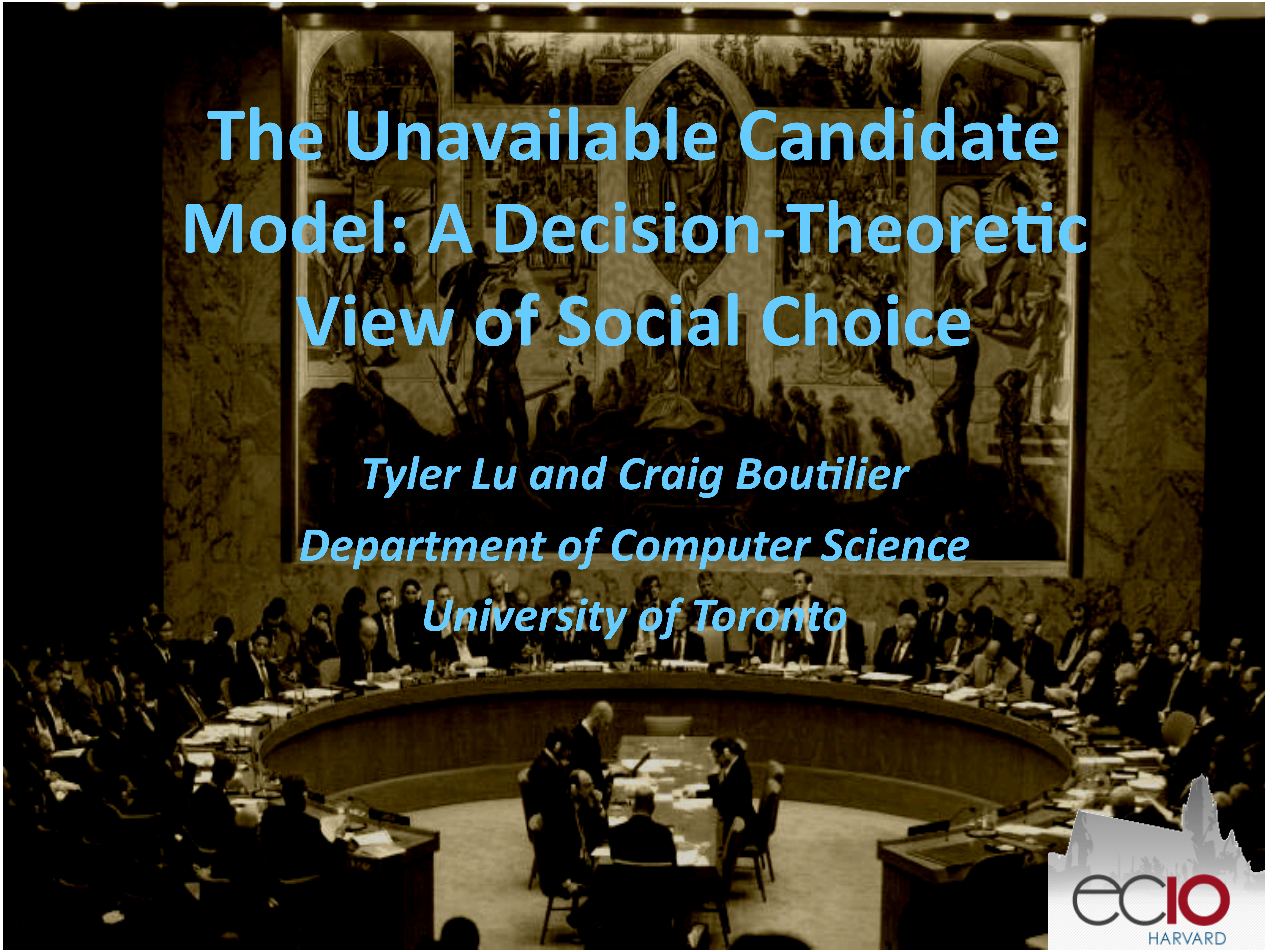


The Unavailable Candidate Model: A Decision-Theoretic View of Social Choice

Tyler Lu and Craig Boutilier

Department of Computer Science

University of Toronto

ec10
HARVARD

Introduction

- Much of social choice theory concerned with aggregating voters' preference ranking (over candidates) into a consensus ranking.
  - Started with Arrow's Theorem (1951)
  - Politics, meta web search, multi-criteria decision making, etc…
- Many aggregation methods: single transferable vote, Borda, Kemeny, …

Introduction

Why need *entire consensus ranking*?

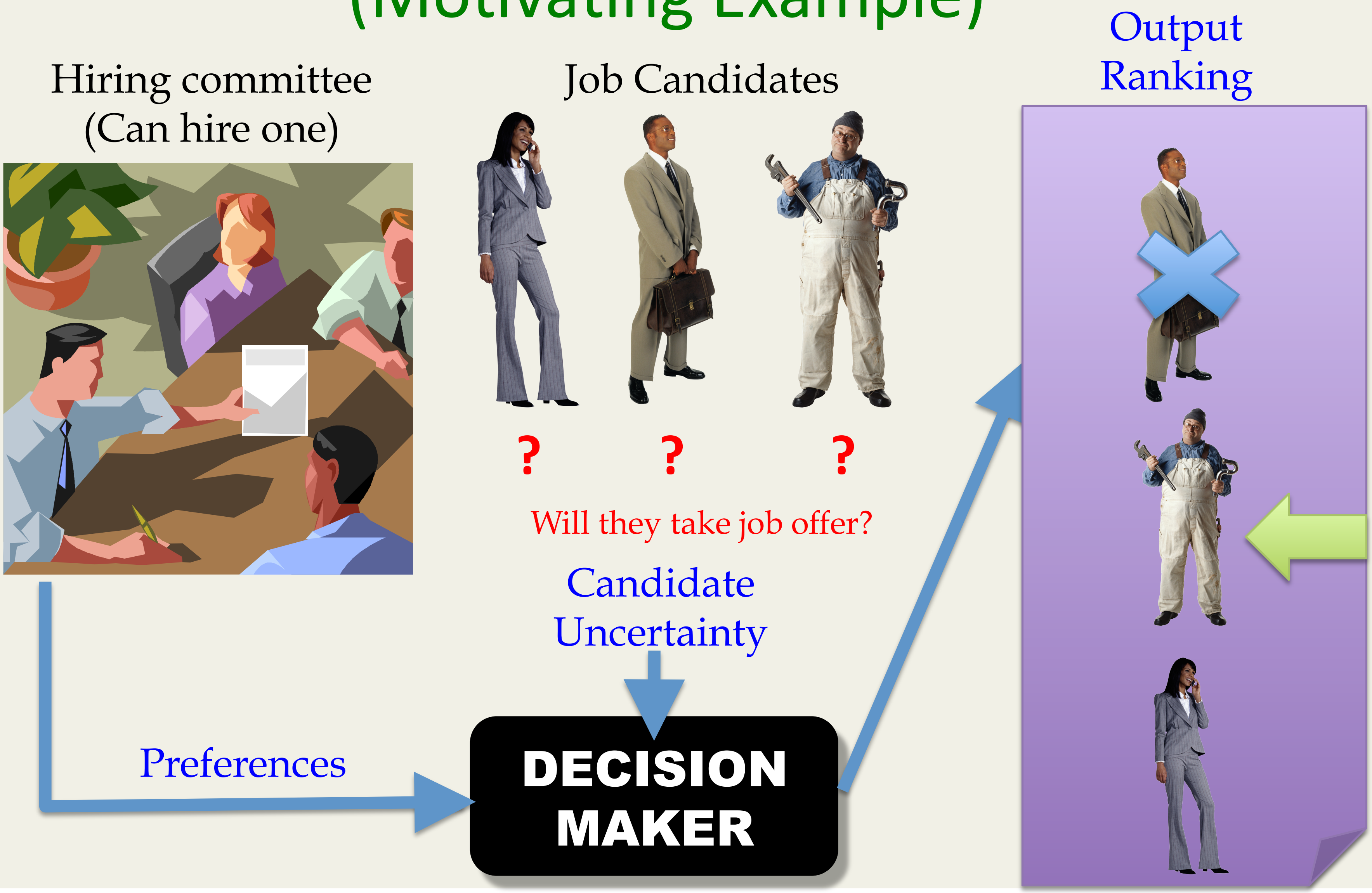
Why use a particular aggregation method (e.g. Kemeny)?

Decision criterion should directly influence how to aggregate rankings

We develop such an approach.

Unavailable Candidate Model

(Motivating Example)



Output Ranking

Hiring committee
(Can hire one)

Job Candidates

? ? ?

Will they take job offer?

Candidate Uncertainty

Preferences

DECISION MAKER

Unavailable Candidate Model

(Recently, independently developed by Baldiga & Green)

- Candidates **$C = \{c_1, \ldots, c_m\}$**
- Voters N = {1..n} with preference profile **$V = (v_1, \ldots, v_n)$** where v_i is a ranking
- Probability P(S) only $S \subseteq C$ are available.
  Can't just select winning candidate!
- Output a *decision policy (aka choice function)*

$$W : 2^C \rightarrow C \cup \{\perp\}$$

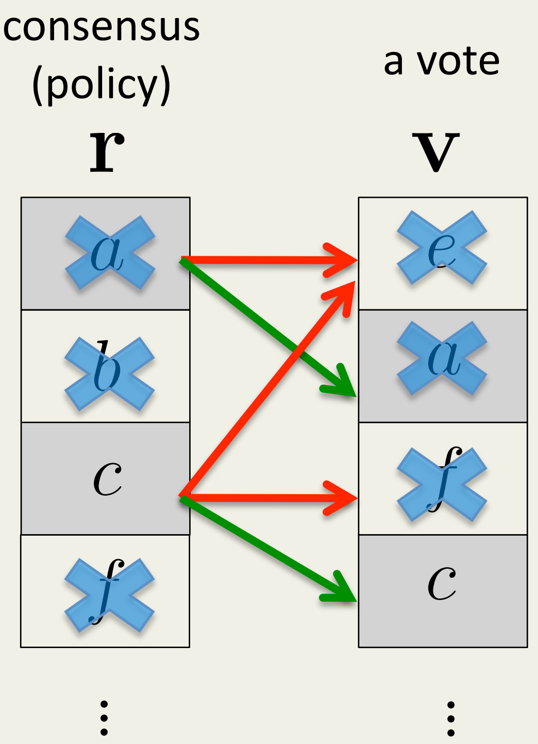
Unavailable Candidate Model

- Consider ranking policies: outputs top available candidate (aka *rationalizable choice function*)
- Other policy classes possible but rankings are compact, interpretable, easy to implement.
- Sometimes legally, or procedurally required
  - National Resident Matching Program, hospitals submit preference ranking (committee in hospital must reach *consensus*).
- Rationalizable approximation to optimal policy

Minimizing Disagreement

Find consensus ranking minimizing expected #disagreements

Disagreement Example



consensus (policy)

a vote

Comparison with Kendall-Tau (Kemeny)

Defn (Kendall-Tau distance)

$$\tau(r,v) = \sum_{\{i,j\}} 1[r,v \text{ disagree on } i,j]$$

Defn (Kemeny consensus)

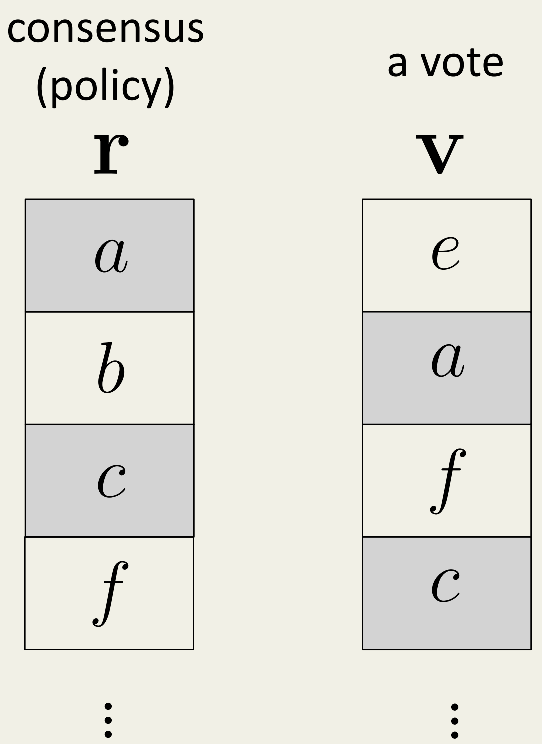
Ranking *r* that minimizes $\kappa(r,V) = \sum_{\ell=1}^{n} \tau(r,v_{\ell})$

Comparison with Kendall-Tau (Kemeny)

Kendall-tau

(=#misordered pairs)

Misordered **{*a*, *e*}** has *same "penalty"* as misordered **{*c*, *f*}**



Our expected #dis.

Prob. of disagreement "**r** outputs *a*, **v** wants *e*"

is larger than

"**r** outputs *c*, **v** wants *f*"

Kendall-tau ignores ***relative importance of misordering*** (how far down in ranking). In most cases, our cost focuses more on disagreements at top

Minimizing Disagreement

- Given *P*, *V*, minimize expected #disagreements

$$\boxed{r^{*} = \operatorname{argmin}_r \mathrm{E}_{S\sim P}\left[\sum_{\ell=1}^{n} \underbrace{\mathbf{1}[r(S) \neq \mathrm{top}(v_\ell, S)]}_{\text{disagreement}}\right]}$$

- Simple *P*: prob. *p* *any* candidate is unavailable
- Can be shown expected #disagreements is

$$\boxed{\sum_{\ell=1}^{n}\sum_{i=1}^{m}(1-p)p^{r(c_i)-1}\left(1-p^{v_\ell(c_i)-t(c_i,r,v_\ell)-1}\right)}$$

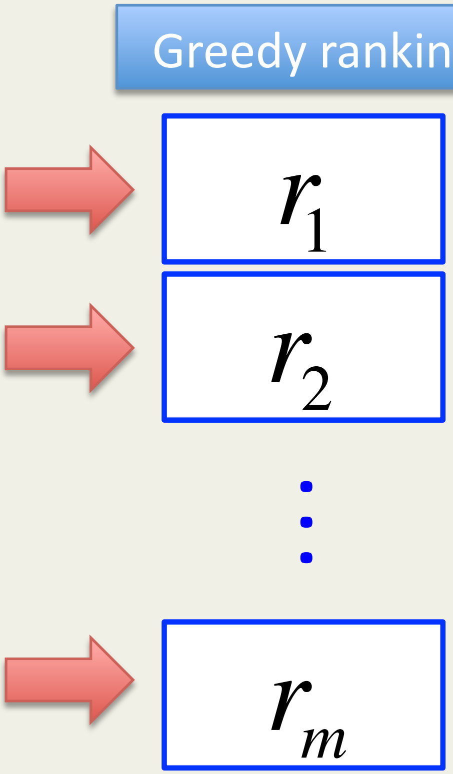
Candidates lower in the ranking contribute less to the expected #disagreements

Computational Optimization

- NP-hard
  - Given *p, V,* and threshold $t \geq 0$, exists ranking *r* with expected #disagreements $\leq t$?
- For optimal ranking
  - Can formulate as integer program
  - Large #vars and constraints (though polynomial)
  - CPLEX is slow to solve
- However, very good greedy algorithm

Greedy Algorithm

- Let $D^k(r, V)$ be expected #disagreements when k^{th} ranked candidate of r is top available



$r_1 = \text{argmin}_c D^1(c\cdots,V)$

$r_2 = \text{argmin}_c D^2(r_1 c\cdots,V)$

(actual algorithm is bit more clever)

Approximation ratio at most $1 + 2p/(1-p)^2$

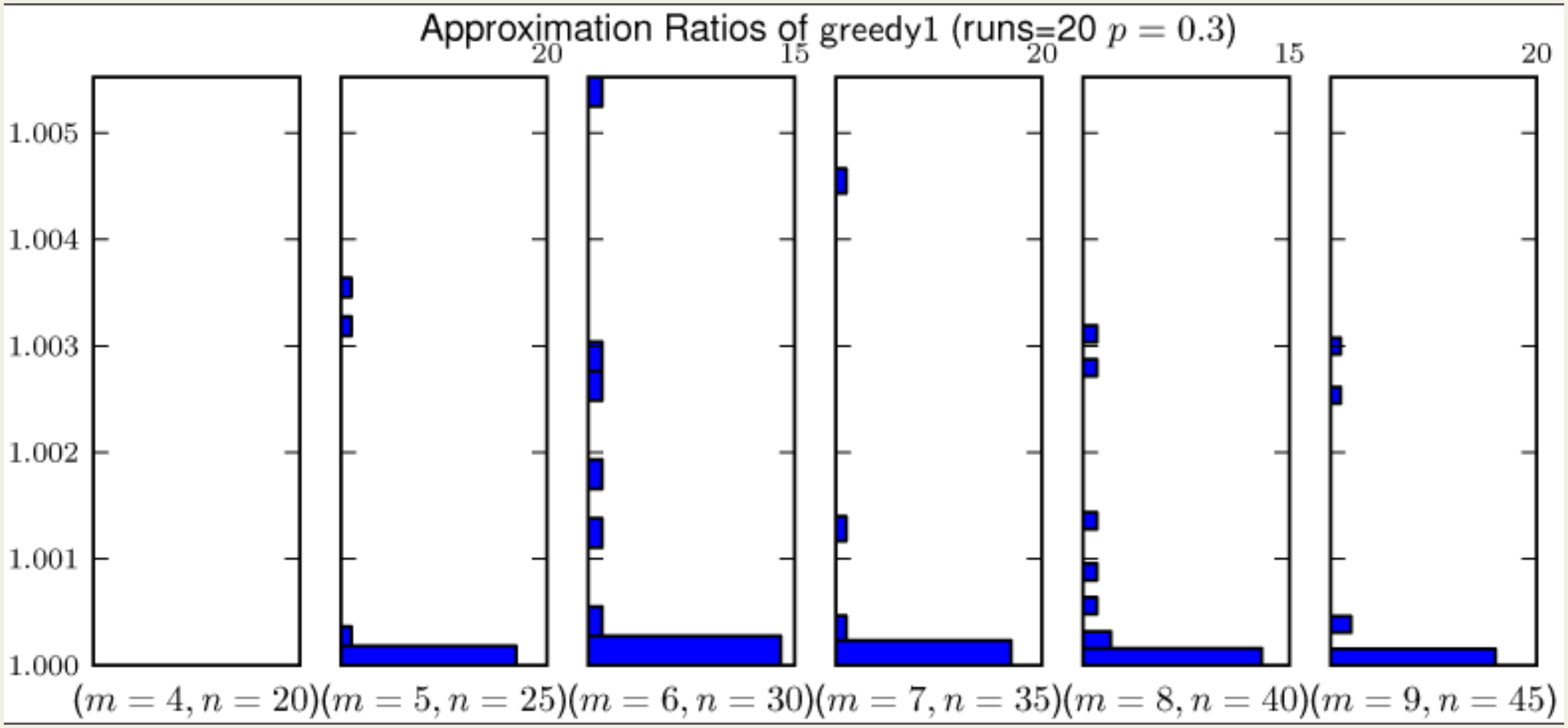
Reminder p is prob. of **unavailable**

Bound likely loose, works very well experimentally!

Greedy Algorithm

Rotated Histograms of greedy performance

Approximation ratios



Varying pairs (m = #candidates, n = #voters)

PolyTime Approximation Scheme

Idea: get top positions of ranking "right"

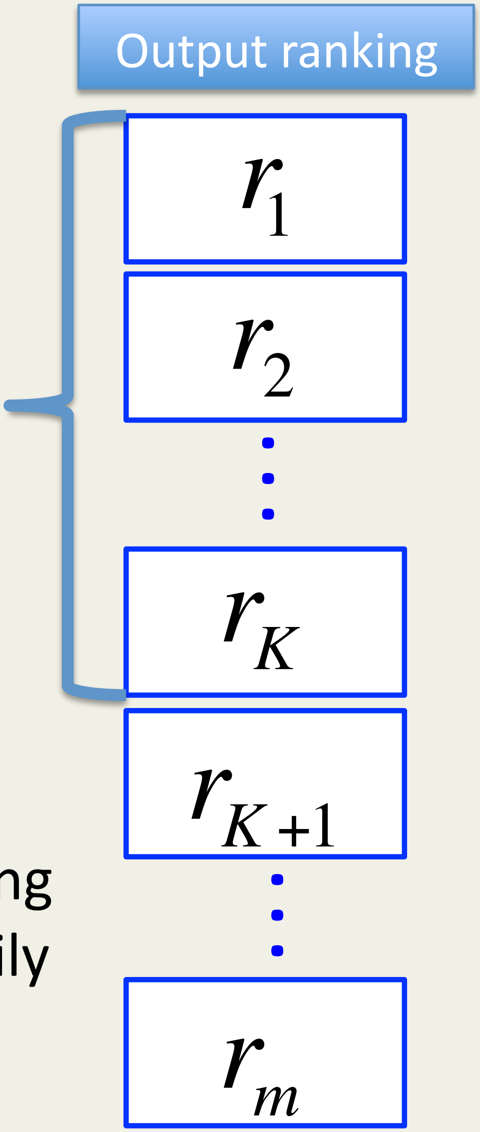
Find top *K* candidates

$$\operatorname{argmin}_{r_1\cdots r_K} \sum_{\ell=1}^{n} \sum_{k=1}^{K} D^k(r_1 \cdots r_K \cdots, V)$$

Approximation ratio at most $1 + 2p^K/(1-p)^2$

Make less than $1 + \varepsilon$ by

$$K = \left\lceil \log\frac{2}{\varepsilon(1-p)^2} \Big/ \log\frac{1}{p} \right\rceil$$



Output ranking

r_1

r_2

$\vdots$

r_K

r_{K+1}

$\vdots$

r_m

Order remaining arbitrarily

Connection to Plurality Voting

- When p = 0 (all candidates available), then top candidate in an optimal ranking is one that receives the largest number of first place votes.
- A consequence of the definition of "disagreement"

Connections to Kemeny

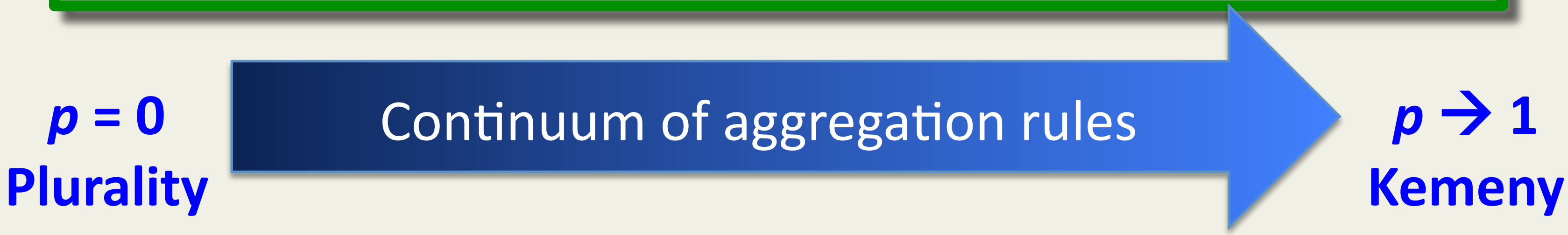
- Focus of much work in computational social choice
- Max likelihood estimator of a distribution with a modal ranking (Young'95, Mallows'57)
  - Votes are I.I.D. samples of "objective" ranking
  - Motivation is ***not*** decision-theoretic
  - Aggregation is *statistical inference*

Connections to Kemeny

Theorem

As p "approaches 1", the following holds

1. Any optimal ranking is also a Kemeny consensus
2. A Kemeny consensus K^* may not minimize expected #disagreements, however,
3. Any K^* has expected #disagreements at most factor of $1 + \varepsilon$ worse than optimal (ε depends on p, m, n)



$p = 0$
Plurality

Continuum of aggregation rules

$p \to 1$
Kemeny

- Can be used to justify Kemeny as "**decision policy under uncertainty**" (in some cases)

Conclusion

Take Home Message

Decision criterion as foundation for rank aggregation

In our case, ranking to minimize expected disagreements under uncertain candidate availability

- Connections to Plurality, Kemeny
- Decision-theoretic justification for Kemeny
- Nice computational properties—good greedy algorithm, PTAS

Future Work

- More general class of distributions (e.g. tractable graphical models)
    - Computing expected disagreements
    - Optimization: approximate, exact
- Disagreement take into account strength of preferences
- How good of an approximation is optimal ranking to optimal policy? *Loss of rationalizing choice functions.*
- Incentive issues – *reporting preferences **and** probabilities*
- Other decision-theoretic models of social choice

Thank you!