{q}}(x,y)) - \mu_0 \ge -\mu_0 - \sum_{c\neq x,y}\mu_c\right) \\
&\ge Pr\left(\sum_{c\neq x,y}\left[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c)) - \mu_c\right] + V_g(w_{\boldsymbol{q}}(y,x)) - V_g(w_{\boldsymbol{q}}(x,y)) - \mu_0 \ge dK\right) \\
&\ge Pr\left(\sum_{c\neq x,y}\left[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c)) - \mu_c\right] + V_g(w_{\boldsymbol{q}}(y,x)) - V_g(w_{\boldsymbol{q}}(x,y)) - \mu_0 \ge K\sqrt{m}\right)
\end{aligned}
$$

Now, we can apply the Berry-Esseen result from the local central limit theorem theory to prove that this probability does not approach zero. For this we use the variance, denoted as $\sigma_{x,c}^2 = Var[V_g(w_{\boldsymbol{q}}(x,c))]$, and for the summand we get the variance of $\sigma_c^2 = \sigma_{xc}^2 + \sigma_{yc}^2 > 0$.[9] Denote $v_m = \sqrt{\sum_{c\neq x,y}\sigma_c^2}$ the total standard deviation, and we get that (denoting $\Phi(x) = Pr_{z\sim\mathcal{N}(0,1)}(z \ge x)$ as the opposite cumulative function of the standard normal distribution):

---

[9] If the variance is 0, we get a rule that has a deterministic score for all items, which results in a full draw, and a random ranking, which was shown to have expected distance of $\frac{1}{2}m^2$, or at least one of the experts is perfect, which contradicts the conditions.

$$\begin{aligned}
Pr(y \succ_{\hat{L}_g} x) \geq & Pr\left(\frac{\sum_{c\neq x,y}[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c)) - \mu_c + V_g(w_{\boldsymbol{q}}(y,x)) - V_g(w_{\boldsymbol{q}}(x,y)) - \mu_0]}{v_m} \geq \frac{K\sqrt{m}}{v_m}\right) \\
&\geq \Phi\left(\frac{K\sqrt{m}}{v_m}\right) - \frac{C_0(\sum_{c\neq x,y} E[|V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c)) - \mu_c|^3] + E[|V_g(w_{\boldsymbol{q}}(y,x)) - V_g(w_{\boldsymbol{q}}(x,y)) - \mu_0|^3])}{v_m^3} \\
&\geq \Phi\left(\frac{K\sqrt{m}}{v_m}\right) - \frac{C_0K(v_m^2 + \sigma_{xy}^2 + \sigma_{yx}^2)}{v_m^3}
\end{aligned}$$

Finally, we can see that the sum v_m can be split into three parts, and every one of these parts has a uniform value. Denoting

$$\begin{cases} s_1 = Var[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c))] > 0 & c \succ x, y \\ s_2 = Var[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c))] > 0 & x \succ c \succ y \\ s_3 = Var[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c))] > 0 & x, y \succ c \end{cases}$$

We get the following:

$$v_m^2 = \sum_{c\neq x,y} Var[V_g(w_{\boldsymbol{q}}(y,c)) - V_g(w_{\boldsymbol{q}}(x,c))] = \sum_{c\succ x,y} s_1 + \sum_{x\succ c\succ y} s_2 + \sum_{x,y\succ c} s_3 \Longrightarrow v_m = \Theta(\sqrt{m})$$

Which also means that for big enough m we have $\sigma_{xy}^2 + \sigma_{yx}^2 \leq v_m^2$ (since these values are constant with respect to m), so we get

$$Pr(y \succ_{\hat{L}_g} x) \geq \Phi\left(\frac{K\sqrt{m}}{v_m}\right) - \frac{C_0K(v_m^2 + \sigma_{xy}^2 + \sigma_{yx}^2)}{v_m^3} \geq \Phi\left(\frac{K\sqrt{m}}{v_m}\right) - \frac{2C_0K}{v_m} \geq \Phi\left(C_1\right) - \frac{C_2K}{\sqrt{m}}$$

Taking $\sqrt{m} > \frac{2C_2K}{\Phi(C_1)}$ we get $Pr(y \succ_{\hat{L}_g} x) \geq \frac{1}{2}\Phi(C_1) > 0$, as needed.

Now, recalling the expected distance is the sum of the switch probabilities, we get that:

$$\begin{aligned}
D(g, \boldsymbol{p}, \boldsymbol{q}) = E\left[d_{KT}(\hat{L}_g, L^*)\right] &= \sum_{x,y\in\binom{I}{2}} Pr(y \succ_{\hat{L}_g} x | x \succ_{L^*} y) \\
&= \sum_{d=1}^{\sqrt{m}} \sum_{x,y\in I_d} Pr(y \succ_{\hat{L}_g} x | x \succ_{L^*} y) + \sum_{d=\sqrt{m}}^{m} \sum_{x,y\in I_d} Pr(y \succ_{\hat{L}_g} x | x \succ_{L^*} y) \\
&\geq \sum_{d=1}^{\sqrt{m}} (m - \sqrt{m})\frac{1}{2}\Phi(C_1) = \frac{1}{2}m\sqrt{m}\Phi(C_1) - \frac{1}{2}m\Phi(C_1) = \Omega(m\sqrt{m}).
\end{aligned}$$

Which completes the proof.

□

Proposition 16. *For any rule g, and a single expert with probability of success* $p \notin \{0, 1\}$, $D(g, \boldsymbol{p}, \boldsymbol{q}) = \Omega(m)$

Proof. First we notice that the optimal rule is neutral, that is for any permutation σ over the items, applying the permutation to the dataset will change the winning order exactly by the permutation itself. Formally, define for any $\sigma : I \longrightarrow I$ permutation, if $\sigma(a) = b, \sigma(c) = d$ ($a \neq c$ different, $b \neq d$), then $\sigma(S_{bd}) = S_{ac}$, and this extends to the entire dataset.

Now a rule g is called neutral iff $g(\sigma(S)) = \sigma(g(S))$. The optimal rule calculated by Proposition 11 is clearly neutral. Define G the set of all social welfare functions, and G_n all neutral social welfare functions. We use the neutrality to focus on a specific true ranking $x \succ_{L^*} y \iff x > y$ (the items are now $I = \{1, ..., m\}$):

$$\begin{aligned}
&Sup_{L^*} Inf_{g\in G} \sum_{x,y\in\binom{I}{2}} Pr(y \succ_g x | x \succ_{L^*} y) = Sup_{L^*} Inf_{g\in G_n} \sum_{x,y\in\binom{I}{2}} Pr(y \succ_g x | x \succ_{L^*} y) \\
&= Inf_{g\in G_n} \sum_{x,y\in\binom{I}{2}} Pr(y \succ_g x | x > y) \geq \sum_{i=1}^{m-1} Inf_{g\in G_n} Pr(i \succ_g i + 1)
\end{aligned}$$

We are now considering only pairs of items that have no other items between them. For any such pair $i, i+1$ consider the permutation $\sigma_i = (i, i+1)$ that only changes the order between these two items. Also denote the event $A = \{\forall j \in N, i \succ_j i+1\}$ and $\sigma(A) = \{\forall j \in N, i+1 \succ_j i\}$. We now get

$$\begin{aligned}
Pr(i \succ_g i+1) &\geq Pr(A)Pr(i \succ_g i+1|A) + Pr(\sigma(A))Pr(i \succ_g i+1|\sigma(A)) \\
&= \prod_{i \in N} p_i Pr(i \succ_g i+1|A) + \prod_{i \in N}(1-p_i)Pr(i \succ_g i+1|\sigma(A)) \\
&= \prod_{i \in N} p_i Pr(i \succ_g i+1|A) + \prod_{i \in N}(1-p_i)Pr(i+1 \succ_g i|A)
\end{aligned}$$

Where the last equality comes from the neutrality of g, as well as the fact that for any item $j \in I$, the distribution of $i \succ_S j$ is the same as the distribution that $i+1 \succ_S j$. We now denote the probability of the rule to conform $p_{conf} = Pr(i \succ_g i+1|A)$ and we get

$$= \prod_{i \in N} p_i p_{conf} + \prod_{i \in N}(1-p_i)(1-p_{conf}) \geq min\{\prod_{i \in N} p_i, \prod_{i \in N}(1-p_i)\} = \Theta(1)$$

Where the last step uses the assumption that $p \notin \{0, 1\}$. We now get that

$$\begin{gathered}
Sup_{L^*} Inf_{g \in G} \sum_{x,y \in \binom{I}{2}} Pr(y \succ_g x | x \succ_{L^*} y) \geq \\
\sum_{i=1}^{m-1} Inf_{g \in G_n} Pr(i \succ_g i+1) \geq \sum_{i=1}^{m-1} \Theta(1) = \Theta(m)
\end{gathered}$$

As needed.

□

G THE PROOF BY [5] IS WRONG

The paper [5] states that the expected distance of the Kemeny winning ranking (in the paper they call it the optimal ranking, and they have n the number of items since there is a single voter, we use here our terminology) from the ground truth is bounded by $O(m)$ for a single voter. To prove that they prove that every order must have higher score than the true order, and that the probability of every ranking having such greater score decreases exponentially with its distance from the ground truth.

We used the same method here. The problem comes when they try to apply the union bound. Before applying the union bound, they prove lemma 3.2 which upper bounds the number of permutations of distance $\leq cm$ to the ground truth for some constant c. They then use the union bound with the same upper bound on the number of permutations of distance *at least cm* from the ground truth.

It should be noted that using the correct union bound will not give the same bound. The number of permutations of distance $\Omega(m)$ from the ground truth is of order $\Theta(m!)$, which is faster than the exponential decay for any constant. This can be proven exactly using the claim in [5] about the number of close permutations: if the number of close permutations is not growing in order of $m!$, then the rest of the permutations must be of order $m!$.

H MAXIMIN VIOLATES REINFORCEMENT FOR $m = 3$

Definition 1. *An unweighted SCF f has the* reinforcement *property if for any two profiles S^1, S^2, if $f(S^1) = f(S^2) = a$, then $f(S') = a$ for $S' := S^1 + S^2$.*

Conitzer and Sandholm [7] provide an example where the Maximin rule violates reinforcement for $m = 4$ items.

Proposition 21. *The Maximin rule violates reinforcement even for $m = 3$.*

Proof. Consider the following two circular profiles with $n = 15$ experts each, in pairwise representation. Edges are from the row item to the column item.

In every row, we highlighted the minimal outgoing edge. The maximin winner is underlined.

S^1	a	b	c	S^2	a	b	c
$\underline{a}$	–	7	9	$\underline{a}$	–	11	**5**
b	8	–	**2**	b	**4**	–	12
c	**6**	13	–	c	10	**3**	–

As we can see, a is the Maximin winner in both profiles. However, joining the profiles, we get the following 30-expert profile:

S'	a	b	c
a	–	18	**14**
b	**12**	–	14
$\underline{c}$	**16**	**16**	–

In profile S', item c beats both a and b, so it is a Condorcet winner and in particular the Maximin winner.

This example also demonstrates where the mismatch in $Pr(S^1|a)Pr(S^2|a)$ vs. $Pr(S'|a)$, with unweighted but named experts:

$$Pr(S'_{bc}|bc) + Pr(S'_{bc}|cb) = p^{14}(1-p)^{16} + p^{16}(1-p)^{14},$$

whereas

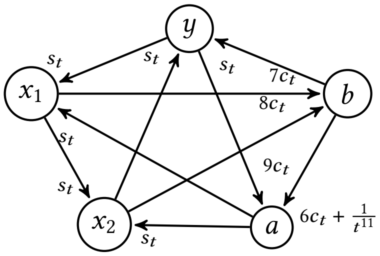
$$\begin{aligned} & Pr(S^1_{bc}|cb)Pr(S^2_{bc}|bc) + Pr(S^1_{bc}|bc)Pr(S^2_{bc}|cb) \\ & = p^{13}(1-p)^2 p^{12}(1-p)^3 + p^2(1-p)^{13}p^3(1-p)^{12} \\ & = p^{25}(1-p)^5 + p^5(1-p)^{25}. \end{aligned}$$

□

I AN EXAMPLE WHERE TIDEMAN WINNER IS NOT MOST LIKELY WINNER IN THE LIMIT

Consider the following dataset (it is equivalent to 10). Since this is not an ordering, we choose for every pair which is the winner by indicating ✓or ✗ in the corresponding column. Notice that experts only disagree on the first 4 columns.

	p_i	$a \succ b$	$a \succ y$	$b \succ x_1$	$b \succ x_2$	$a \succ x_1$
expert 1	$\frac{t+2^{\frac{1}{t^{11}}}}{1+t+2^{\frac{1}{t^{11}}}}$	✗	✓	✓	✓	✓
experts 2-5	$\frac{t}{1+t}$	✓	✗	✗	✗	✓
experts 6-7	$\frac{t}{1+t}$	✗	✗	✓	✗	✓
expert 8	$\frac{t}{1+t}$	✗	✗	✗	✓	✓
experts 9-10	$\frac{t}{1+t}$	✗	✗	✗	✗	✓



	p_i	$a \succ x_2$	$b \succ y$	$y \succ x_1$	$y \succ x_2$	$x_1 \succ x_2$
expert 1	$\frac{t+2^{\frac{1}{t^{11}}}}{1+t+2^{\frac{1}{t^{11}}}}$	✓	✓	✓	✗	✓
experts 2-5	$\frac{t}{1+t}$	✓	✓	✓	✗	✓
experts 6-7	$\frac{t}{1+t}$	✓	✓	✓	✗	✓
expert 8	$\frac{t}{1+t}$	✓	✓	✓	✗	✓
expert 9-10	$\frac{t}{1+t}$	✓	✓	✓	✗	✓

Notice the Tideman score for y, x_1, x_2 is $20\log(t) + \frac{2}{t^{11}}$ and for a, b is $10\log(t), 10\log(t) - \frac{2}{t^{11}}$ respectively, which means either a or b wins as $t \longrightarrow \infty$ since they are the clear Tideman winners over the other items (the gap will surpass the needed constant for some t). Between a and b our rule does not tell us who is the winner, while the Tideman winner is b. Total experts weights is $W^* = 10\log(t) + \frac{1}{t^{11}}$ (one expert is slightly better than the rest, and votes with the majority only when voting $b \succ a$).

We prove here that as $t \longrightarrow \infty$, the difference $\kappa(b) - \kappa(a) \longrightarrow \infty$, proving that a is the most likely winner. For this we fully calculate $\kappa(a), \kappa(b)$, up to the necessary terms (they both have a single summand of order 10 and 8 multiplied by terms that approach 1, with b having a third higher-order term bigger than a):

$$\kappa(a) \leq t^{10} + t^8 2^{-\frac{1}{t^{11}}} + 14t^2 2^{\frac{1}{t^{11}}}$$

$$\kappa(b) \geq t^{10} 2^{-\frac{2}{t^{11}}} + t^8 2^{-\frac{3}{t^{11}}} + t^6 2^{-\frac{1}{t^{11}}}$$

We thus obtain that:

$$\kappa(b) - \kappa(a) \geq t^{10}(2^{-\frac{3}{t^{11}}} - 1) + t^8(2^{-\frac{3}{t^{11}}} - 2^{-\frac{1}{t^{11}}}) + t^6 - 14t^2$$

Since the terms multiplying the highest order terms approach zero faster than the highest order terms, the most significant remaining term is t^6, and we get convergence to infinity as required.

Full calculation:

$$\begin{aligned}
\kappa(a) &= \sum_{J \subseteq I/\{a\}} 2^{\sum_{c\in J} w^*(c,a) - w^*(a,c)} \\
&= 2^0 + 2^{2\log(t)+\frac{1}{t^{11}}} + 2^{8\log(t)-\frac{1}{t^{11}}} + 2^{-10\log(t)-\frac{1}{t^{11}}} + 2^{-10\log(t)-\frac{1}{t^{11}}} + 2^{10\log(t)} + 2^{-8\log(t)} + 2^{-8\log(t)} \\
&+ 2^{-2\log(t)-\frac{2}{t^{11}}} + 2^{-2\log(t)-\frac{2}{t^{11}}} + 2^{-20\log(t)-\frac{2}{t^{11}}} + 2^{-12\log(t)-\frac{3}{t^{11}}} \\
&+ 2^{-18\log(t)-\frac{1}{t^{11}}} + 2^{-\frac{1}{t^{11}}} + 2^{-\frac{1}{t^{11}}} + 2^{-10\log(t)-\frac{2}{t^{11}}} \\
&\leq t^{10} + t^8 2^{-\frac{1}{t^{11}}} + 14t^2 2^{\frac{1}{t^{11}}} && \text{(combine all low terms and keep only the 3 highest terms)}
\end{aligned}$$

$$\begin{aligned}
\kappa(b) &= \sum_{J \subseteq I/\{b\}} 2^{\sum_{c\in J} w^*(c,b) - w^*(b,c)} \\
&= 2^0 + 2^{-2\log(t)-\frac{1}{t^{11}}} + 2^{-10\log(t)-\frac{1}{t^{11}}} + 2^{4\log(t)-\frac{1}{t^{11}}} + 2^{6\log(t)-\frac{1}{t^{11}}} \\
&+ 2^{-12\log(t)-\frac{2}{t^{11}}} + 2^{2\log(t)-\frac{2}{t^{11}}} + 2^{4\log(t)-\frac{2}{t^{11}}} \\
&+ 2^{-6\log(t)-\frac{2}{t^{11}}} + 2^{-4\log(t)-\frac{2}{t^{11}}} + 2^{10\log(t)-\frac{2}{t^{11}}} + 2^{-\frac{3}{t^{11}}} \\
&+ 2^{8\log(t)-\frac{3}{t^{11}}} + 2^{-6\log(t)-\frac{3}{t^{11}}} + 2^{-8\log(t)-\frac{3}{t^{11}}} + 2^{-2\log(t)-\frac{4}{t^{11}}} \\
&\geq t^{10} 2^{-\frac{2}{t^{11}}} + t^8 2^{-\frac{3}{t^{11}}} + t^6 2^{-\frac{1}{t^{11}}} && \text{(keep only 3 highest order terms)}
\end{aligned}$$

As required.