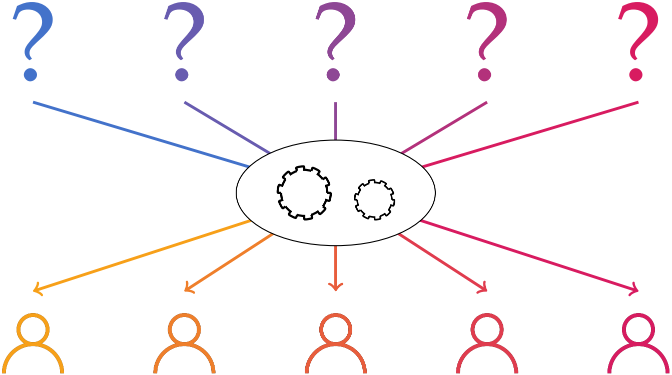
Truth-Tracking in Voting

Computational Social Choice 2021

Simon Rey
Institute for Logic, Language and Computation
University of Amsterdam

Two Views on Voting

Axiomatic approach: Studying voting rules through the normative properties they satisfy.



Epistemic approach: Studying voting rules through their ability to recover the ground truth.

Plan of Today

We will explore a new perspective on voting: the *epistemic approach*.

The first result we will discuss is the well-known *Condorcet jury theorem*.

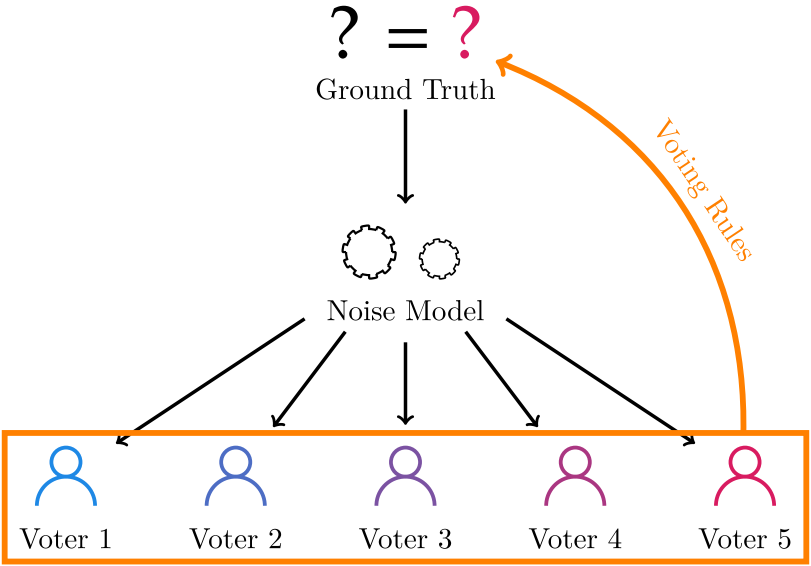
From there, we will move towards the systematic analysis of voting rules as *maximum likelihood estimators*.

We will then investigate the question of the number of samples needed for rules to perform well.

Our last technical analysis will focus on rules being *robust* against classes of noise models.

Elkind and Slinko "Rationalizations of Voting Rules" *Handbook of Computational Social Choice* (2016)

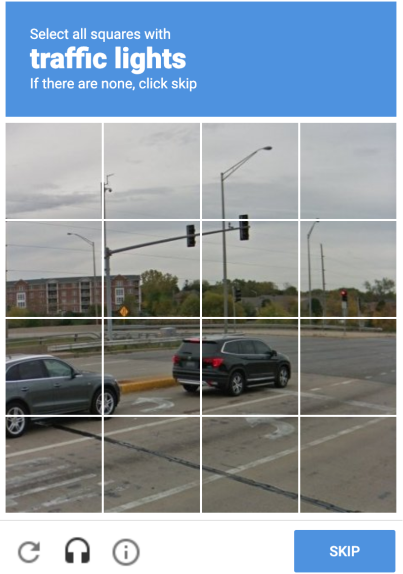
Noise Models



Ground Truth

The ground truth hypothesis states that there exists an *objectively best* outcome for any given election.

Making this assumption is not always justified... but for some cases it is highly relevant. In particular when it comes to *crowd-sourcing*.



The Model

- $N = \{1, \ldots, n\}$ is the set of voters;
- $A = \{a, b, c, \ldots\}$ is the set of alternatives;
- $\mathcal{O} \subseteq 2^A$ is the set of all admissible outcomes (can consists of only singletons, all sets of projects fitting within the budget limit, ...);
- $o^\star \in \mathcal{O}$ is the ground truth of the election;
- $\mathcal{B}$ is the set of all admissible ballots (rankings, approval, ...);
- Each voter $i \in N$ submits a ballot $b_i \in \mathcal{B}$, giving raise to a profile $\boldsymbol{b} = (b_1, \ldots, b_n)$;
- A noise model $\mathcal{M} : \mathcal{O} \to \Pi(\mathcal{B})$ is a function mapping the ground truth $o^\star$ to a probability distribution over the ballots $\mathcal{M}(o^\star)$ ($\Pi(\mathcal{B})$ is the set of all probability distribution over $\mathcal{B}$); $\mathcal{M}(o^\star)(b)$ is the probability for a specific ballot $b \in \mathcal{B}$ to be generated;
- For $\boldsymbol{b}$, let $\mathcal{M}(o^\star)(\boldsymbol{b}) = \prod_{b \in \boldsymbol{b}} \mathcal{M}(o^\star)(b)$, i.e. ballots are *i.i.d.*;
- A voting rule $F : \mathcal{B}^n \to 2^{\mathcal{O}} \setminus \{\emptyset\}$ takes as input a profile $\boldsymbol{b}$ and returns a set of admissible outcomes $F(\boldsymbol{b})$.

1. Simple Case: Two Candidates

Condorcet's Noise Model

Consider an election with two alternatives, $A = \{a, b\}$. The goal is to select one of the two: $\mathcal{O} = \{\{a\}, \{b\}\}$. Voters submit plurality ballots (i.e., submit the name of one alternative): $\mathcal{B} = \{a, b\}$.

Condorcet's noise model parametrized by $p \in [0, 1]$ is defined such that for every ballot $b \in \mathcal{B}$ we have:

$$\mathcal{M}_p^{Cond}(o^\star)(b) = \begin{cases} p & \text{if } b = o^\star, \\ 1 - p & \text{otherwise.} \end{cases}$$

Remember that the ballots of the voters are sampled *identically and independently* (i.i.d. assumption).

Condorcet "Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix" *Imprimerie Royale* (1785)

Condorcet Jury Theorem

Theorem:

Under Condorcet's noise model $\mathcal{M}_p^{Cond}$, if $1/2 < p \leq 1$, then the *majority rule* selects the ground truth with probability *1* as $n \rightarrow +\infty$.

➥ Formalizes the *wisdom of the crowd*.

Condorcet "Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix" *Imprimerie Royale* (1785)
Young "Condorcet's theory of voting" *American Political science review* (1988)

Proof of the Condorcet Jury Theorem

Proposition: Strong law of large numbers

Let $X_1, X_2, \ldots$ be pairwise *independent identically distributed* random variables with $\mu = \mathbb{E}(X_i) < \infty$. Let $S_n = X_1 + \ldots + X_n$. Then, ${}^{S_n}/_n$ converges *almost surely* to μ as $n \rightarrow +\infty$.

Let a and b be the two alternatives, with a being the ground truth.

Consider a voter $i \in N$ to be a random variable $X_i \in \{0, 1\}$ with:

- $\mathbb{P}(X_i = 1) = p$ (candidate a);
- $\mathbb{P}(X_i = 0) = 1 - p$ (candidate b).

Let $S_n = \sum_{i \in N} X_i$. The majority rules is correct iff ${}^{S_n}/_n > {}^1/_2$.

By the above, ${}^{S_n}/_n$ *converges to p with probability 1* as $n \rightarrow +\infty$.

Since $p > {}^1/_2$, the result follows. ∎

Extensions of the Condorcet Jury Theorem

Assumptions of the Condorcet jury theorem can be relaxed in different ways:

- What if we do *not know the skills* of the voters?
- What if voters are *asymmetric*?
- What if voters behave *strategically*?

See the following references for a discussion on that (and also on some further topics).

Nitzan "Collective preference and choice" *Cambridge University Press* (2009)

Dietrich and Spiekermann "Jury Theorems" *The Routledge Handbook of Social Epistemology* (2019)

Towards the Maximum Likelihood Approach

The Condorcet jury theorem tells us that using the majority rule is a good choice under Condorcet's noise model.

↪ Let us pursue this idea of defining the best rule to use when we know the way agents form their preferences (the noise model).

2. Maximum Likelihood Estimators

Likelihood

Likelihood: Given a parametrized probability distribution and a set of observations, how likely was it that the observations have been generated by the probability distribution with a given parameter?

Observations
(Profile)

$$L(\boldsymbol{b}, o^\star) = \prod_{b \in \boldsymbol{b}} \mathcal{M}(o^\star)(b)$$

Parameter of the model
(Ground truth)

Probabilistic model
(Noise model)

Maximum Likelihood Estimator

Maximum likelihood estimators: Select over all ground truths, the one such that the likelihood of the profile is the highest.

$$F(\boldsymbol{b}) = \arg\max_{o^\star \in \mathcal{O}} L(\boldsymbol{b}, o^\star) = \arg\max_{o^\star \in \mathcal{O}} \prod_{b \in \boldsymbol{b}} \mathcal{M}(o^\star)(b)$$

↳ Using this formula as it is may not always be practical. Can we interpret *known rules* as maximum likelihood estimators (MLE)?

Positive Result: Positional Scoring Rules

<u>THEOREM</u>:

Any *positional scoring rule* can be interpreted as an *MLE*.

Proof: Assume voters submit rankings over the alternatives. Let $o^\star$ be the ground truth winner. Consider the PSR with score function s.

Suppose voter $i \in N$ ranks $o^\star$ at position $r_i(o^\star)$ with probability proportional to $2^{s(r_i(o^\star))}$. The likelihood of $\boldsymbol{b}$ and o is such that:

$$L(\boldsymbol{b}, o) \propto \prod_{b_i \in \boldsymbol{b}} 2^{s(r_i(o))} = 2^{\sum_{b_i \in \boldsymbol{b}} s(r_i(o))}$$

The MLE selects then the alternative with the highest score. ■

Conitzer and Sandholm "Common voting rules as maximum likelihood estimators" *Proc. of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)* (2005)

Negative Result: Weak Reinforcement

THEOREM:

If F fails *weak reinforcement*, it never is an *MLE*.

Weak reinforcement: $\forall \boldsymbol{b}, \boldsymbol{b}', F(\boldsymbol{b}) = F(\boldsymbol{b}') \implies F(\boldsymbol{b} \oplus \boldsymbol{b}') = F(\boldsymbol{b})$.

Proof: Suppose F fails weak reinforcement. Consider $\boldsymbol{b}$ and $\boldsymbol{b}'$ such that $F(\boldsymbol{b}) = F(\boldsymbol{b}') = o$ but $F(\boldsymbol{b} \oplus \boldsymbol{b}') \neq o$.

Let $\mathcal{M}$ be such that $o = \arg\max_{o^\star \in \mathcal{O}} L(\boldsymbol{b}, o^\star) = \arg\max_{o^\star \in \mathcal{O}} L(\boldsymbol{b}', o^\star)$.

Then, since ballots are independently distributed (for the last $=$):

$$F(\boldsymbol{b} \oplus \boldsymbol{b}') \neq o = \arg\max_{o^\star \in \mathcal{O}} L(\boldsymbol{b}, o^\star) L(\boldsymbol{b}', o^\star) = \arg\max_{o^\star \in \mathcal{O}} L(\boldsymbol{b} \oplus \boldsymbol{b}', o^\star).$$

Conitzer and Sandholm "Common voting rules as maximum likelihood estimators" *Proc. of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)* (2005)

The General Picture

	MLE for Winner	¬MLE for Winner
MLE for Ranking	Scoring rules	Weird rules
¬MLE for Ranking	STV	Copeland, ranked pairs

↳ The MLE requirement is quite strong. Let's explore different properties to distinguish between the rules.

Conitzer and Sandholm "Common voting rules as maximum likelihood estimators" *Proc. of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)* (2005)

3. Sample Complexity

A New Question

How many ballots does a voting rule need to recover the ground truth?

↳ *Sample complexity* gives an answer to that question.

Caragiannis, Procaccia, and Shah "When do Noisy Votes Reveal the Truth?" *ACM Transactions on Economics and Computation* (2016)

Accuracy of a Rule

Accuracy of a rule for a ground truth: The probability that a rule returns the ground truth given a fixed number of samples from a noise model.

Noise model

$$Acc(F, k, o^\star) = \sum_{\boldsymbol{b} \in \mathcal{B}^k} \mathcal{M}(o^\star)(\boldsymbol{b}) \times \mathbb{1}_{F(\boldsymbol{b})=o^\star}$$

All profiles of size k

Indicator on F returning $o^\star$

Accuracy of a Rule

Accuracy of a rule for a ground truth:

$$Acc(F, k, o^\star) = \sum_{\boldsymbol{b} \in \mathcal{B}^k} \mathcal{M}(o^\star)(\boldsymbol{b}) \times \mathbb{1}_{F(\boldsymbol{b}) = o^\star}$$

Accuracy of a rule: The worst-case accuracy for any ground truth given a fixed number of samples.

$$Acc(F, k) = \min_{o^\star \in \mathcal{O}} \sum_{\boldsymbol{b} \in \mathcal{B}^k} \mathcal{M}(o^\star)(\boldsymbol{b}) \times \mathbb{1}_{F(\boldsymbol{b}) = o^\star}$$

Sample Complexity

Accuracy of a rule:

$$Acc(F, k) = \min_{o^\star \in \mathcal{O}} \sum_{\boldsymbol{b} \in \mathcal{B}^k} \mathcal{M}(o^\star)(\boldsymbol{b}) \times \mathbb{1}_{F(\boldsymbol{b}) = o^\star}$$

Sample complexity: The smallest number of samples needed to achieve an accuracy close to 1, i.e., the smallest number of samples needed for a rule to return the ground truth with probability $1 - \epsilon$.

$$\mathcal{SC}(F, \epsilon) = \min \{k \in \mathbb{N} \mid Acc(F, k) \geq 1 - \epsilon\}$$

Mallows' Model

Let ballots be *rankings* over the alternatives. Mallows' model takes as parameters a distance d (Kendall-tau here) and a level of noise $\gamma \in [0, 1]$ such that:

$$\mathcal{M}^{Mall}(o^\star)(b) = \frac{\gamma^{d(b,o^\star)}}{Z_\gamma}$$

$$d(b, o^\star) = \sum_{x,y \in A} \mathbb{1}_{o^\star(x) > o^\star(y)} \times \mathbb{1}_{b(y) > b(x)}$$

Z_γ is a normalization factor that is *independent of the ground truth* with the Kendall-tau distance (the distance defined above).

For $\gamma = 0$, only the ground truth has a *non-zero probability* to occur.
For $\gamma = 1$, all rankings are *equally likely* to occur.

Mallows "Non-Null Ranking Models. I" *Biometrika* (1957)

The Kemeny Rule

Definition:

The *Kemeny rule* KEM takes as input a profile $\boldsymbol{b} = (b_1, \ldots, b_n)$ and is such that:

$$KEM(\boldsymbol{b}) = \arg\min_{o \in \mathcal{O}} \sum_{i \in N} d(b_i, o)$$

$$d(b, o) = \sum_{x, y \in A} \mathbb{1}_{o(x) > o(y)} \times \mathbb{1}_{b(y) > b(x)}$$

The Kemeny Rule is Optimal for Mallows' Model

THEOREM:

Given $\epsilon > 0$, the *Kemeny rule* with uniform tie-breaking has optimal sample complexity in Mallows' model. For every rule F, we have:

$$\mathcal{SC}(KEM, \epsilon) \leq \mathcal{SC}(F, \epsilon).$$

Maybe unsurprising as the Kemeny rule is an *MLE* for Mallows' model.

↳ This is however not always the case: being an MLE does not always imply having optimal sample complexity.

Caragiannis, Procaccia, and Shah "When do Noisy Votes Reveal the Truth?" *ACM Transactions on Economics and Computation* (2016)

Lu and Boutilier "Learning Mallows Models with Pairwise Preferences" *ICML* (2011)

The Kemeny Rule is Optimal for Mallows' Model

Define $\mathit{TotalAcc}(F, k) = \sum_{o \in \mathcal{O}} \mathit{Acc}(F, k, o)$.

Assume the following two points (proofs are rather technical):

- *Lemma 1:* $\mathit{Acc}(\mathit{KEM}, k, o) = \mathit{Acc}(\mathit{KEM}, k, o')$, for all k, o, o'.
- *Lemma 2:* $\mathit{TotalAcc}(\mathit{KEM}, k) \geq \mathit{TotalAcc}(F, k)$, for all F, k.

Fix $\mathcal{SC}(\mathit{KEM}, \epsilon) = k$. There is $\hat{o} \in \mathcal{O}$ s.t. $\mathit{Acc}(\mathit{KEM}, k-1, \hat{o}) < 1 - \epsilon$.

From *Lemma 1*: $\mathit{Acc}(\mathit{KEM}, k-1, o) < 1 - \epsilon$ *for all* $o \in \mathcal{O}$.

Hence, $\mathit{TotalAcc}(\mathit{KEM}, k-1) < m!(1 - \epsilon)$ (there are $m!$ rankings).

For any F, *Lemma 2* gives us:

$$\mathit{TotalAcc}(F, k-1) \leq \mathit{TotalAcc}(\mathit{KEM}, k-1) < m!(1 - \epsilon).$$

By the pigeonhole principle, there is $o \in \mathcal{O}$ s.t. $\mathit{Acc}(F, k-1, o) < 1 - \epsilon$.

Thus $\mathcal{SC}(F, \epsilon) \geq k = \mathcal{SC}(\mathit{KEM}, \epsilon)$. ■

Number of Samples Required

<u>Theorem</u>:

For any $\epsilon > 0$, the *Kemeny rule* returns the ground truth with probability $1 - \epsilon$ given $O(\ln(|A|/\epsilon))$ samples from Mallows' model.

The *plurality rule* sometimes requires *exponentially* many samples for Mallows' model.

Positional scoring rules with *distinct* weights require a *polynomial* number of samples from Mallows' model.

Caragiannis, Procaccia, and Shah "When do Noisy Votes Reveal the Truth?" *ACM Transactions on Economics and Computation* (2016)

Towards an Axiomatic Perspective

For now, all properties we have studied apply to a specific noise model.
Can we say something about classes of noise models?

↪ This leads to an axiomatic analysis of the truth-tracking approach.

4. Robustness to Noise

Accuracy in the Limit

Definition:

A rule F is *accurate in the limit* for a given noise model $\mathcal{M}$ if for every $\epsilon > 0$, there exists an n_ϵ such that for every profile of size at least n_ϵ, F returns the *ground truth* with probability $1 - \epsilon$:

$$\lim_{n \to +\infty} Acc(F, n) = 1.$$

↳ This is a *normative* requirement (an axiom): any relevant rule should be able to recover the ground truth given sufficiently many samples from the noise model.

Caragiannis, Procaccia, and Shah "When do Noisy Votes Reveal the Truth?" *ACM Transactions on Economics and Computation* (2016)

Monotone Robust Rules

We will focus on classes of noise models. We will classify them using distances between ballots and outcomes.

For a given distance d between ballots and outcomes, a noise model $\mathcal{M}$ is *d-monotonic* if for any two ballots b, b' and any $o^\star$, we have:

$$\mathcal{M}(o^\star)(b) > \mathcal{M}(o^\star)(b') \iff d(b, o^\star) < d(b', o^\star).$$

We are interested in voting rules that are accurate in the limit for sets of "similar" noise models, where similarity is defined by the above.

Definition:

A rule is *monotone robust* against a distance d if it is accurate in the limit for *every* d-monotonic noise model.

Uniquely Robust Rules

Theorem:

Modal ranking (outputting the ranking that has been submitted the highest number of time) is the only scoring rule that is monotone robust against *all* distances.

Modal counting (outputting the approval ballot that has been submitted the highest number of time) is the only approval-based multiwinner rule that is monotone robust against *all* distances.

Caragiannis, Procaccia, and Shah "Modal Ranking: A Uniquely Robust Voting Rule" *Proc. of the 28th AAAI Conference on Artificial Intelligence (AAAI)* (2014)
Caragiannis, Kaklamanis, Karanikolas, and Krimpas "Evaluating Approval-Based Multiwinner Voting in Terms of Robustness to Noise" *IJCAI* (2020)

5. Conclusion

Summary

We have introduced the truth-tracking approach to voting.

We started simple with the *Condorcet jury theorem* which lead us to the idea of looking at voting rules as *maximum likelihood estimators*. We showed that some rules can be interpreted as MLE and some others cannot.

We then investigated the question of *how many samples* are needed to recover the ground truth. We saw that the Kemeny rule has optimal *sample complexity* for Mallows' model.

Finally we connected the axiomatic approach to the truth-tracking one by discussing the idea of *robustness against a distance*.

What's next? Judgment aggregation, a more general framework for aggregating information that subsumes the voting framework.