Computing Optimal Bayesian Decisions for Rank Aggregation via MCMC Sampling

David Hughes, Kevin Hwang and Lirong Xia

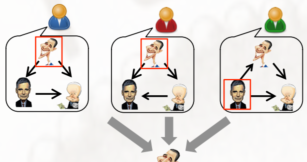
Presented by Zhibing Zhao

Rensselaer

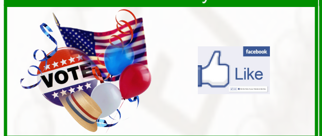
Introduction

Social Choice (Rank Aggregation)

Presidential elections
- Alternatives: candidates
- Agents: voters
- Profile: preferences of agents
- Mechanism: plurality rule
- Decision: president



Democracy



Classical social choice

The axiomatic approach [Arrow-50]

Truth

Statistical approaches

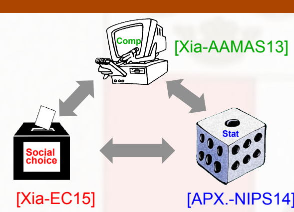
MLE; MAP

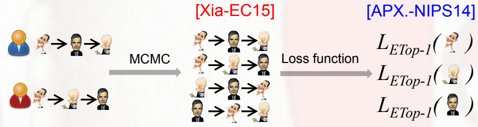
Our Goal and Approach

Mechanisms
- 2 alternatives: majority rule
- Kemeny's rule (for ranking)

Goal
- Computational
- Axiomatic
- Statistical

Our approach





Statistical decision-theoretic framework for social choice

Inputs
- statistical model: Θ, S, $\Pr_\theta(s)$
- decision space: D
- loss function: $L(\theta, d)\in\mathbb{R}$

unknown ground truth | decision to make

The mechanism

f: Profiles⟶D with minimum Bayesian expected loss:

$f(R) \in \text{argmin}_d E_{\theta|R} L(\theta,d)$

Complexity

Theorem: if BayesianLoss for Mallows' model w.r.t. the exact Top-1 loss function and uniform prior has a polynomial-time approximation algorithm with constant approximation ratio, then P=NP

Mallows' Model [Mallows-1957]

- Fix the dispersion ϕ<1
- Parameter space
  - all full rankings over alternatives
- Sample space
  - i.i.d. generated full rankings
- Probabilities: given a ground truth ranking W, generate a ranking V

  $\Pr_W(V) \propto \phi^{\text{Kendall}(V,W)}$

Decision Space Examples

- **All alternatives** (resolute voting)
- Power set of alternatives excluding the empty set (irresolute voting)
- All full rankings over alternatives

Loss Function Examples

- **Top-k**
  - $L_{ETop\text{-}k}$(W,d)=0, when d is ranked within top k
  - $L_{ETop\text{-}k}$(W,d)=1, otherwise
- Other smooth functions

References

[Mallows 1957] Colin L. Mallows. Non-null ranking model. *Biometrika*, 44(1/2): 114–130, 1957.
[Sinclair 1992] Alistair Sinclair. Improved Bounds for Mixing Rates of Markov Chains and Multicommodity Flow. *Combinatorics, Probability and Computing*, 1(4):351–370, 1992.
[Sinclair 1989] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, 1989.
[Jerrum 1996] Mark Jerrum and Alistair Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In Dorit S. Hochbaum, editor, *Approximation algorithms for NP-hard problems*, pages 482–519. PWS Publishing Company, 1996.
[Liu 1996] Jun S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119, 1996.

Our Approach

Metropolis-Hastings MCMC Algorithms

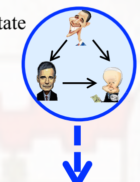
MCMC Algorithms

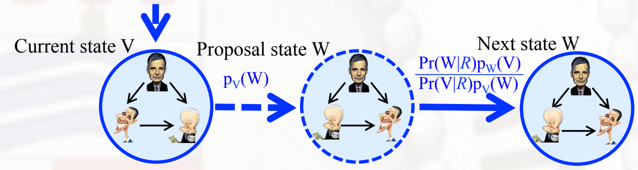
- Given profile R
- Generate N samples Q from a Markov chain with stationary distribution Pr(|R)
- Compute $\text{argmin}_d \Sigma_{\theta\in Q} L(\theta,d)$

Metropolis-Hastings Sampling

- Initial state – a ranking
- Given proposal distributions $\{p_V(\cdot):V\in\Theta\}$
  - $p_V(W) = p_W(V)$
- In each step
  - Current state V
  - Proposal next state W
  - The next state is W with probability

    $\Pr(W|R)\, p_W(V)/\Pr(V|R)\, p_V(W)$
  - Otherwise the next state is V



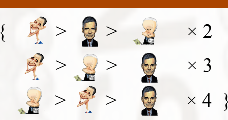


Theoretical Guarantee

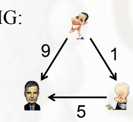
- The performance of the algorithm closely depends on the mixing time of the MC
- Theorem. The mixing time of the M-H algorithm for Mallows' model for any starting state is

  $\phi^{-k_{max}}$ Poly(other inputs)

- Profile:



- WMG:

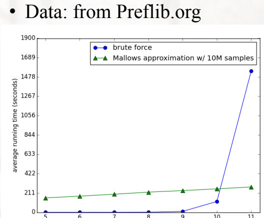


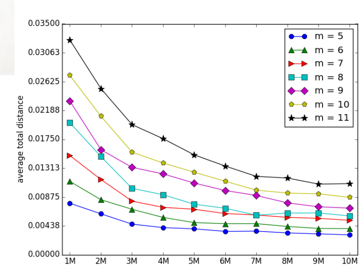
k_{max}=14

- k_{max}: maximum cut in the weighted majority graph
- ϕ: dispersion parameter of the model
- Proof uses techniques from
  - novel analysis of a set of canonical paths [Sinclair, 1992; Jerrum and Sinclair 1989, 1996; Liu 1996]

Experiments

- Mallows' model
- D: single winner
- L: Top-1 loss function
- Data: from Preflib.org





Summary & Future Research

Summary

- Mallows' model and Condorcet's model
- Bayesian loss is NP-hard to approximate
- New MCMC algorithms
  - Mallows' model: sample complexity/running time only exponential in k_{max}
  - Condorcet's model: polynomial sample complexity

Future Research

- Performance improvement of Markov chain sampler in practice
- Design and analysis of other Markov chain samplers
- Evaluation of the Markov chain approach