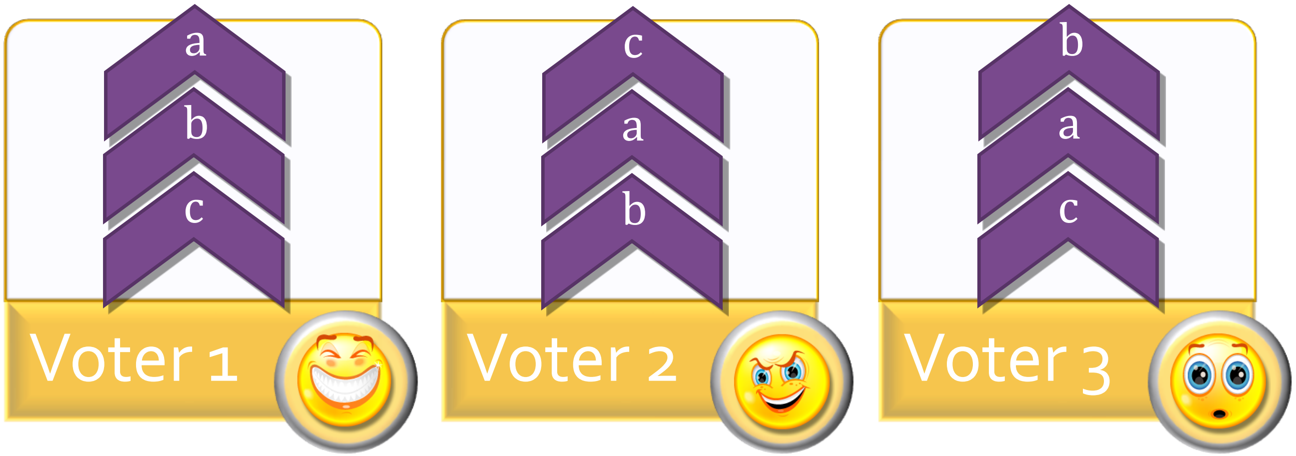
When do noisy votes reveal the truth?

Ioannis Caragiannis
University of Patras

Joint work with Ariel Procaccia and Nisarg Shah

Voting theory

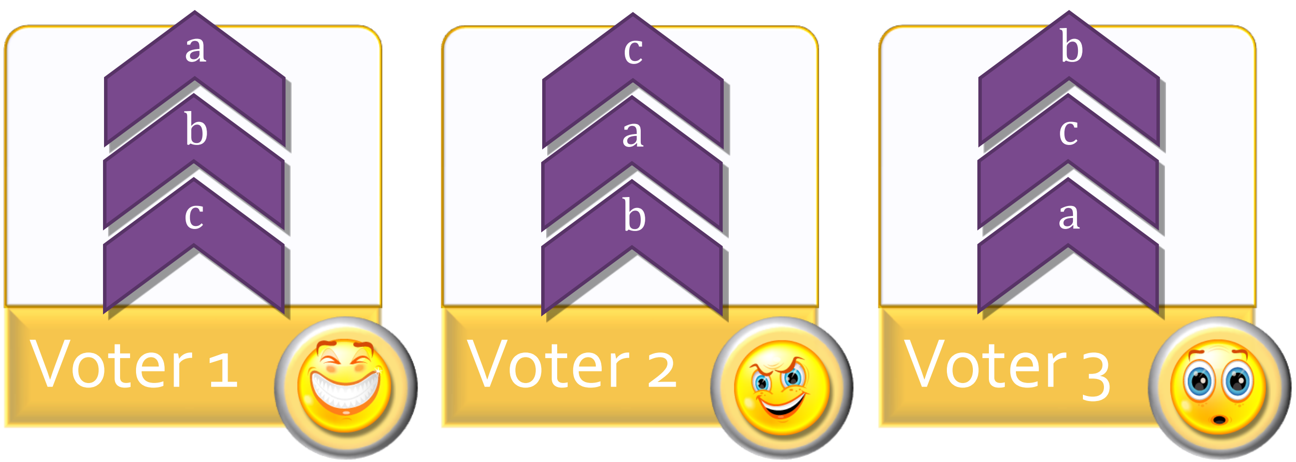
- n voters
- m alternatives
- Voters rank the alternatives
- Profile: one ranking per voter



- Voting rule: takes a profile as input and returns a winning alternative or an aggregate ranking

Voting theory

- n voters
- m alternatives
- Voters rank the alternatives
- Profile: one ranking per voter



- Voting rule: takes a profile as input and returns a winning alternative or an aggregate ranking

The setting

- There is a ground truth
- Noisy votes around this truth are drawn according to a prob. distribution
- Can a voting rule discover the truth?

- Questions:
  - How many samples do we need in the Mallows prob. distribution?
  - For more general distributions, can voting rules discover the truth with infinitely many samples?

Practical motivation

- Human computation
  - EteRNA, Foldit, Crowdsourcing, etc.
  - How many users/ workers are required?

- Judgment aggregation
  - Jury system, experts ranking, etc.
  - How many experts are required?

Some notation

- A: the set of m alternatives
  - E.g., A = {a,b,c}
- L(A): the set of all possible rankings of the alternatives in A
- $\sigma \in L(A)$: a vote (ranking of the alternatives)
  - E.g., $\sigma = b > c > a$

Noise models

- An example (assuming a true ranking σ^*)
  - Rank any pair of alternatives as in σ^* with prob. p and incorrectly with prob. 1-p
  - If no ranking is defined, repeat

- Proposed by Condorcet; today known as Mallows model
- $\Pr[\sigma|\sigma^*] \sim \exp(-\Theta(d_{KT}(\sigma,\sigma^*)))$
  - where d_{KT} is the Kendall tau distance

Kendall what…? Distance functions between rankings

- A function $d: L(A)^2 \to R_{\geq 0}$ is called a distance function when
  - $d(\sigma,\sigma') = 0$ if and only if $\sigma = \sigma'$
  - $d(\sigma,\sigma') = d(\sigma',\sigma)$
  - $d(\sigma,\sigma') \leq d(\sigma,\tau)+d(\tau,\sigma')$
- Example:
  - Kendall tau: number of disagreements between all pairs of alternatives
  - Other: footrule, max displacement, Caley, Hamming

Example

- $\sigma_1 = a > b > c > d > e$
- $\sigma_2 = b > c > a > e > d$

- Kentall tau = 3
- Footrule = 6
- Max displacement = 2

Voting rules

- n voters
- Voting rule r: $L(A)^n \rightarrow L(A)$
  - defined for all values of $n>0$
- $\pi \in L(A)^n$: a profile of the votes

- The voting rule computes an aggregate ranking for each profile

- Can also be randomized
- Also known as social welfare functions

Voting rules: some examples

- Positional scoring rules:
  - Scoring vector $(\alpha_1,\alpha_2,\ldots,\alpha_m)$
  - Each voter awards α_1 points to his most preferred alternative, α_2 points to his second most preferred one, and so on
  - The alternatives are sorted in the descending order of their total points
- Kemeny:
  - Compute the ranking that minimizes the sum of Kendall tau distances from all votes

What is the sample complexity of voting rules in the Mallows model?

- Sample complexity $N^r(\epsilon)$
  - minimum number of samples so that the accuracy (i.e., the probability that voting rule r returns the ground truth) is at least $1-\epsilon$
- Theorem: Kemeny with uniform tie breaking has optimal Mallows sample complexity over all randomized voting rules

Sample complexity bounds

- Theorem: Kemeny has Mallows sample complexity $O(\log(m/\epsilon))$
- Theorem: Plurality has exponential Mallows sample complexity
- Theorem: All scoring rules with adjacent scores differing by at most U and at least L have Mallows sample complexity $\mathrm{poly}(m, U/L, \log 1/\epsilon)$
  - Borda, Harmonic

- Open question: What is the optimal Mallows sample complexity among scoring rules?

What about other voting rules?

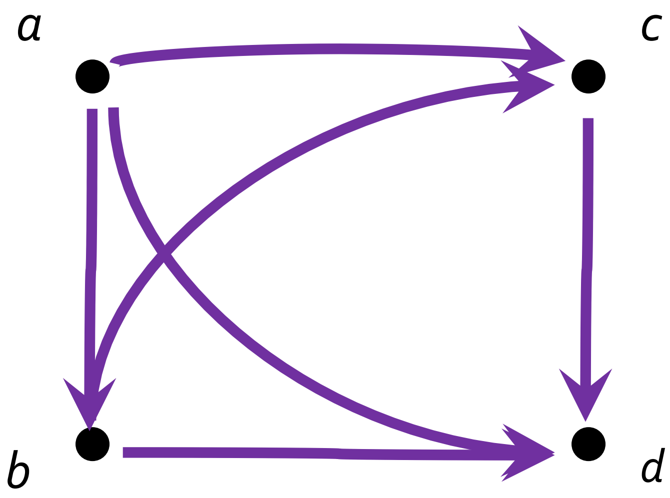
- Two general classes of voting rules:
  - Pairwise majority consistent (PM-c) voting rules
  - Position dominance consistent (PD-c) voting rules
- When there is "consensus" among the voters about the winning ranking, the voting rule should return this ranking as well
- So, the two classes will be defined by two different definitions of "consensus"

Pairwise majority consistent (PM-c) voting rules

- Given a profile π, its pairwise majority (PM) graph is defined as follows:
  - The alternatives are the nodes
  - For any pair of nodes a and b, there is a directed edge from a to b if the majority of the voters prefer a to b
- When the PM-graph of π is complete and acyclic, it reduces to a ranking σ
- A voting rule r is PM-c if r(π)= σ whenever the PM-graph of the profile π reduces to the ranking σ

An example

voter 1	voter 2	voter 3	voter 4	voter 5
a	*b*	*a*	*d*	*a*
b	*c*	*b*	*b*	*d*
c	*d*	*c*	*a*	*b*
d	*a*	*d*	*c*	*c*



Outcome
a
b
c
d

Position dominance consistent (PD-c) voting rules

- An alternative a position-dominates another alternative b if, for every i, a appears more times in the first i positions of votes than b
- PD-graph: directed graph indicating position domination between alternatives
- When the PD-graph of π is complete, it reduces to a ranking σ
- A voting rule r is PD-c if $r(\pi) = \sigma$ whenever the PD-graph of π reduces to σ

PM-c vs PD-c rules

- No voting rule can be both PM-c and PD-c

4 votes	2 votes	3 votes	2 votes
a	b	b	c
b	a	c	a
c	c	a	b

  - The PM-c and PD-c graphs reduce to a>b>c and b>a>c respectively

- PM-c: Kemeny, Ranked pairs, Copeland, Schulze
- PD-c: scoring rules, Bucklin

More sample complexity bounds

- Any PM-c rule has Mallows sample complexity $O(\log(m/\epsilon))$
- Exponential bounds for PD-c rules

Why just Mallows?

- Let's generalize: define the probabilities $\Pr[\sigma|\sigma^*]$ in a different way
- A noise model is called monotonic with respect to a distance function d if (for true ranking σ^*)
  - $d(\sigma,\sigma^*) < d(\sigma',\sigma^*)$ implies $\Pr[\sigma|\sigma^*] > \Pr[\sigma'|\sigma^*]$ and $d(\sigma,\sigma^*) = d(\sigma',\sigma^*)$ implies $\Pr[\sigma|\sigma^*] = \Pr[\sigma'|\sigma^*]$

What should we hope for with general noise models?

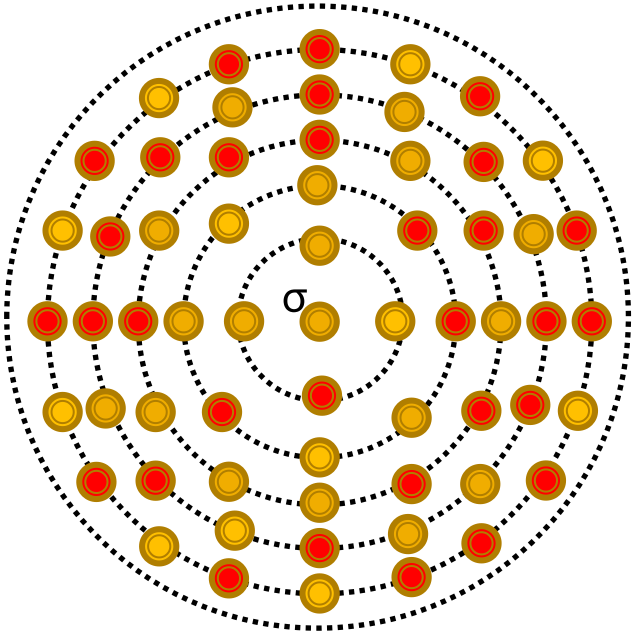
- Sample complexity can be huge
- What about accuracy in the limit?
  - We require that the true ranking is returned with prob. 1 when a voting rule is applied on infinitely many samples
  - For example, all rules mentioned are accurate in the limit for Mallows
- A voting rule is d-monotone robust if it is accurate in the limit for all d-monotonic noise models

Two characterization results

- When all PM-c/PD-c voting rules are d-monotone robust?
  - i.e., accurate in the limit for every d-monotonic noise model
- Theorem: All PM-c voting rules are d-monotone robust iff d is majority concentric (MC)
- Theorem: All PD-c voting rules are d-monotone robust iff d is position concentric (PC)

Majority concentric (MC) distances

- For any integer k, ranking σ and alternatives a, b such a > b in σ, the number of rankings with a > b at distance at most k is not smaller than the number of rankings with b > a



Position concentric (PC) distances

- A bit more technical definition that takes into account the appearances of the alternatives in the i top positions of rankings

Have we generalized enough?

- Corollary: All PM-c and PD-c voting rules are d-monotone robust iff d is both MC and PC
- Kentall tau, footrule, and max displacement are both MC and PC

Last slide

- Summary of contribution
  - Sample complexity of voting rules in the Mallows model
  - Generalizations to other noise models using the relaxed requirement of accuracy in the limit
- Very recent work
  - Modal ranking: monotone-robust wrt all distance functions