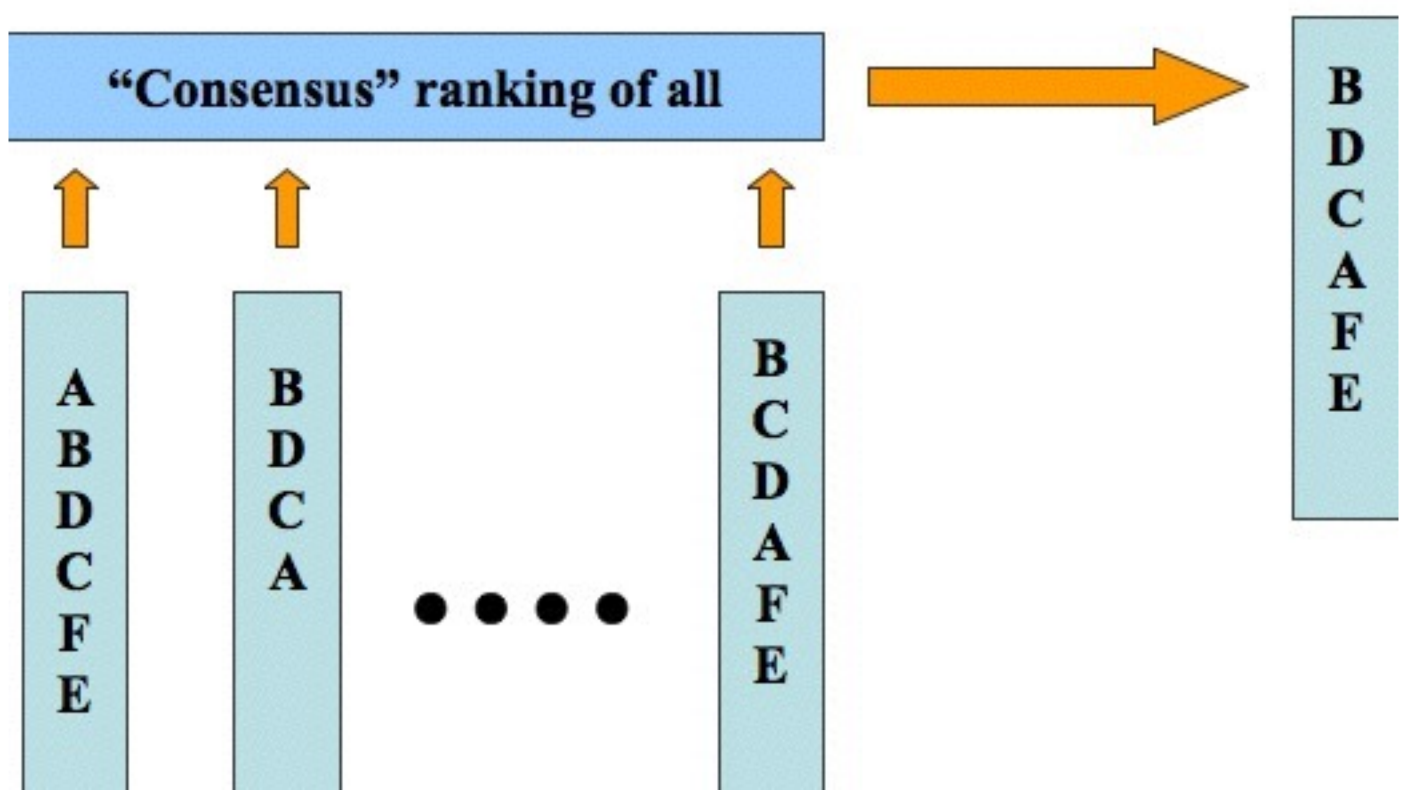
The Distributional Rank Aggregation Problem, and an Axiomatic Analysis

Pradeep Ravikumar
University of Texas at Austin
On leave at Vatic Labs

Joint work with Adarsh Prasad, Harsh Pareek

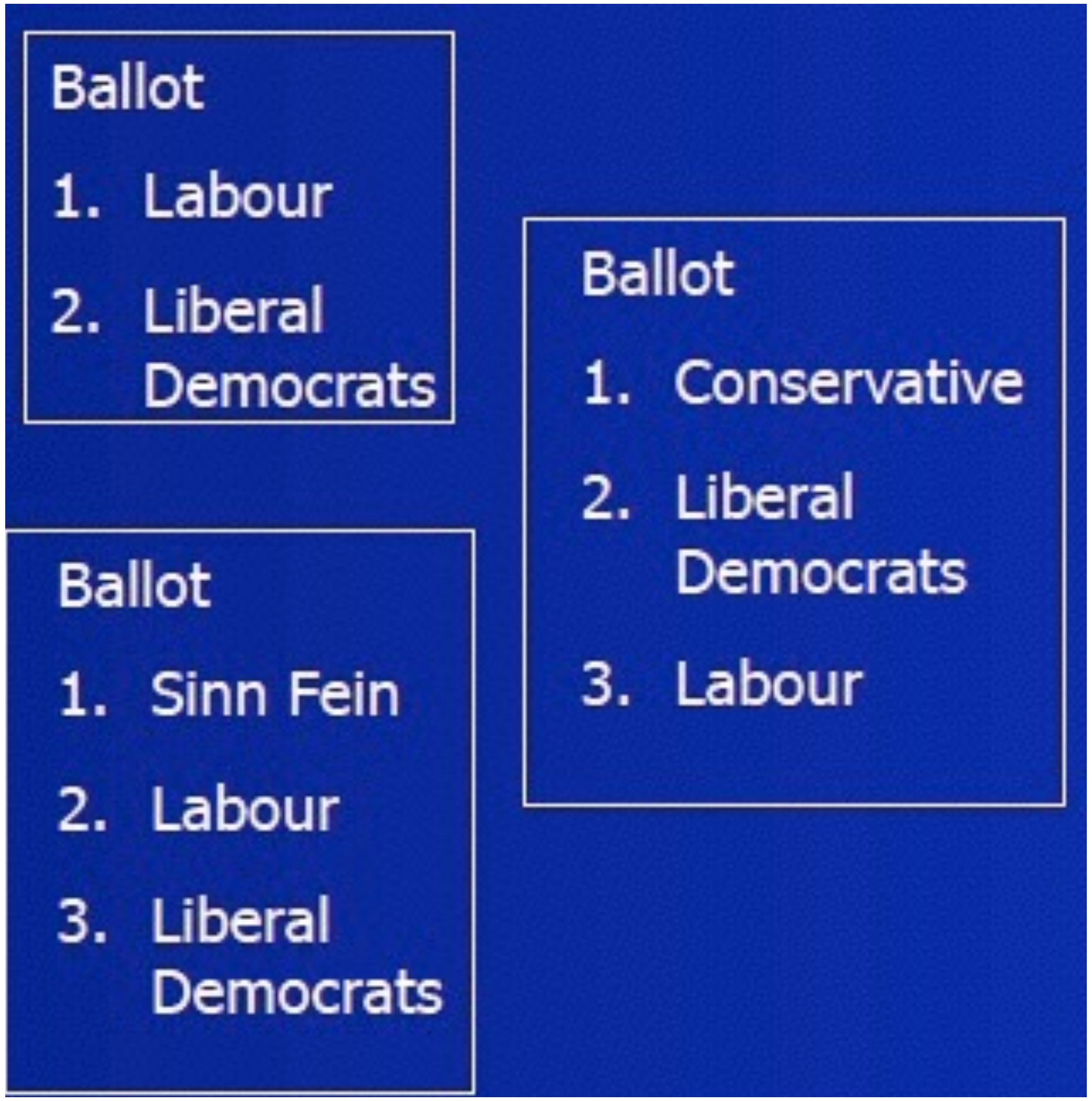
Rank Aggregation

- A set of agents provide their ranked preferences over a fixed set of alternatives
- We wish to aggregate them into a consensus ranking



Application: Elections

- Selecting representatives in elections



Source: David Williamson, The Rank Aggregation Problem

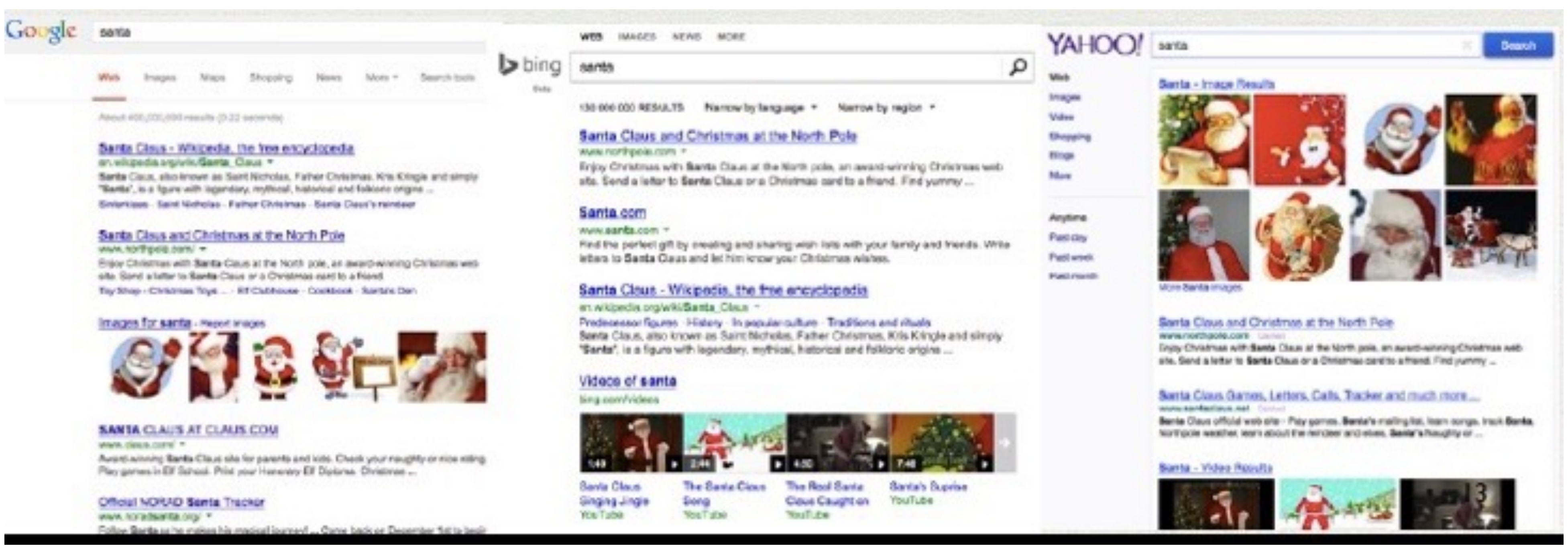
Application: Competitions

- Aggregating scores in competitions

						Aggregation		
Rank	Judge A	Judge B	Judge C	Judge D	Judge E	sum	average	rank
item 1	5	6	2	9	1	23	4.6	4
item 2	7	3	6	5	7	28	5.6	8
item 3	5	3	2	3	3	16	3.2	2
item 4	2	9	6	9	3	29	5.8	9
item 5	7	3	9	1	7	27	5.4	6
item 6	1	9	6	5	6	27	5.4	6
item 7	2	8	1	5	7	23	4.6	4
item 8	2	6	9	4	10	31	6.2	10
item 9	10	1	2	5	1	19	3.8	3
item 10	7	1	2	1	3	14	2.8	1

Application: Search Engines

- Meta-search engines that aggregate rankings from different search engines



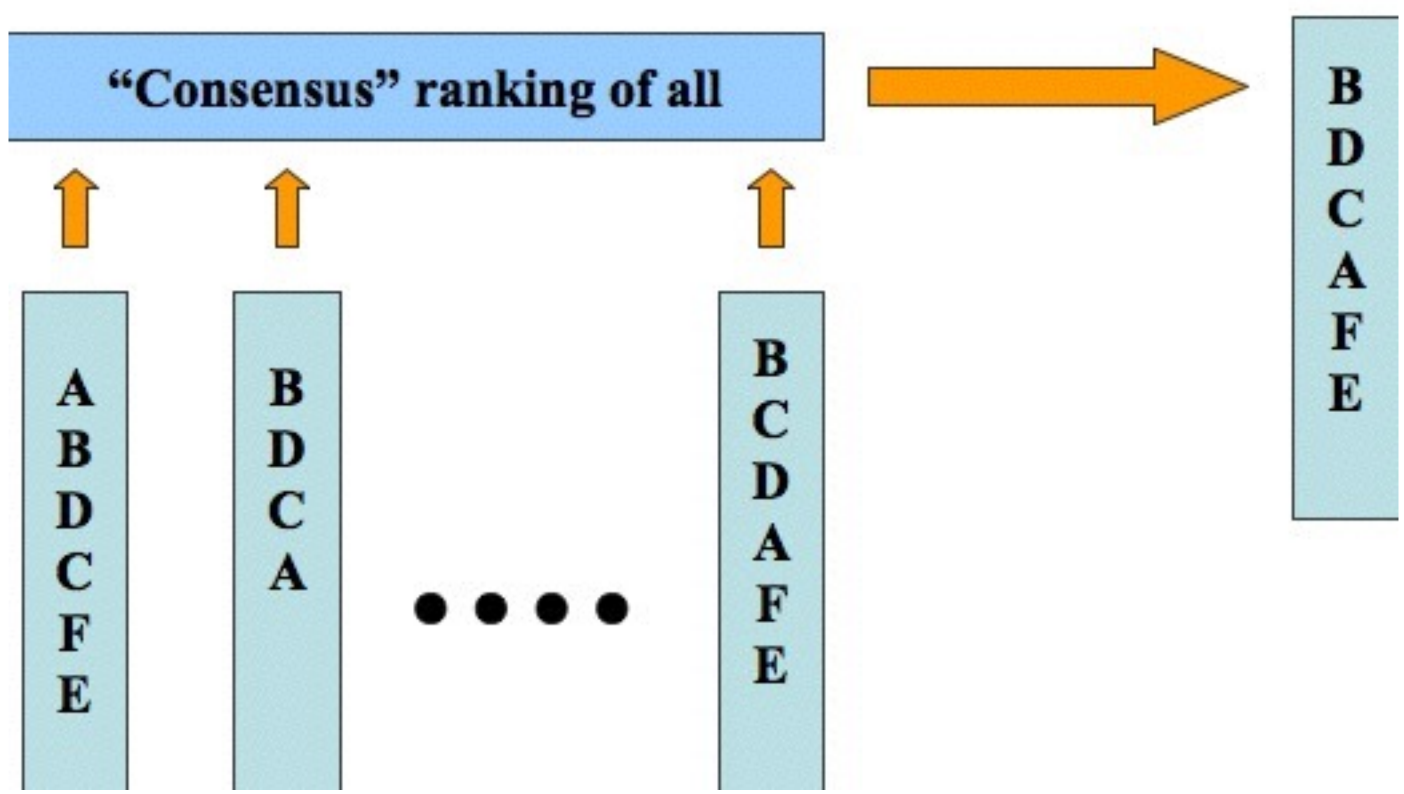
Source: Aris Gionis, Algorithmic Methods for Data Mining

Rank Aggregation

- Has been studied in varied communities
- **Statistics:** Modeling distributions over permutations e.g. Mallows Model
- **Social Choice/Welfare Theory:** Normative Axioms
- **Theoretical Computer Science:** Distance based procedures e.g. Kemeny Rule using the Kendall-Tau distance; NP-hardness, approximation results
- **Information Retrieval:** Meta Search

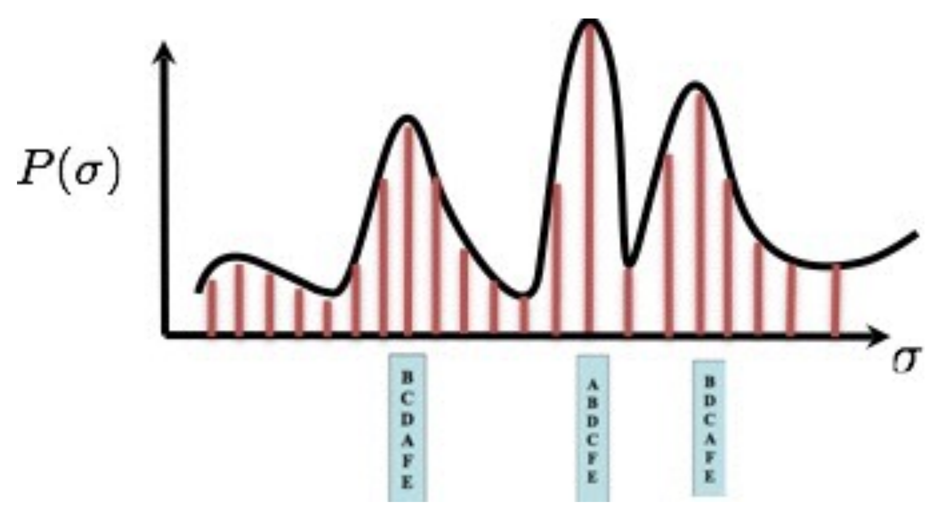
Distributional Rank Aggregation

- In real-world, no access to voters
- Access to only the **histogram** of the ranking preferences
  - how many voters for each ranking preference



Distributional Rank Aggregation

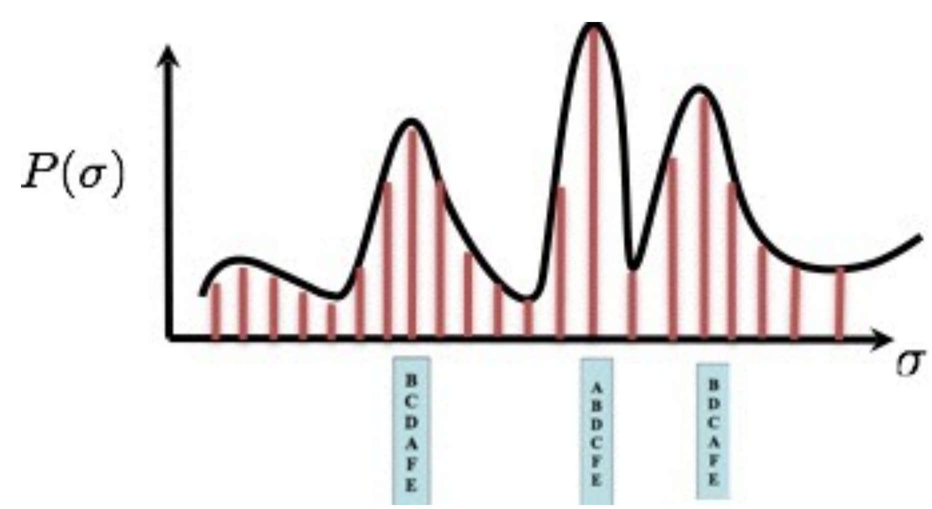
- In real-world, no access to voters
- Access to only the **histogram** of the ranking preferences
  - how many voters for each preference
- We term this rank aggregation variant that takes as input only the histogram as **Distributional Rank Aggregation**.



Distributional Rank Aggregation

- We term this rank aggregation variant that takes as input only the histogram as **Distributional Rank Aggregation**.

- Note that this is distinct from a statistical assumption: we do not assume that the ranking preferences of voters are drawn from any distribution (such as the Mallows model)



A Normative Approach
(instead of a decision-theoretic approach)

- How do we evaluate a distributional rank aggregation algorithm?

A Normative Approach (instead of a decision-theoretic approach)

- How do we evaluate a distributional rank aggregation algorithm?

- As is standard in machine learning, we could use a decision theoretic/utility maximization approach

A Normative Approach (instead of a decision-theoretic approach)

- How do we evaluate a distributional rank aggregation algorithm?

- As is standard in machine learning, we could use a decision theoretic/utility maximization approach
  - A **utility function** (or on the flip side, a **loss function**) assigning a score to any consensus ranking given a target consensus ranking

A Normative Approach (instead of a decision-theoretic approach)

- How do we evaluate a distributional rank aggregation algorithm?

- As is standard in machine learning, we could use a decision theoretic/utility maximization approach

  - A **utility function** (or on the flip side, a **loss function**) assigning a score to any consensus ranking given a target consensus ranking

  - But which evaluation metric/utility to use?

A Normative Approach
(instead of a decision-theoretic approach)

- How do we evaluate a distributional rank aggregation algorithm?

- As is standard in machine learning, we could use a decision theoretic/utility maximization approach

  - A **utility function** (or on the flip side, a **loss function**) assigning a score to any consensus ranking given a target consensus ranking

  - But which evaluation metric/utility to use?

- Alternatively, a normative approach: evaluate a rank aggregation algorithm by whether, and to what extent, it satisfies **reasonable axioms**

A Normative Approach (instead of a decision-theoretic approach)

- How do we evaluate a distributional rank aggregation algorithm?

- As is standard in machine learning, we could use a decision theoretic/utility maximization approach

    - A **utility function** (or on the flip side, a **loss function**) assigning a score to any consensus ranking given a target consensus ranking

    - But which evaluation metric/utility to use?

- Alternatively, a normative approach: evaluate a rank aggregation algorithm by whether, and to what extent, it satisfies **reasonable axioms**

- Results in this vein have been obtained for clustering by Ackerman et al (NIPS 2008, 2010)

Social Welfare Axioms

- What are good properties that an aggregation procedure should satisfy?
- Dates back to 1700s French philosophers and mathematicians.

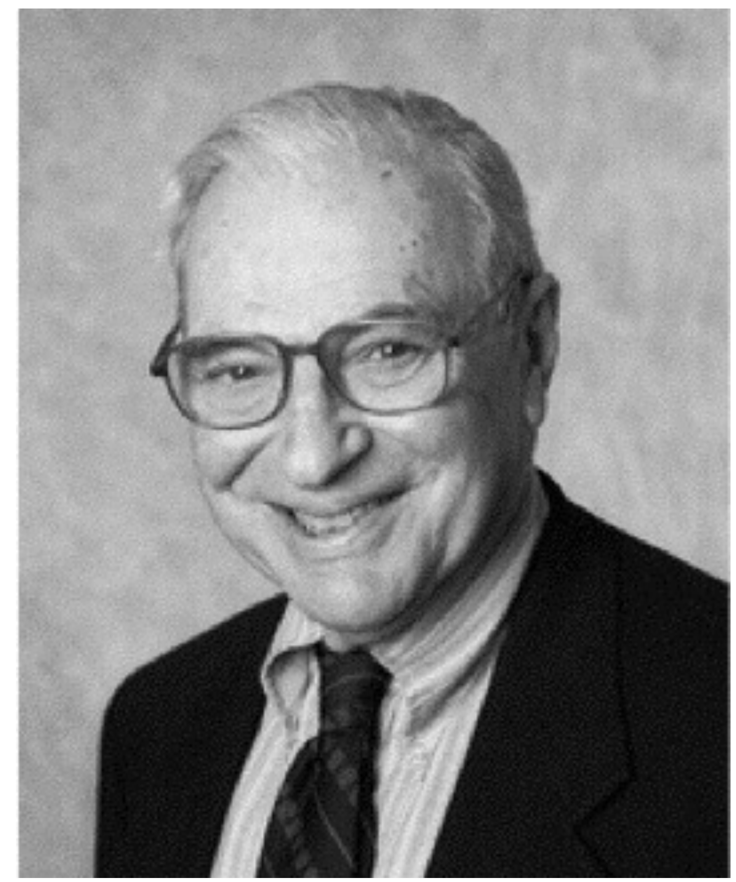
Condorcet

vs

Borda

Social Welfare Axioms

- Dates back to 1700s French philosophers and mathematicians.
- More modern attempts to construct an aggregation system that satisfies a set of natural axioms
- **Impossibility Result [Arrow(1951)]**: reasonable axioms cannot be simultaneously satisfied.



Kenneth Arrow

Social Welfare Axioms

- Impossibility Results that state reasonable axioms cannot be simultaneously satisfied
- Axioms are typically qualitatively stated, and prone to mis-interpretation: lack of quantitative characterization
- In particular, for any aggregation procedure, notion of approximate satisfiability of an axiom is missing.

Outline for rest of talk

- **Translate** social choice axioms to the distributional ranking setting.
- **Characterize** axioms quantitatively
  - Reduce misinterpretation.
  - Understand the underlying connections between axioms.
  - Relaxed-variants of axioms
- **Finesse** Impossibility Results:
  - Show it is possible to satisfy reasonable axioms simultaneously, if approximately

Translating Social Choice Axioms to setting of Distributional Rank Aggregation

Translating Social Choice Axioms

- Given distribution P over ranking preferences, what properties should the consensus ranking σ^*_P satisfy?

Pareto Efficiency

- For every item pair x and y, if everyone prefers x to y, then x is preferred to y in the resulting social preference order.

Pareto Efficiency

- For every pair x and y of alternatives, if everyone prefers x to y, then x is preferred to y in the resulting social preference order.

$$\text{If } \forall\sigma : \sigma(x) < \sigma(y), \text{ then } \sigma^*_P(x) < \sigma^*_P(y).$$

in the support of P

Consensus ranking given distribution P

Independence of Irrelevant Alternatives (IIA)

- If voters change their preferences, but keep their relative positions of x and y, then the relative positions of x and y in the aggregation should still remain the same

Independence of Irrelevant Alternatives (IIA)

- If voters change their preferences, but keep their relative positions of x and y, then the relative positions of x and y in the aggregation should still remain the same

Let $P_{x<y} = \sum_{\sigma:\sigma(x)<\sigma(y)} P(\sigma)$ be the marginal probability of x being preferred over y.

If for any two distributions P and Q, $P_{x<y} = Q_{x<y}$, then

Independence of Irrelevant Alternatives (IIA)

- If voters change their preferences, but keep their relative positions of x and y, then the relative positions of x and y in the aggregation should still remain the same

Let $P_{x<y} = \sum_{\sigma:\sigma(x)<\sigma(y)} P(\sigma)$ be the marginal probability of x being preferred over y.

If for any two distributions P and Q, $P_{x<y} = Q_{x<y}$, then

$$sign\left[\sigma_P^*(x) - \sigma_P^*(y)\right] = sign\left[\sigma_Q^*(x) - \sigma_Q^*(y)\right].$$

Majority Rule

- For any set of voters, if an alternative x is ranked in the top position in strictly more than half the votes, then in the final aggregation, x should be ranked at the highest position

Majority Rule

- For any set of voters, if an alternative x is ranked in the top position in strictly more than half the votes, then in the final aggregation, x should be ranked at the highest position

Let $P^1(x) = \sum_{\sigma:\sigma(x)=1} P(\sigma)$ denote the marginal probability of x being ranked first.

If $P^1(x) > \frac{1}{2}$, then $\sigma^*_P(x) = 1$.

Condorcet Criterion

- If there exists an item x that is preferred over every other item y, by strictly more than half of the voters, then x should be ranked at the highest position in the aggregation.

Condorcet Criterion

- If there exists an item x that is preferred over every other item y, by strictly more than half of the voters, then x should be ranked at the highest position in the aggregation.

For any distribution P, if $P_{x<y} > \frac{1}{2} \; \forall \; y \in \mathcal{X} \backslash \{x\}$, then $\sigma_P^*(x) = 1$.

Relaxed Variants of Axioms

Independence of Irrelevant Alternatives (IIA)

- Exact Axiom:

Let $P_{x<y} = \sum_{\sigma:\sigma(x)<\sigma(y)} P(\sigma)$ be the marginal probability of x being preferred over y.

If for any two distributions P and Q, $P_{x<y} = Q_{x<y}$, then

$$sign\left[\sigma_P^*(x) - \sigma_P^*(y)\right] = sign\left[\sigma_Q^*(x) - \sigma_Q^*(y)\right].$$

Independence of Irrelevant Alternatives (IIA)

- Exact Axiom:

Let $P_{x<y} = \sum_{\sigma:\sigma(x)<\sigma(y)} P(\sigma)$ be the marginal probability of x being preferred over y.

If for any two distributions P and Q, $P_{x<y} = Q_{x<y}$, then

$$sign\left[\sigma_P^*(x) - \sigma_P^*(y)\right] = sign\left[\sigma_Q^*(x) - \sigma_Q^*(y)\right].$$

- epsilon-relaxed Axiom:

If for any two distributions P and Q, $P_{x<y} = Q_{x<y} = \gamma$, where γ satisfies:

$$|\gamma - \frac{1}{2}| \geq \epsilon, \epsilon > 0,$$

then $sign\left[\sigma_P^*(x) - \sigma_Q^*(y)\right] = sign\left[\sigma_P^*(x) - \sigma_Q^*(y)\right].$

Majority Rule

- Exact Axiom:

Let $P^1(x) = \sum_{\sigma:\sigma(x)=1} P(\sigma)$ denote the marginal probability of x being ranked first.

If $P^1(x) > \frac{1}{2}$, then $\sigma^*_P(x) = 1$.

Majority Rule

- Exact Axiom:

Let $P^1(x) = \sum_{\sigma:\sigma(x)=1} P(\sigma)$ denote the marginal probability of x being ranked first.

If $P^1(x) > \frac{1}{2}$, then $\sigma^*_P(x) = 1$.

- epsilon-relaxed Axiom:

If $P^1(x) \geq \frac{1}{2} + \epsilon\ ; \epsilon > 0$, then $\sigma^*_P(x) = 1$.

Condorcet Rule

- Exact Axiom:

For any distribution P, if $P_{x<y} > \frac{1}{2} \; \forall \; y \in \mathcal{X} \backslash \{x\}$, then $\sigma^*_P(x) = 1$.

Condorcet Rule

- Exact Axiom:

For any distribution P, if $P_{x<y} > \frac{1}{2} \; \forall \; y \in \mathcal{X} \backslash \{x\}$, then $\sigma^*_P(x) = 1$.

- epsilon-relaxed Axiom:

For any distribution P, if $P_{x<y} \geq \frac{1}{2} + \epsilon \; ; \epsilon > 0 \; \forall \; y \in X \backslash \{x\}$, $\sigma^*_P(x) = 1$.

Distributional Impossibility Theorem

Any distributional rank aggregation procedure can be represented as:

$$\sigma_P^* = \arg\min_{\sigma \in S_n} g(\sigma, P)$$

for some $g : S_n \times \mathcal{P}_n \mapsto \mathbb{R}$.

Theorem (Distributional Impossibility Theorem)

For $n \geq 2$, if $g(\sigma, P)$ is a continuous function of P for each fixed σ, both Universality and Pareto cannot be satisfied simultaneously.

(unique, deterministic, complete ranking)

Finessing Impossibility Theorems

Theorem (**Distributional Possibility Theorem**)

The set of positional loss functions which admits the following axioms is non-empty: exact versions of Pareto and Monotonicity, relaxed axioms ϵ-IIA, ϵ-Strong Condorcet, ϵ'-Majority Rule

Expected Loss Minimization

Consider rank aggregation procedures which minimize the expected value of a discrepancy measure (which we denote by ℓ) over P:

$$\sigma^*_{\ell,P} = \arg\min_{\sigma \in S_n} g(\sigma, P) = \arg\min_{\sigma \in S_n} \mathbb{E}_{\sigma' \sim P}[\ell(\sigma, \sigma')]$$

Positional Scoring Loss

Consider rank aggregation procedures which minimize the expected value of a discrepancy measure (which we denote by ℓ) over P:

$$\sigma^*_{\ell,P} = \underset{\sigma\in S_n}{\arg\min}\, g(\sigma, P) = \underset{\sigma\in S_n}{\arg\min}\, \mathbb{E}_{\sigma'\sim P}[\ell(\sigma,\sigma')]$$

Positional Scoring Loss.
A loss ℓ_h is a Positional Scoring Loss iff it can be decomposed as $\ell_h(\sigma,\sigma') = \sum_x h(\sigma'(x))\,\sigma(x)$, where $h : [n] \mapsto \mathbb{R}$.

Positional Scoring Loss

Consider rank aggregation procedures which minimize the expected value of a discrepancy measure (which we denote by ℓ) over P:

$$\sigma^*_{\ell,P} = \underset{\sigma \in S_n}{\arg\min}\, g(\sigma, P) = \underset{\sigma \in S_n}{\arg\min}\, \mathbb{E}_{\sigma' \sim P}[\ell(\sigma, \sigma')]$$

Positional Scoring Loss.
A loss ℓ_h is a Positional Scoring Loss iff it can be decomposed as $\ell_h(\sigma, \sigma') = \sum_x h(\sigma'(x))\, \sigma(x)$, where $h : [n] \mapsto \mathbb{R}$.

Examples:

Borda Count: $h(i) = n - i$

Plurality Rule: $h(i) = 1$ *if* $i = 1, 0$ *else*

Positional Scoring Rules

Name	$h : [n] \mapsto \mathbb{R}$	ϵ for IIA/Cond.	ϵ for Maj.
Borda Count	$h(i) = n - i$	$1/2 - 1/n$	$1/2 - 1/n$
Plurality Rule	$h(i) = 1$ *if* $i = 1, 0$ *else*	$1/2$	0
Anti-Plurality Rule	$h(n) = 0; h(i) = 1 \; \forall i \neq n$	$1/2$	$1/2$
Log Rule	$h(i) = -\log(i)$	$\frac{\log(n)}{2\log(n^2/(n-1))}$	$\frac{\log(n)}{2\log(n^2/(n-1))}$
Squared Rule	$h(i) = -i^2$	$\frac{n^2-4}{2(n^2+2)}$	$\frac{n^2-4}{2(n^2+2)}$

- "Axiomatic Characterization" of Positional Scoring Rules

Finessing Impossibility Theorems

Theorem (**Borda Count Optimality Theorem**)

For any fixed n, Borda count is optimal w.r.t. the ϵ-Strong Condorcet condition and ϵ-IIA, i.e. has the least ϵ among all positional loss functions.

Summary

- Distributional Rank Aggregation: given access only to histogram of ranking preferences
- We translate classical social choice axioms to the distributional ranking setting.
- Quantitative characterization of axioms:
  - Reduce misinterpretation
  - Relaxed-variants of axioms
- Finesse Impossibility Results:
  - Show it is possible to satisfy reasonable axioms simultaneously, if approximately