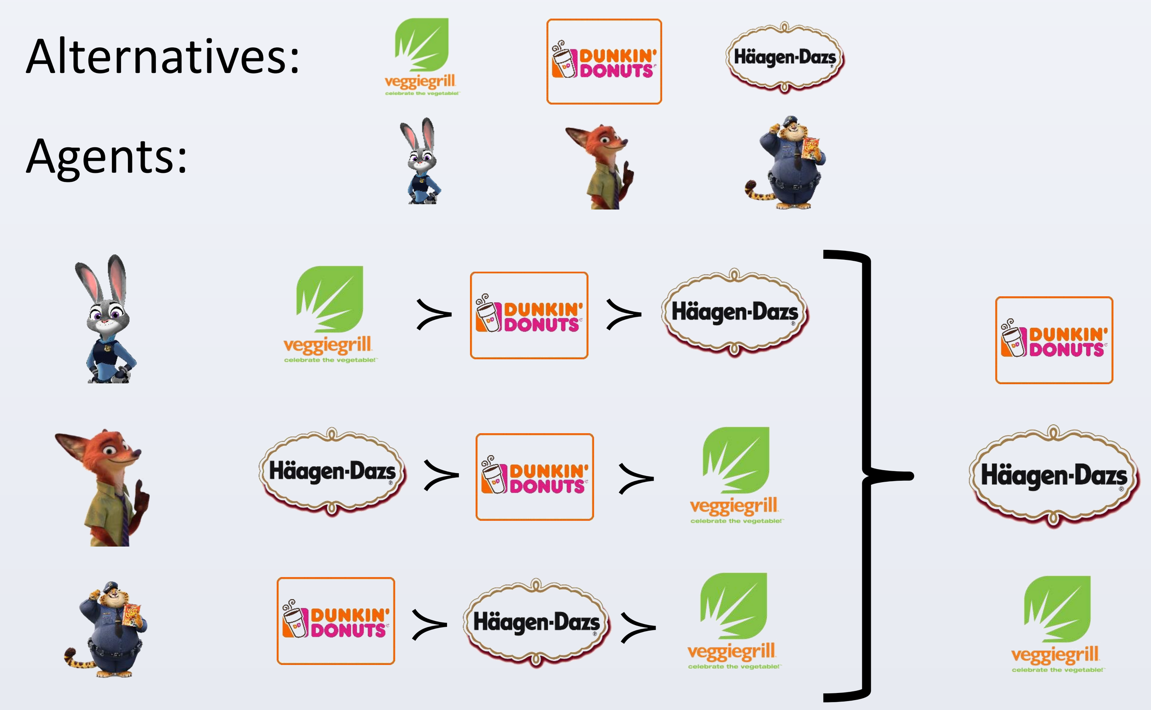
Composite Marginal Likelihood Methods for Random Utility Models

Zhibing Zhao and Lirong Xia

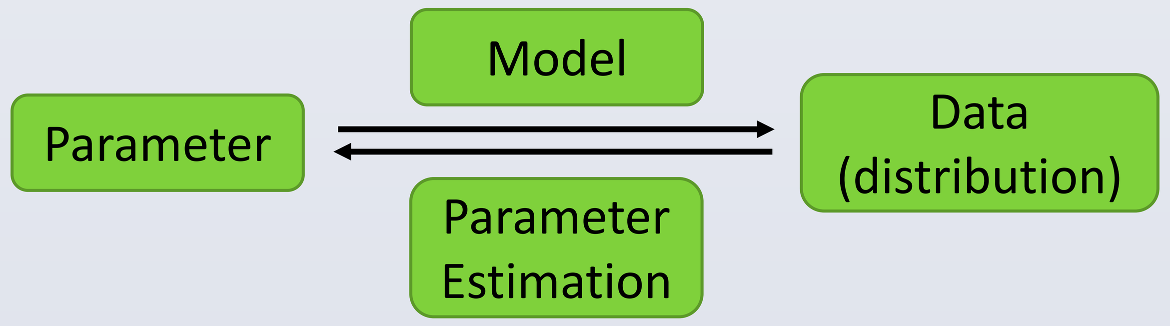
Rensselaer Polytechnic Institute

Introduction

Rank Aggregation



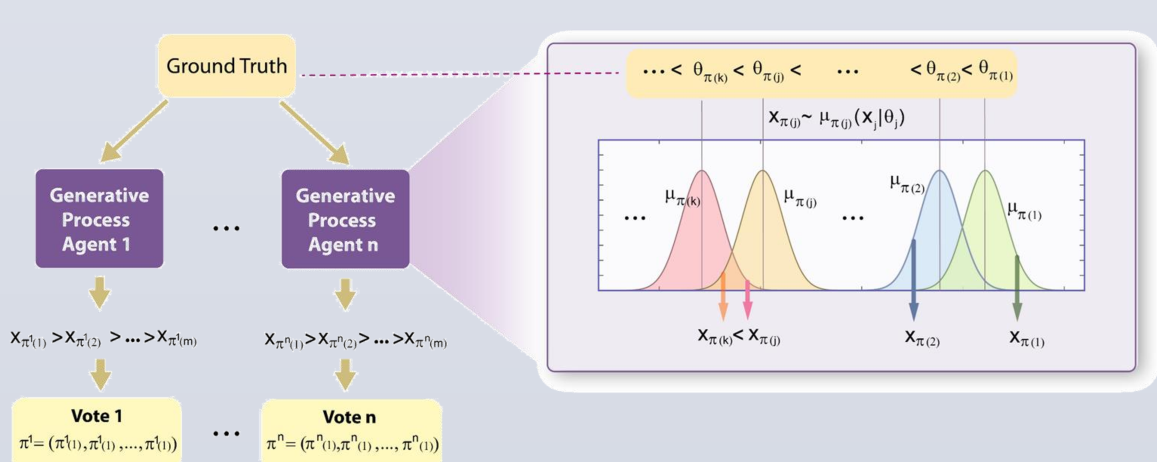
Ranking Models



Random Utility Models

Parameter space: $\Theta_{RUM} = \{\vec{\theta} | \theta_1, \ldots \theta_m\}$

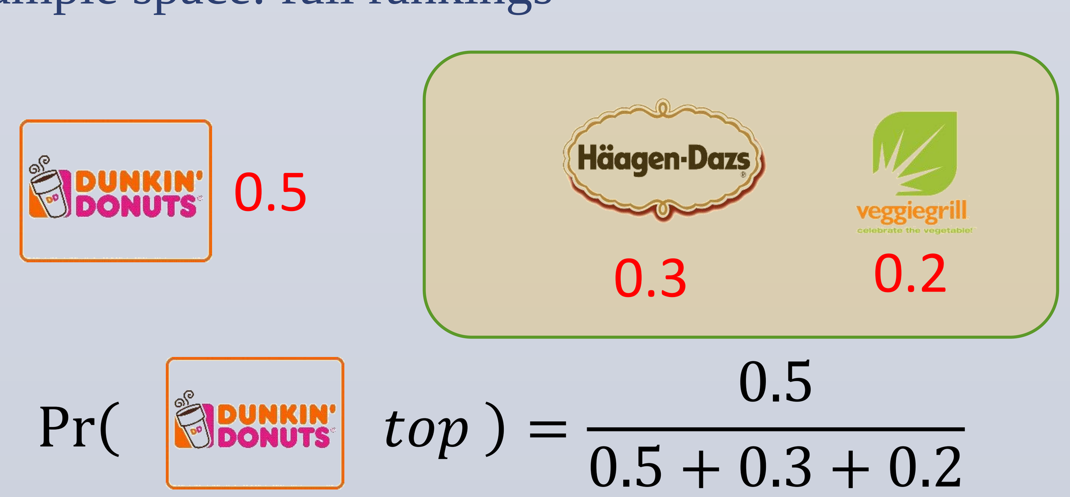
Sample space: full rankings



Plackett-Luce Model

Parameter space: $\Theta_{\mathrm{PL}} = \{\vec{\theta} | \theta_1, \ldots \theta_m\}$

Sample space: full rankings



Rank-Breaking

	Dunkin' Donuts	Haagen-Dazs	Veggiegrill
Dunkin' Donuts		2	2
Haagen-Dazs	1		2
Veggiegrill	1	1	

Composite Marginal Likelihood

$$\vec{\theta}^* = \arg\max \sum_{i \neq j} \log \Pr(a_i \succ a_j | \vec{\theta})^{\kappa_{ij} w_{ij}}$$

+ strictly concave
+ asymptotically normal
+ fast
+ easy to implement

- restricted to full rankings (future direction)

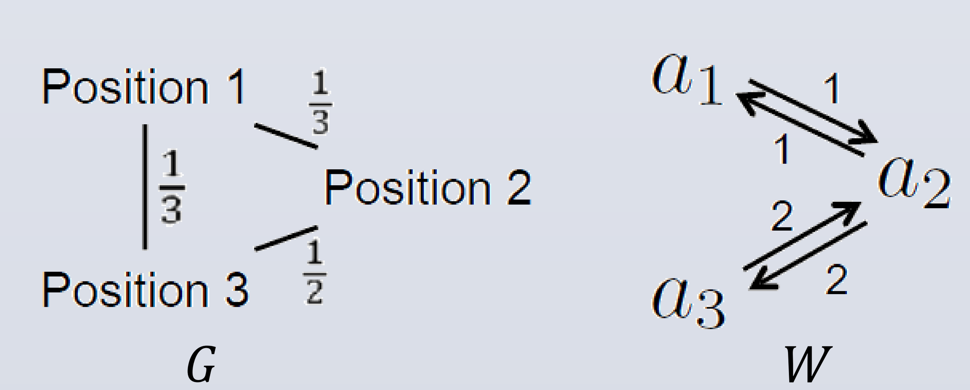
Theoretical Results

High-level message:

Breaking: symmetric RUMs: uniform breaking

Plackett-Luce: weighted union of position-k breakings

CML weights: connected and symmetric



Theorems 1 & 2: Strict log-concavity is preserved under convolution and marginalization.

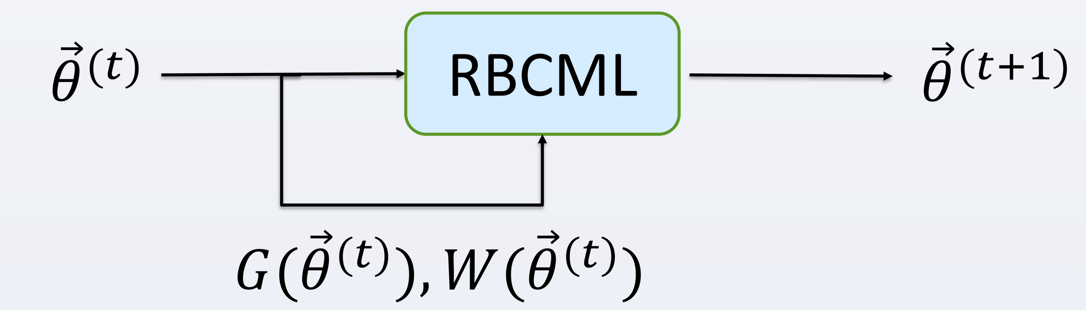
Theorems 3 & 4: (strongly connected $W \otimes G(P)$ is desired.) For Plackett-Luce model or RUMs where CDF of each utility distribution is strictly log-concave, the composite likelihood function (objective function of RBCML) is strictly log-concave if and only if $W \otimes G(P)$ is weakly connected. RBCML is bounded if and only if $W \otimes G(P)$ is strongly connected.

Theorem 5: RBCML is consistent and asymptotic normal.

Theorems 6 & 7: When CML weight is uniform, RBCML is consistent if and only if (i) for Plackett-Luce model, the breaking is weighted union of position-k breakings; (ii) for symmetric RUMs, the breaking is uniform.

Theorems 8 & 9: RBCML is consistent if and only if W is connected and symmetric and (i) for Plackett-Luce model, the breaking is weighted union of position-k breakings; (ii) for symmetric RUMs, the breaking is uniform.

Adaptive RBCML



Input: Profile P of n rankings, number of iterations T, the heuristic of breaking $G(\vec{\theta})$ and weights $W(\vec{\theta})$.

Output: estimated parameter $\vec{\theta}^*$

Initialize $\vec{\theta}^{(0)} = \vec{0}$

For $t = 1$ **to** T **do**

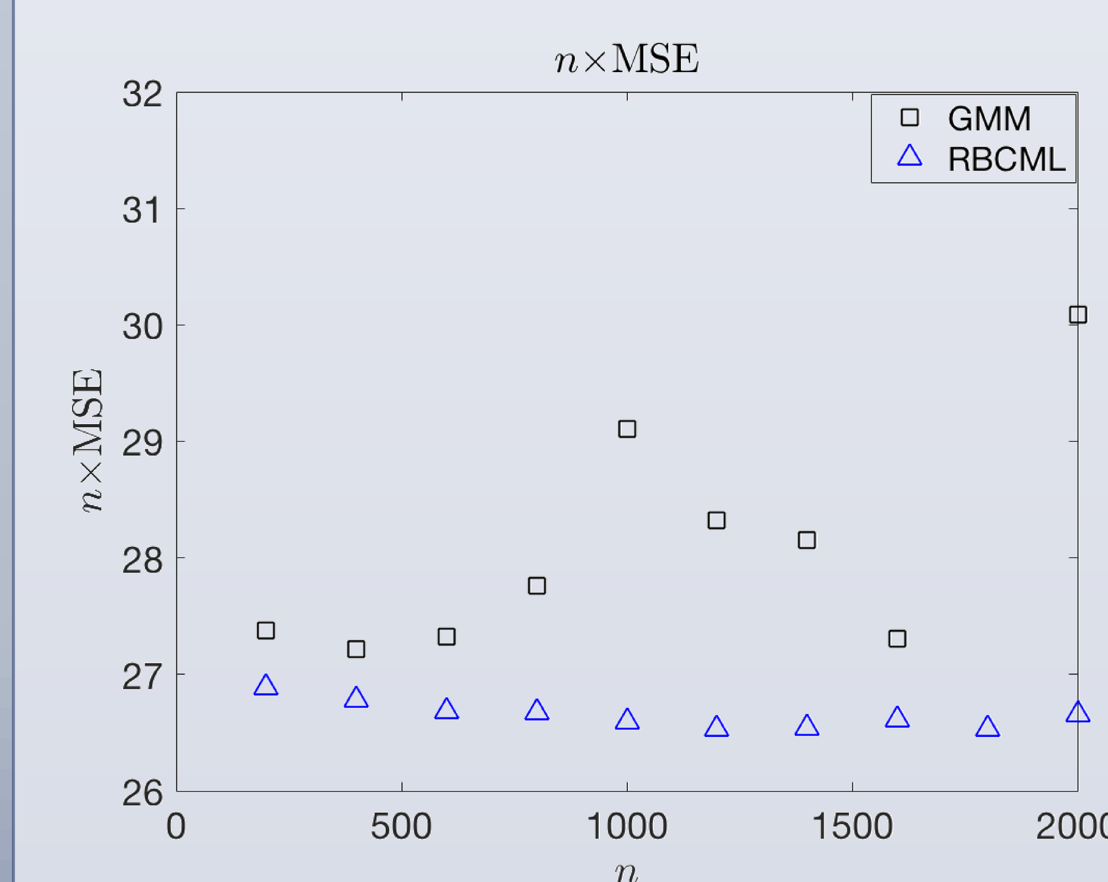
Compute $G(\vec{\theta}^{(t-1)})$ and $W(\vec{\theta}^{(t-1)})$

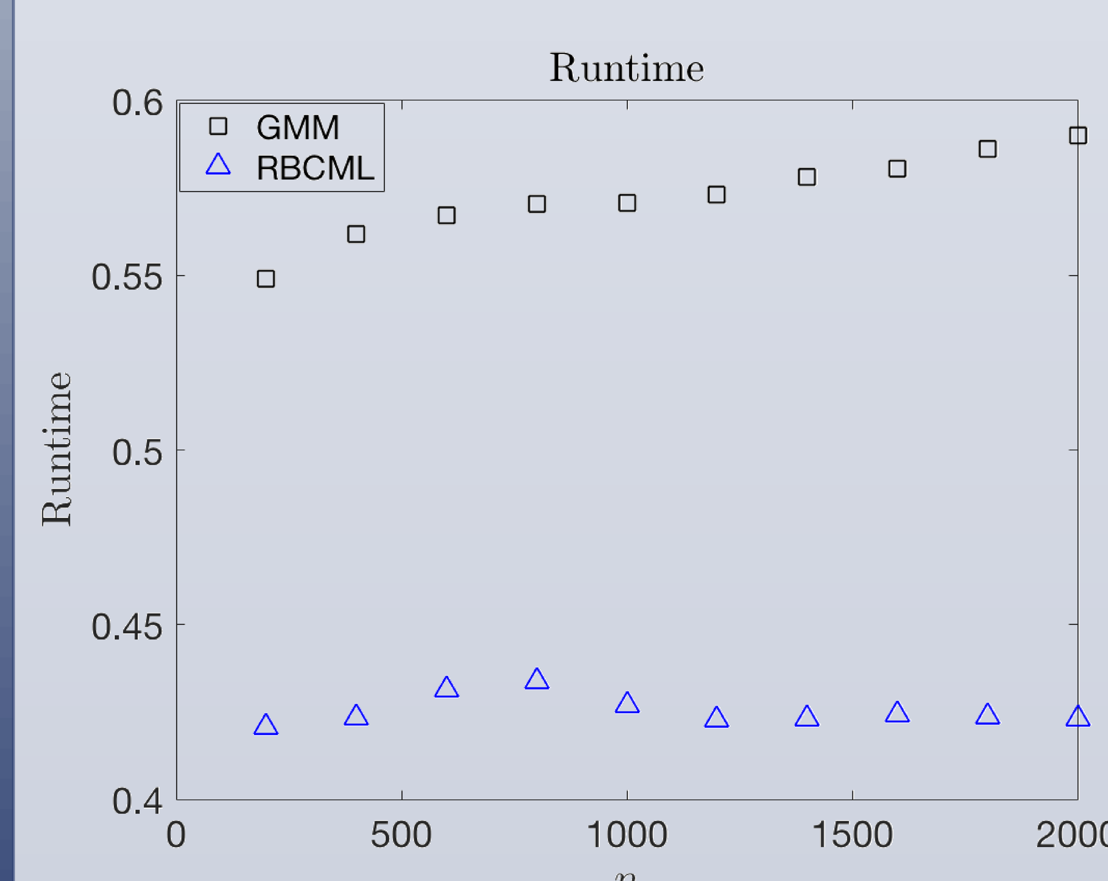
Estimate $\vec{\theta}^{(t)}$ using $G(\vec{\theta}^{(t-1)})$ and $W(\vec{\theta}^{(t-1)})$ using RBCML

End for

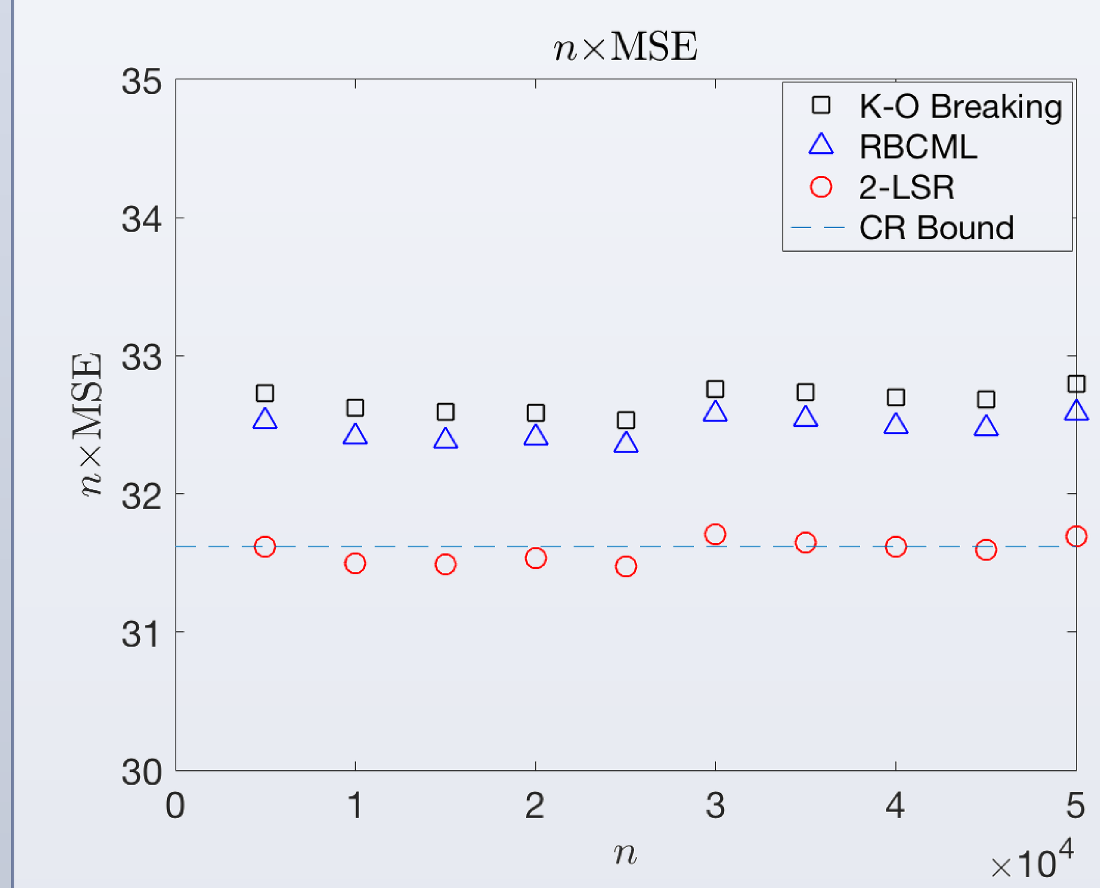
Experiments

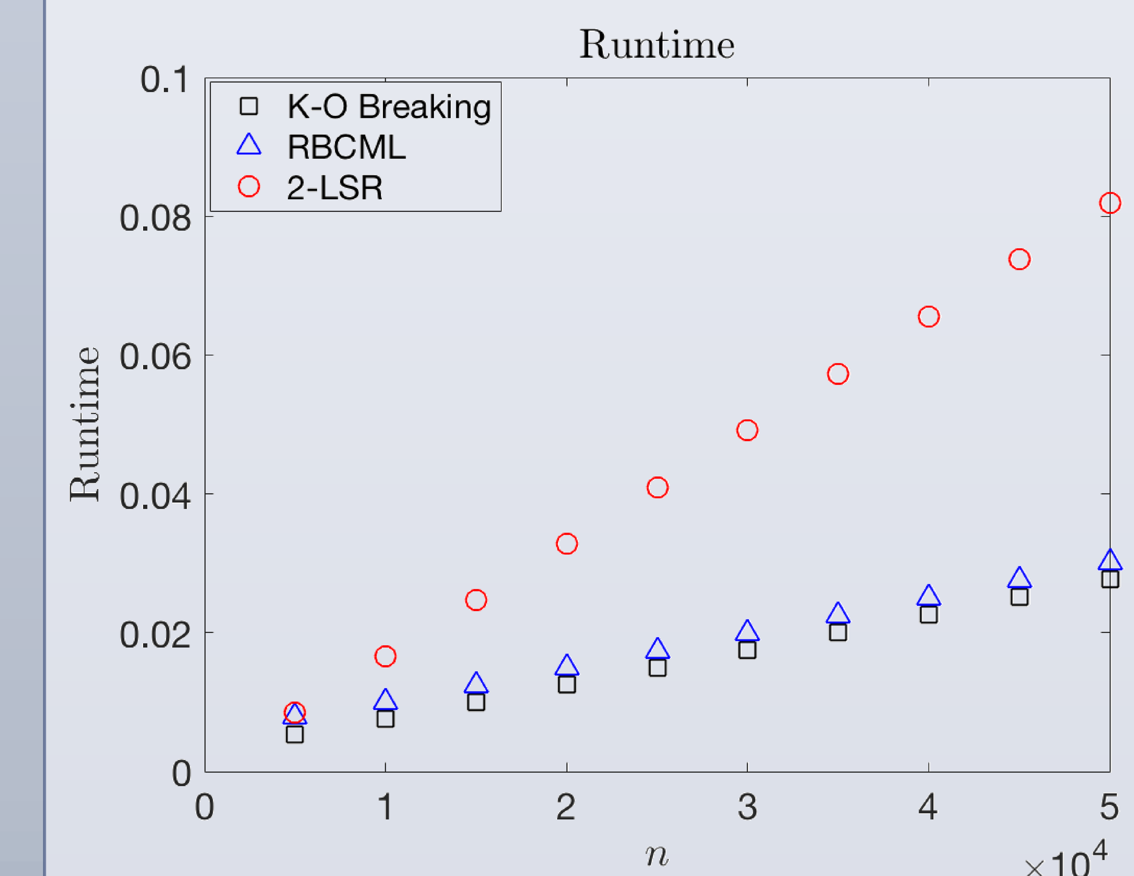
Gaussian Random Utility Model





Plackett-Luce Model





Summary and Future Work

RBCML: fast and accurate due to strict concavity and asymptotic normality.

Future Work: to extend RBCML to partial orders.

References

Hossein Azari Soufiani, David C. Parkes, and Lirong Xia, "Computing Parametric Ranking Models via Rank-Breaking", In proceedings of the 31st International Conference on Machine Learning, 2014.

Lucas Maystre and Matthias Grossglauser, "Fast and Accurate Inference of Plackett-Luce Models", in Advances in Neural Information Processing Systems, 2015.

Ashish Khetan and Sewoong Oh, "Data-Driven Rank Breaking for Efficient Rank Aggregation", in Journal of Machine Learning Research, 2016.

zhaozb08@gmail.com

http://homepages.rpi.edu/~zhaoz6/