# A Selective Literature Review of the Truth Tracking Approach in Computational Social Choice

Simon Rey

November 26, 2020

# 

# ???????

# 

# Voting Theory



# 









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

# ??????

# ? ? ? ? ?

# $\land$ $\land$ $\land$ $\land$ $\land$







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

Noise Models

# ? = ? Ground Truth

Noise Models



Noise Model

Noise Models





• Condorcet Jury Theorem

- Condorcet Jury Theorem
- Maximum Likelihood Approach

- Condorcet Jury Theorem
- Maximum Likelihood Approach
- Sample Complexity

- Condorcet Jury Theorem
- Maximum Likelihood Approach
- Sample Complexity
- Robusteness to noise

- Condorcet Jury Theorem
- Maximum Likelihood Approach
- Sample Complexity
- Robusteness to noise

From specific noise models to classes of noise models

# 1. Simple Case: Two Candidates
















































Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



Accuracy: 60%



#### Increasing Number of Voters with Accuracy 60%


#### THEOREM:

For an election with *two candidates* and *n* voters, if the voters correctly identify the ground truth with probability 1/2 and do so*independently*, then the*majority rule* $selects the ground truth with probability 1 as <math>n \to +\infty$ .

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

• Young "Condorcet's theory of voting" (1988)

#### THEOREM:

For an election with *two candidates* and *n* voters, if the voters correctly identify the ground truth with probability 1/2 and do so*independently*, then the*majority rule* $selects the ground truth with probability 1 as <math>n \to +\infty$ .

→ This is the first application of the maximum likelihood approach that we know of!

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

• Young "Condorcet's theory of voting" (1988)

# 2. Maximum Likelihood Approach



# Maximum Likelihood Approach

- Basic Definitions and First Results

$$L(\boldsymbol{V},\boldsymbol{\theta}) = \prod_{\boldsymbol{V} \in \boldsymbol{V}} \mathbb{P}(\boldsymbol{V} \mid \boldsymbol{\theta})$$

#### Likelihood







$$R(\mathbf{V}) = \arg\max_{\theta} L(\mathbf{V}, \theta) = \arg\max_{\theta} \prod_{V \in \mathbf{V}} \mathbb{P}(V \mid \theta)$$

#### Maximum Likelihood Estimator



#### Maximum Likelihood Estimator



#### Maximum Likelihood Estimator



|                     | MLE for Winner                        | not MLE for Winner                          |
|---------------------|---------------------------------------|---|
| MLE for Ranking     | Scoring rules: Borda, veto, plurality | Weird rules                                 |
| not MLE for Ranking | STV                                   | Bucklin, Copeland,<br>maximin, ranked pairs |

Conitzer and Sandholm "Common voting rules as maximum likelihood estimators" (2006)
Conitzer, Rognlie, and Xia "Preference Functions that Score Rankings and Maximum Likelihood Estimation" (2009)

### Maximum Likelihood Approach

└─ The Case of Approval Ballots

Ground truth:  $\sigma^{\star}$ 













#### <u>THEOREM</u>:

With the *Kendall tau* distance the set of MLE best alternatives coincides with the set of approval winners.

• Procaccia and Shah "Is Approval Voting Optimal Given Approval Votes?" (2015)

#### <u>THEOREM</u>:

With the *Kendall tau* distance the set of MLE best alternatives coincides with the set of approval winners.

With *plurality* or *veto* ballots, approval voting is an MLE for every "relevant" distance.

• Procaccia and Shah "Is Approval Voting Optimal Given Approval Votes?" (2015)

#### Looking for Specific Objectives

Select k alternatives so to maximize the probability of containing:

- the top alternative of the ground truth ranking,
- Ithe top k alternatives of the ground truth ranking,
- Ithe top k alternatives of the ground truth ranking in the right order.

• Procaccia, Reddi, and Shah "A maximum likelihood approach for selecting sets of alternatives" (2012)

#### Looking for Specific Objectives

Select k alternatives so to maximize the probability of containing:

- the top alternative of the ground truth ranking,
- Ithe top k alternatives of the ground truth ranking,
- Ithe top k alternatives of the ground truth ranking in the right order.

THEOREM:

All three objectives are NP-hard to achieve under Mallows' model.

They are *tractable* in very *noisy situation* ( $\gamma \approx 1$ ).

• Procaccia, Reddi, and Shah "A maximum likelihood approach for selecting sets of alternatives" (2012)

# 3. Sample Complexity



Sample Complexity

- Some Definitions

$$Acc(R,k) = \left(\sum_{\boldsymbol{V} \in \mathcal{L}(A)^{k}} \mathbb{P}(\boldsymbol{V} \mid \sigma^{\star}) \mathbb{P}(R(\boldsymbol{V}) = \sigma^{\star})\right)$$

Accuracy of rule Rwith k samples  $\left(\sum_{\boldsymbol{V}\in\mathcal{L}(A)^{k}}\mathbb{P}(\boldsymbol{V}\mid\sigma^{\star})\mathbb{P}(R(\boldsymbol{V})=\sigma^{\star})\right)$ Acc(R, k) =











$$Acc(R, k) = \min_{\sigma^{\star} \in \mathcal{L}(A)} \left( \sum_{\boldsymbol{V} \in \mathcal{L}(A)^{k}} \mathbb{P}(\boldsymbol{V} \mid \sigma^{\star}) \mathbb{P}(R(\boldsymbol{V}) = \sigma^{\star}) \right)$$

$$\mathcal{SC}(R,\epsilon) = \min \left\{ k \mid Acc(R,k) \ge 1 - \epsilon \right\}$$

### Sample Complexity

└─ Sample Complexity in Practice
Given  $\epsilon > 0$ , the *Kemeny rule* with uniform tie-breaking is such that for Mallows' mode and for every rule *R*, we have:

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

For any  $\epsilon > 0$ , the *Kemeny rule* returns the ground truth with probability  $1 - \epsilon$  given  $\mathcal{O}(\ln(|A|/\epsilon))$  and no rule can do better.

For any  $\epsilon > 0$ , the Kemeny rule returns the ground truth with probability  $1 - \epsilon$  given  $\mathcal{O}(\ln(|A|/\epsilon))$  and no rule can do better.

➡ Also holds for pairwise-majority consistent rules.

- 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.

## 4. Robustness to Noise



Robustness to Noise

— Definitions, Again!

### DEFINITION:

A rule *R* is *accurate in the limit* for a noise model if for every  $\epsilon > 0$ , there exists  $n_{\epsilon}$  such that for every profile of size at least  $n_{\epsilon}$ , *R* returns the *ground truth* with probability  $1 - \epsilon$ .

#### DEFINITION:

A noise model is *d*-monotonic if for any  $\sigma$ ,  $\sigma'$ , we have:

$$\mathbb{P}(\sigma \mid \sigma^{\star}) > \mathbb{P}(\sigma' \mid \sigma^{\star}) \Longleftrightarrow d(\sigma, \sigma^{\star}) < d(\sigma', \sigma^{\star}).$$

#### DEFINITION:

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

## **Robustness to Noise**

Pairwise Majority Consistent Rules

















#### DEFINITION:

A rule *R* is PM-consistent if it outputs the *Condorcet order* when the PM graph is *complete* and *acyclic*.

## Majority Concentric Distances



## Majority Concentric Distances



## Majority Concentric Distances



DEFINITION:

A distance *d* if *majority concentric* if for every  $\sigma$ , every *a*, *b* such that  $a \succ_{\sigma} b$  and every *k* we have:

$$|\eta_{\mathbf{a}\succ \mathbf{b}}^{\mathbf{k}}(\sigma)| \geq |\eta_{\mathbf{b}\succ \mathbf{a}}^{\mathbf{k}}(\sigma)|$$

All *PM consistent rules* are monotone robust against *d* if and only if *d* is *majority concentric*.

*Multiwinner approval voting* is *d*-monotone robust if and only if *d* is *majority concentric*.

• Caragiannis, Kaklamanis, Karanikolas, and Krimpas "Evaluating Approval-Based Multiwinner Voting in Terms of Robustness to Noise" (2020) **Robustness to Noise** 

Gloablly Robust Rules

### <u>Theorem</u>:

*Modal ranking* is the only generalized scoring rule that is monotone robust against *all* distances.

• Caragiannis, Procaccia, and Shah "Modal Ranking: A Uniquely Robust Voting Rule." (2014)

• Caragiannis, Kaklamanis, Karanikolas, and Krimpas "Evaluating Approval-Based Multiwinner Voting in Terms of Robustness to Noise" (2020)

## Uniquely Robust Rules

#### <u>Theorem</u>:

*Modal ranking* is the only generalized scoring rule that is monotone robust against *all* distances.

THEOREM:

*Modal counting* is the only ABCC multiwinner rule that is monotone robust against *all* distances.

• Caragiannis, Procaccia, and Shah "Modal Ranking: A Uniquely Robust Voting Rule." (2014)

• Caragiannis, Kaklamanis, Karanikolas, and Krimpas "Evaluating Approval-Based Multiwinner Voting in Terms of Robustness to Noise" (2020)

# 5. Conclusion and Future Directions



• *Maximum Likelihood Approach:* Which outcome should we select given that agents form their preferences following a specific noise model?

- *Maximum Likelihood Approach:* Which outcome should we select given that agents form their preferences following a specific noise model?
- *Sample Complexity:* How many samples do we need to achieve a suitable accuracy?

- *Maximum Likelihood Approach:* Which outcome should we select given that agents form their preferences following a specific noise model?
- *Sample Complexity:* How many samples do we need to achieve a suitable accuracy?
- *Robustness to Noise:* Does the rule return the ground truth with high probability when there are infinitely many ballots? Is it true for classes of noise model based on a distance?

- Generalizations of the Condorcet Jury Theorem
- Epistemic social choice literature in Economics, Political Science ....
- Other estimators, criteria, objectives, ...

- Bovens and Rabinowicz "Democratic answers to complex questions-an epistemic perspective" (2006)
- Pivato "Voting Rules as Statistical Estimators" (2013)
- Xia "Statistical Properties of Social Choice Mechanisms" (2014)
- Elkind and Slinko "Rationalizations of Voting Rules" (2016)
- Pivato "Realizing epistemic democracy" (2019)

- Developing epistemic approaches in more complex voting settings:
  - *Multiwinner voting*: Generalizing the work of Caragiannis et al. (2020) to non-ABCC rules (Phragmen for instance)

- Developing epistemic approaches in more complex voting settings:
  - *Multiwinner voting*: Generalizing the work of Caragiannis et al. (2020) to non-ABCC rules (Phragmen for instance)
  - Settings with Constrained Outcomes: Participatory Budgeting, Judgment Aggregation...

- Developing epistemic approaches in more complex voting settings:
  - *Multiwinner voting*: Generalizing the work of Caragiannis et al. (2020) to non-ABCC rules (Phragmen for instance)
  - Settings with Constrained Outcomes: Participatory Budgeting, Judgment Aggregation...
- Investigating different probability models with non-uniform distributions, dependencies to other features of the models...

- Developing epistemic approaches in more complex voting settings:
  - *Multiwinner voting*: Generalizing the work of Caragiannis et al. (2020) to non-ABCC rules (Phragmen for instance)
  - Settings with Constrained Outcomes: Participatory Budgeting, Judgment Aggregation...
- Investigating different probability models with non-uniform distributions, dependencies to other features of the models...
- Looking into the links between various complexity classes: elicitation complexity, sample complexity, communication complexity...