Composite Marginal Likelihood Methods for Random Utility Models

Zhibing Zhao [1] **Lirong Xia** [1]

Abstract

We propose a novel and flexible *rank-breaking-then-composite-marginal-likelihood (RBCML)* framework for learning *random utility models (RUMs)*, which include the *Plackett-Luce* model. We characterize conditions for the objective function of RBCML to be strictly log-concave by proving that strict log-concavity is preserved under convolution and marginalization. We characterize necessary and sufficient conditions for RBCML to satisfy consistency and asymptotic normality. Experiments on synthetic data show that RBCML for Gaussian RUMs achieves better statistical efficiency and computational efficiency than the state-of-the-art algorithm and our RBCML for the Plackett-Luce model provides flexible tradeoffs between running time and statistical efficiency.

1. Introduction

How to model rank data and how to make optimal statistical inferences from rank data are important topics at the interface of statistics, computer science, and economics. *Random utility models (RUMs)* (Thurstone, 1927) are one of the most widely-applied statistical models for rank data. In an RUM, each alternative a_i is parameterized by a utility distribution μ_i. Agents' rankings are generated in two steps. In the first step, a *latent utility* u_i for each alternative a_i is generated from μ_i. In the second step, the alternatives are ranked w.r.t. their utilities u_i in descending order. The logit model and the probit model, which are very popular in statistics and economics, both have random utility interpretations.

While providing better fitness to the rank data (Azari Soufiani et al., 2012; Zhao et al., 2018b), general RUMs are computationally hard to tackle due to the lack of closed-form formulas for the likelihood function. The only known exception is the *Plackett-Luce model* (Plackett, 1975; Luce, 1959), which is the RUM with Gumbel distributions. RUMs, especially the Plackett-Luce model, have been widely applied to model and predict human behavior (McFadden, 2000), where the standard case of *discrete choice models* can be viewed as the Plackett-Luce model restricted to top choices. Other notable recent applications include elections (Gormley & Murphy, 2008), crowdsourcing (Pfeiffer et al., 2012), recommender systems (Wang et al., 2016), preference elicitation (Azari Soufiani et al., 2013b; Zhao et al., 2018a), marketing (Berry et al., 1995), health care (Bockstael, 1999), transportation (Bhat et al., 2007), and security (Yang et al., 2011).

Recently there has been a growing interest in designing faster and more accurate algorithms for RUMs. Many algorithms in previous work share the following *rank-breaking-then-optimization* architecture. First, rank data are converted to pairwise comparison data. Second, based on the pairwise comparisons, various optimization algorithms are designed to estimate the ground truth (Negahban et al., 2012; Azari Soufiani et al., 2013a; 2014; Chen & Suh, 2015; Khetan & Oh, 2016b;a).

Pairwise data are often obtained from rank data by applying *rank-breaking*, which allows for a smooth tradeoff between computational efficiency and statistical efficiency (Azari Soufiani et al., 2013a; 2014; Khetan & Oh, 2016b;a). Given m alternatives, a rank-breaking scheme is modeled by a weighted undirected graph $\mathcal{G}$ (see Figure 1 for an example) over $\{1,\dots,m\}$ (the vertices are positions in a ranking), such that for any ranking R over the m alternatives and any distinct $i_1, i_2 \le m$, we obtain $g_{i_1 i_2}$ (the weight on the edge $\{i_1, i_2\}$ in $\mathcal{G}$) pairwise comparisons between alternatives at positions i_1 and i_2 of R.

Our Contributions. By leveraging the celebrated *composite marginal likelihood (CML)* methods (Lindsay, 1988; Varin, 2008), we propose a novel and flexible *rank-breaking-then-CML* framework. Given an RUM, our framework, denoted by RBCML$(\mathcal{G},\mathcal{W})$, is defined by a weighted rank-breaking graph $\mathcal{G}$ and a CML-weight vector $\mathcal{W}=\{w_{i_1 i_2}: i_1, i_2 \le m, i_1 \ne i_2\}$, which contains one non-negative weight for each pair of alternatives (a_{i_1}, a_{i_2}). We note that both $\mathcal{G}$ and $\mathcal{W}$ are the algorithm designer's choices. Given rank data P, we compute $\vec{\theta}$ to maximize the following *com-*

[1]Computer Science Department, Rensselaer Polytechnic Institute, Troy, NY, USA. Correspondence to: Zhibing Zhao <zhaoz6@rpi.edu>, Lirong Xia <xial@cs.rpi.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

posite log-likelihood function.

$$\text{CLL}_{\mathcal{M}}(\vec{\theta}, P) = \sum_{i_1 \neq i_2} (\kappa_{i_1 i_2} w_{i_1 i_2} \ln p_{i_1 i_2}(\vec{\theta}))$$

Here $\vec{\theta}$ represents the parameters of RUM. Given $\mathcal{G}$, $\kappa_{i_1 i_2}$ is the percentage of pairwise comparisons $a_{i_1} \succ a_{i_2}$ in the data. $p_{i_1 i_2}(\vec{\theta})$ is the probability of $a_{i_1} \succ a_{i_2}$ under RUM with $\vec{\theta}$, which is the total probability of generating a ranking with $a_{i_1} \succ a_{i_2}$ given $\vec{\theta}$. We note that the RBCML framework is very general because any combination of $\mathcal{G}$ and $\mathcal{W}$ can be used. A breaking graph $\mathcal{G}$ is *uniform*, if all edges have the same weight. Let $\mathcal{G}_{\text{u}}$ denote the breaking graph whose weights are all 1. A CML-weight vector $\mathcal{W}$ is *symmetric*, if for all $i_1 \neq i_2$, we have $w_{i_1 i_2} = w_{i_2 i_1}$. $\mathcal{W}$ is *uniform*, if all weights are 1, denoted by $\mathcal{W}_{\text{u}}$.

Theoretical contributions. For convenience we let position-k breaking denote the breaking that consists of all unit-weight edges between position k and all positions after k. E.g. the position-1 breaking consists of all unit-weight pairwise comparisons in positions $\{(1,2),(1,3),\ldots,(1,m)\}$. A weighted union of position-k breakings is a breaking that has the same weight (possibly zero) for each k. An example is shown in Figure 1, which is the union of 1/3 position-1 breaking and 1/2 position-2 breaking. Our theoretical results carry the following message about "good" RBCMLs.

We should use RBCML$(\mathcal{G}, \mathcal{W})$ with connected and symmetric $\mathcal{W}$. For Plackett-Luce model, we should use a breaking $\mathcal{G}$ that is the weighted union of multiple position-k breakings. For RUMs with symmetric utility distributions, we should use $\mathcal{G}_u$.

The message is established via a series of theorems (Theorems 1, 2, 5, 8, and 9). Theorems 1 and 2, which prove that strict log-concavity is preserved under convolution and under marginalization, are of independent interest.

Algorithmic contributions. Experiments on synthetic data for Gaussian RUMs, where each utility distribution is Gaussian, show that RBCML$(\mathcal{G}_{\text{u}}, \mathcal{W}_{\text{u}})$ achieves better statistical efficiency and computational efficiency than the GMM algorithm by Azari Soufiani et al. (2014). For the Plackett-Luce model, we propose an RBCML with a heuristic $\mathcal{W}_H$. We compare our RBCML for the Plackett-Luce model with the consistent rank-breaking algorithm by Khetan & Oh (2016b) and the I-LSR algorithm by Maystre & Grossglauser (2015) via experiments on synthetic data and show that our RBCML provides a tradeoff between statistical efficiency and computational efficiency.

Related Work and Discussions. Our RBCML framework leverages the strengths of rank breaking and CML. The major advantage of CML is that often marginal likelihood functions are much easier to optimize than the full likelihood function. However, for RUMs, even computing the marginal likelihood may take too much time, as CML needs to count the number of pairwise comparisons between alternatives in the rankings, which takes $O(m^2 n)$ time, where m is the number of alternatives and n is the number of rankings. Therefore, standard CML becomes inefficient when m or n are large. RBCML overcomes such inefficiency by applying rank-breaking. The computational complexity of rank-breaking can be $O(kmn)$ for any $k \leq m$. Often a tradeoff between computational efficiency and statistical efficiency must be made.

RBCML generalizes the algorithm proposed by Khetan & Oh (2016b), which focused on the Plackett-Luce model and whose optimization technique turns out to be CML with $\mathcal{W}_{\text{u}}$.[1] The comparison between RBCML and other related work is summarized in Table 1.

Our theorems on *strict* log-concavity of composite likelihood function generalize Hunter (2004)'s result, which was proved for Plackett-Luce with $\mathcal{G}_{\text{u}}$ and $\mathcal{W}_{\text{u}}$. Our results can be applied to not only other $\mathcal{W}$'s under Plackett-Luce, but also other RUMs where the PDFs of utility distributions are strictly log-concave, e.g. Gaussians. Technically, proving our results for general RUMs is much more challenging due to the lack of closed-form formulas for the likelihood function. Another line of previous work proved (non-strict) log-concavity for special cases of RBCML (Azari Soufiani et al., 2012; Khetan & Oh, 2016a;b). Again, our theorems are stronger because (1) our theorems work for a more general class of RBCML, and (2) strict log-concavity is more desirable than log-concavity because the formal implies the uniqueness of the solution.

The key step in our proofs is the preservation of strict log-concavity under convolution (Theorem 1) and marginalization (Theorem 2). Surprisingly, we were not able to find these theorems in the literature, despite that it is well-known that (non-strict) log-concavity and strong log-concavity are preserved under convolution and marginalization (Saumard & Wellner, 2014). Our proofs of Theorems 1 and 2 are based on a careful examination of the condition for equality in the Prékopa-Leindler inequality proved by Dubuc (1977). We believe that Theorems 1 and 2 are of independent interest.

Xu & Reid (2011) provided sufficient conditions for general CML methods to satisfy consistency and asymptotic normality. Unfortunately, some of the conditions by Xu & Reid (2011) do not hold for RBCML. Therefore, we derive new proof of consistency and asymptotic normality for RBCML.

Khetan & Oh (2016b;a) provide sufficient conditions on rank-breakings for CML with $\mathcal{W}_{\text{u}}$ to be consistent under

[1] Khetan & Oh (2016b)'s algorithm works for special partial orders. In this paper, we only focus on comparisons between RBCML and their algorithms restricted to linear orders.

Algorithms	Breaking	Optimization	RUM
(Azari Soufiani et al., 2013a)	Uniform	GMM	Plackett-Luce
(Azari Soufiani et al., 2014)	Uniform	GMM	RUMs with sym. distributions
(Khetan & Oh, 2016b;a)	any	CML($\mathcal{W}_u$)	Plackett-Luce
RBCML	any	general CML	Plackett-Luce and RUMs with sym. distributions

Table 1. RBCML vs. previous work. GMM stands for Generalized Method of Moments.

the Plackett-Luce model. It is an open question what are all consistent rank-breakings for CML, even with $\mathcal{W}_u$. We answer this question for Plackett-Luce (Theorem 8), as well as a large class of other RUMs (Theorem 9), and for all $\mathcal{W}$'s.

2. Preliminaries

Let $\mathcal{A} = \{a_1, a_2, \cdots, a_m\}$ denote the set of m alternatives. Let $\mathcal{L}(\mathcal{A})$ denote the set of all linear orders (rankings) over $\mathcal{A}$. A ranking $R \in \mathcal{L}(\mathcal{A})$ is denoted by $a_{i_1} \succ a_{i_2} \succ \ldots \succ a_{i_m}$, where a_{i_1} is ranked at the top, a_{i_2} is ranked at the second position, etc. We write $a \succ_R b$ if a is ranked higher than b in R. Let $P = \{R_1, R_2, \ldots, R_n\}$ denote the collection of n rankings, called a *preference profile*.

Definition 1 (Random utility models (RUMs)) *A random utility model $\mathcal{M}$ over $\mathcal{A}$ associates each alternative a_i with a utility distribution $\mu_i(\cdot|\vec{\theta}_i)$. The parameter space is $\Theta = \{\vec{\theta} = \{\vec{\theta}_i | i = 1, 2, \ldots, m\}\}$. The sample space is $\mathcal{L}(\mathcal{A})^n$. Each ranking is generated i.i.d. in two steps. First, for each $i \le m$, a latent utility u_i is generated from $\mu_i(\cdot|\vec{\theta}_i)$ independently, and second, the alternatives are ranked according to their utilities in the descending order. Given a parameter $\vec{\theta}$, the probability of generating $R = a_{i_1} \succ a_{i_2} \succ \ldots \succ a_{i_m}$ is*

$$\Pr_{\mathcal{M}}(R|\vec{\theta}) = \int_{-\infty}^{\infty} \int_{u_{i_m}}^{\infty} \cdots \int_{u_{i_2}}^{\infty} \mu_{i_m}(u_{i_m}|\vec{\theta}_{i_m}) \cdots \mu_{i_1}(u_{i_1}|\vec{\theta}_{i_1}) du_{i_1} du_{i_2} \cdots du_{i_m}$$

In this paper, we focus on the *location family*, where the shapes of the utility distributions are fixed and each utility distribution μ_i is only parameterized by its mean, denoted by θ_i. Let π_i denote the distribution obtained from $\mu_i(\cdot|\theta_i)$ by shifting the mean to 0. For the location family, we have $\pi_i(u_i|\theta_i) = \pi(u_i - \theta_i)$. Because shifting the means of all alternatives by the same distance will not affect the distribution of the rankings, **w.l.o.g. we let $\theta_m = 0$ throughout the paper**. Moreover, we assume that the PDF of each utility distribution is continuous and positive everywhere. We further say that an RUM is *symmetric* if the PDF of each utility distribution is symmetric around its mean. We use Gaussian RUMs to denote the RUMs where all utility distributions are Gaussian.

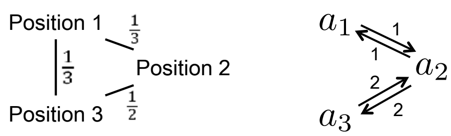
For any combination of m probability distributions $\pi_1, \ldots, \pi_m$ whose means are 0, we let RUM$(\pi_1, \ldots, \pi_m)$ denote the RUM location family where the shapes of utility distributions are $\pi_1, \ldots, \pi_m$. For any probability distribution π whose mean is 0, let RUM(π) denote the RUM where the shapes of all utility distributions are π.

Given a profile P and a parameter $\vec{\theta}$, we have $\Pr_{\mathcal{M}}(P|\vec{\theta}) = \prod_{j=1}^n \Pr_{\mathcal{M}}(R_j|\vec{\theta})$. Because all utilities are drawn independently, the probability of pairwise comparison is $\Pr_{\mathcal{M}}(a_{i_1} \succ a_{i_2}|\vec{\theta}) = \int_{-\infty}^{\infty} \int_{u_{i_2}}^{\infty} \mu_{i_1}(u_{i_1}|\vec{\theta}) \mu_{i_2}(u_{i_2}|\vec{\theta}) du_{i_1} du_{i_2}$.

Example 1 (Plackett-Luce model as an RUM) *Let $\mu_i(\cdot|\theta_i)$ be the Gumbel distribution where $\mu_i(x_i|\theta_i) = e^{-(x_i-\theta_i)-e^{-(x_i-\theta_i)}}$. For any ranking $R = a_{i_1} \succ a_{i_2} \succ \ldots \succ a_{i_m}$, we have $\Pr_{PL}(R|\vec{\theta}) = \prod_{t=1}^{m-1} \frac{e^{\theta_{i_t}}}{\sum_{l=t}^m e^{\theta_{i_l}}}$. The probability of $a_{i_1} \succ a_{i_2}$ under the Plackett-Luce model is $\Pr_{PL}(a_{i_1} \succ a_{i_2}|\vec{\theta}) = \frac{e^{\theta_{i_1}}}{e^{\theta_{i_1}} + e^{\theta_{i_2}}}$.*

A *weighted (rank-)breaking* $\mathcal{G} = \{g_{ii'} : i < i' \le m\}$ can be represented by a weighted undirected graph over positions $\{1, \ldots, m\}$, such that for any $g_{ii'} > 0$, there is an edge between i and i' whose weight is $g_{ii'}$. We say that $\mathcal{G}$ is *uniform*, if all weights are the same. Let $\mathcal{G}_u$ denote the the uniform breaking where all weights are 1. For any $1 \le k \le m-1$, the position-k breaking is the graph where for any $l > k$, there is an edge with weight 1 between k and l. For any $\vec{\theta} \in \mathbb{R}^{m-1}$, any weighted rank-breaking $\mathcal{G}$, any pair of alternatives a_{i_1}, a_{i_2}, let $\mathcal{G}_{a_{i_1} \succ a_{i_2}}(R) = g_{ii'}$ such that a_{i_1} and a_{i_2} are ranked at the ith position and the i'th position in R, respectively. Given a profile P, we define $\kappa_{i_1 i_2} = \frac{\sum_{j=1}^n \mathcal{G}_{a_{i_1} \succ a_{i_2}}(R_j)}{n}$, and let $\bar{\kappa}_{i_1 i_2} = E[\kappa_{i_1 i_2}|\vec{\theta}]$. We note that $\kappa_{i_1 i_2}$ is a function of the preference profile. $\bar{\kappa}_{i_1 i_2}$ is the expected $\kappa_{i_1 i_2}$ value for perfect data given $\vec{\theta}$, which means that it is a function of the ground truth parameter $\vec{\theta}$.

Example 2 *Let $m = 3, n = 2$. The profile $P = \{a_1 \succ a_2 \succ a_3, a_3 \succ a_2 \succ a_1\}$. Let $\mathcal{G} = \{g_{12} = g_{13} = \frac{1}{3}, g_{23} = \frac{1}{2}\}$ as shown in Figure 1 (a). Then we have $\kappa_{12} = \kappa_{13} = \frac{1}{3}/n = \frac{1}{6}$, $\kappa_{23} = \frac{1}{2}/n = \frac{1}{4}$, $\kappa_{32} = \kappa_{31} = \frac{1}{3}/n = \frac{1}{6}$, $\kappa_{21} = \frac{1}{2}/n = \frac{1}{4}$.*



(a) $\mathcal{G}$. (b) $\mathcal{W}$.

Figure 1. A rank-breaking $\mathcal{G}$ and a CML-weight vector $\mathcal{W}$.

3. Composite Marginal Likelihood Methods

Let $\mathcal{W} = \{w_{ii'} : a_i, a_{i'} \in \mathcal{A}\}$ denote a CML-weight vector. We say that $\mathcal{W}$ is *symmetric*, if for any pair of alternatives $a_i, a_{i'}$, we have $w_{ii'} = w_{i'i} > 0$. We say that $\mathcal{W}$ is *uniform*, if all $w_{ii'}$'s are equal. Let $\mathcal{W}_{\text{u}}$ denote a uniform $\mathcal{W}$.

We note that vertices in $\mathcal{W}$ corresponds to the alternatives while vertices in $\mathcal{G}$ corresponds to positions in a ranking. For example, vertex i in $\mathcal{W}$ corresponds to a_i, while vertex i in $\mathcal{G}$ corresponds to the ith position in a ranking.

Example 3 *A symmetric $\mathcal{W}$ is shown in Figure 1 (b), where $w_{12} = w_{21} = 1$ and $w_{23} = w_{32} = 2$.*

Given $\mathcal{G}$ and $\mathcal{W}$, we propose the *rank-breaking-then-CML* framework for RUMs, denoted by RBCML$(\mathcal{G}, \mathcal{W})$, to be the maximizer of composite log-marginal likelihood, which is defined below.

Definition 2 (Composite marginal likelihood for RUMs) *Given an RUM $\mathcal{M}$, for any preference profile P and any θ, let $p_{i_1 i_2}(\vec{\theta}) = \Pr_{\mathcal{M}}(a_{i_1} \succ a_{i_2} | \vec{\theta})$. The composite marginal likelihood is $CL_{\mathcal{M}}(\vec{\theta}, P) = \prod_{i_1 \neq i_2} (p_{i_1 i_2}(\vec{\theta}))^{\kappa_{i_1 i_2} w_{i_1 i_2}}$. The composite log-marginal likelihood becomes:*

$$CLL_{\mathcal{M}}(\vec{\theta}, P) = \sum_{i_1 \neq i_2} \kappa_{i_1 i_2} w_{i_1 i_2} \ln p_{i_1 i_2}(\vec{\theta}) \qquad (1)$$

We let RBCML$(\mathcal{G}, \mathcal{W})(P) = \arg\max_{\vec{\theta}} \text{CLL}_{\mathcal{M}}(\vec{\theta}, P)$. For the Plackett-Luce model the composite (log-)marginal likelihood has a closed-form formula.

Definition 3 (CML for Plackett-Luce) *For any $\vec{\theta}$ and preference profile P, the composite marginal likelihood for the Plackett-Luce model is $CL_{PL}(\vec{\theta}, P) = \prod_{i_1 < i_2} (\frac{e^{\theta_{i_1}}}{e^{\theta_{i_1}} + e^{\theta_{i_2}}})^{\kappa_{i_1 i_2} w_{i_1 i_2}} (\frac{e^{\theta_{i_2}}}{e^{\theta_{i_1}} + e^{\theta_{i_2}}})^{\kappa_{i_2 i_1} w_{i_2 i_1}}$. The composite log-marginal likelihood is*

$$CLL_{PL}(\vec{\theta}, P) = \sum_{i_1 < i_2} (\kappa_{i_1 i_2} w_{i_1 i_2} \theta_{i_1} + \kappa_{i_2 i_1} w_{i_2 i_1} \theta_{i_2} - (\kappa_{i_1 i_2} w_{i_1 i_2} + \kappa_{i_2 i_1} w_{i_2 i_1}) \ln(e^{\theta_{i_1}} + e^{\theta_{i_2}})) \qquad (2)$$

The first order conditions are, for all i, $\frac{\partial \text{CLL}_{\text{PL}}(\vec{\theta}, P)}{\partial \theta_i} = \sum_{i' \neq i} (\kappa_{ii'} w_{ii'} - (\kappa_{ii'} w_{ii'} + \kappa_{i'i} w_{i'i}) \frac{e^{\theta_i}}{e^{\theta_i} + e^{\theta_{i'}}})$.

Example 4 *Continuing Example 2 and Example 3,*

$$CLL_{PL}(\vec{\theta}, P) = \frac{1}{6}\theta_1 + \frac{1}{4}\theta_2 - (\frac{1}{6} + \frac{1}{4}) \ln(e^{\theta_1} + e^{\theta_2}) + \frac{1}{2}\theta_2 - (\frac{1}{2} + \frac{1}{3}) \ln(e^{\theta_2} + 1)$$

By solving the first order conditions, we have $e^{\theta_1} = 1$ and $e^{\theta_2} = 1.5$. So the outcome of RBCML is $\theta_1 = 0$, $\theta_2 = \ln 1.5$. We recall that $\theta_3 = 0$ in this paper.

4. Preservation of Strict Log-Concavity

Definition 4 (Log-concavity and strict log-concavity) *A function $f(\vec{x}) > 0$ is* log-concave *if $\forall 0 < \lambda < 1$, we have $f(\lambda \vec{x} + (1 - \lambda)\vec{y}) \geq f(\vec{x})^{\lambda} f(\vec{y})^{1-\lambda}$. If the inequality is always strict, then f is* strictly log-concave.

Theorem 1 (Preservation under convolution) *Let $f(x)$ and $g(x)$ be two continuous and strictly log-concave functions on $\mathbb{R}$. Then $f * g$ is also strictly log-concave.*

Proof: The proof is done by examining the equality condition for the Prékopa-Leindler inequality. Let $h = f * g$, namely, for any $y \in \mathbb{R}$, $h(y) = \int_{\mathbb{R}} f(y - x) g(x) dx$. Because f and g are continuous, so does h. To prove the strict log-concavity of h, it suffices to prove that for any different $y_1, y_2 \in \mathbb{R}$, $h(\frac{y_1 + y_2}{2}) > \sqrt{h(y_1) h(y_2)}$.

Suppose for the sake of contradiction that this is not true. Since log-concavity preserves under convolution (Saumard & Wellner, 2014), h is log-concave. So, there exist $y_1 < y_2$ such that $h(\frac{y_1 + y_2}{2}) = \sqrt{h(y_1) h(y_2)}$. Let $\Lambda(x, y) = f(y - x) g(x)$. We further define

$$H(x) = \Lambda(x, \frac{y_1 + y_2}{2}) = f(\frac{y_1 + y_2}{2} - x) g(x)$$
$$F(x) = \Lambda(x, y_1) = f(y_1 - x) g(x)$$
$$G(x) = \Lambda(x, y_2) = f(y_2 - x) g(x)$$

Because (non-strict) log-concavity is preserved under convolution, $\Lambda(x, y)$ is log-concave. We have that for any $x \in \mathbb{R}$, $H(x) \geq \sqrt{F(x) G(x)}$. The Prékopa-Leindler inequality asserts that

$$\int_{\mathbb{R}} H(x) dx \geq \sqrt{\int_{\mathbb{R}} F(x) dx \int_{\mathbb{R}} G(x) dx} \qquad (3)$$

Because $h(\frac{y_1 + y_2}{2}) = \int_{\mathbb{R}} H(x) dx$, $h(y_1) = \int_{\mathbb{R}} F(x) dx$, $h(y_2) = \int_{\mathbb{R}} G(x) dx$, and $h(\frac{y_1 + y_2}{2}) = \sqrt{h(y_1) h(y_2)}$, (3) becomes an equation. It was proved by Dubuc (1977) that: there exist $a > 0$ and $b \in \mathbb{R}$ such that the following conditions hold almost everywhere for $x \in \mathbb{R}$ (see the translation of Dubuc's result in English by Ball & Böröczky (2010)). 1. $F(x) = a H(x + b)$, 2. $G(x) = a^{-1} H(x - b)$.

The first condition means that for almost every $x \in \mathbb{R}$,

$$f(y_1 - x)g(x) = af(\frac{y_1 + y_2}{2} - x - b)g(x + b)$$
$$\iff \frac{g(x)}{g(x+b)} = a\frac{f(\frac{y_1+y_2}{2} - x - b)}{f(y_1 - x)} \quad (4)$$

The second condition means that for almost all $x \in \mathbb{R}$, $f(y_2 - x)g(x) = a^{-1}f(\frac{y_1+y_2}{2} - x + b)g(x - b) \iff \frac{g(x-b)}{g(x)} = a\frac{f(y_2-x)}{f(\frac{y_1+y_2}{2}-x+b)}$. Therefore, for almost all $x \in \mathbb{R}$,

$$\frac{g(x)}{g(x+b)} = a\frac{f(y_2 - x - b)}{f(\frac{y_1+y_2}{2} - x)} \quad (5)$$

Combining (4) and (5), for almost every $x \in \mathbb{R}$ we have

$$\frac{g(x)}{g(x+b)} = a\frac{f(y_2 - x - b)}{f(\frac{y_1+y_2}{2} - x)} = a\frac{f(\frac{y_1+y_2}{2} - x - b)}{f(y_1 - x)} \quad (6)$$

Because $f(x)$ is strictly log-concave, for any fixed $c \neq 0$, $\frac{f(x+c)}{f(x)}$ is strictly monotonic. Because $y_1 \neq y_2$ and $y_2 - x - b - (\frac{y_1+y_2}{2} - x) = \frac{y_1+y_2}{2} - x - b - (y_1 - x) = \frac{y_2-y_1}{2} - b$, we must have that $\frac{y_2-y_1}{2} - b = 0$, namely $b = \frac{y_2-y_1}{2}$. Therefore, (6) becomes $\frac{g(x)}{g(x+\frac{y_2-y_1}{2})} = a$ for almost every $x \in \mathbb{R}$, which contradicts the strict log-concavity of g. This means that $h = f * g$ is strictly log-concave. ■

Theorem 2 (Preservation under marginalization) *Let $h(x, y)$ be a strictly log-concave function on $\mathbb{R}^2$. Then $\int_{\mathbb{R}} h(x, y)dx$ is strictly log-concave on $\mathbb{R}$.*

Again, the proof is done by examining the equality condition for the Prékopa-Leindler inequality. All missing proofs can be found in the supplementary material.

5. Strict Log-Concavity of CML

For any profile P, let $G(P)$ denote the weighted directed graph where each represents an alternative. For any $1 \leq i \neq i' \leq m$, the weight on the edge from i to i' is $\kappa_{ii'}$. A weighted directed graph is *(weakly) connected*, if after removing the directions on all edges, the resulting undirected graph is connected. A weighted directed graph is *strongly connected*, if there is a directed path with positive weights between any pair of vertices. Given any pair of weighted graphs G_1 and G_2, we let $G_1 \otimes G_2$ denote the weighted graph where the weights on each edge is the multiplication of the weights of same edge in G_1 and G_2.

Theorem 3 *Given any profile P, the composite likelihood function for Plackett-Luce, i.e. $CL_{PL}(\vec{\theta}, P)$, is strictly log-concave if and only if $\mathcal{W} \otimes G(P)$ is weakly connected. $\arg\max_{\vec{\theta}} CL_{PL}(\vec{\theta}, P)$ is bounded if and only if $\mathcal{W} \otimes G(P)$ is strongly connected.*

The proof is similar to the log-concavity of likelihood for BTL by (Hunter, 2004). For general RUMs we prove a similar theorem.

Theorem 4 *Let $\mathcal{M}$ be an RUM where the CDF of each utility distribution is strictly log-concave. Given any profile P, the composite likelihood function for $\mathcal{M}$, i.e. $CL_{\mathcal{M}}(\vec{\theta}, P)$, is strictly log-concave if and only if $\mathcal{W} \otimes G(P)$ is weakly connected. $\arg\max_{\vec{\theta}} CL_{\mathcal{M}}(\vec{\theta}, P)$ is bounded if and only if $\mathcal{W} \otimes G(P)$ is strongly connected.*

Proof sketch: It is not hard to check that when $\mathcal{W} \otimes G(P)$ is not connected, there exist $\vec{\theta}^{(1)}$ and $\vec{\theta}^{(2)}$ such that for any $0 < \lambda < 1$ we have $\text{CLL}_{\text{PL}}(\vec{\theta}^{(1)}, P) = \text{CLL}_{\text{PL}}(\vec{\theta}^{(2)}, P) = \lambda\text{CLL}_{\text{PL}}(\vec{\theta}^{(1)}, P) + (1-\lambda)\text{CLL}_{\text{PL}}(\vec{\theta}^{(2)}, P)$, which violates strict log-concavity. Suppose $\mathcal{W} \otimes G(P)$ is weakly connected, it suffices to prove for any $i_1 \neq i_2$, $\Pr(a_{i_1} \succ a_{i_2}|\vec{\theta})$ is strictly log-concave. We can write this as an integral over $u_{i_2} - u_{i_1}$: $\Pr(u_{i_1} > u_{i_2}|\vec{\theta}) = \int_0^\infty \Pr(u_{i_2} - u_{i_1} = s|\vec{\theta})ds$.

Let $\pi^*_{i_2}(\cdot|\vec{\theta})$ denote the flipped distribution of $\pi_{i_2}(\cdot|\vec{\theta})$ around $x = s$, then we have $\pi^*_{i_2}(s - x|\vec{\theta}) = \pi_{i_2}(s + x|\vec{\theta})$. Further we have $\Pr(u_{i_1} > u_{i_2}|\vec{\theta}) = \int_0^\infty \int_{-\infty}^\infty \pi_{i_1}(x|\theta_{i_1})\pi_{i_2}(x + s|\theta_{i_2})dxds = \int_0^\infty \pi_{i_1} * \pi^*_{i_2} ds$. By Theorem 1, $\pi_{i_1} * \pi^*_{i_2}$ is strictly log-concave. Then we prove that tail probability of a strictly log-concave distribution is also strictly log-concave.

The proof for boundedness is similar to the proof of a similar condition for BTL by Hunter (2004). ■

6. Asymptotic Properties of RBCML

Given any RUM $\mathcal{M}$ and any parameter $\vec{\theta}$, we define $\text{ELL}_{\mathcal{M}}(\vec{\theta}) = E[\text{CLL}_{\mathcal{M}}(\vec{\theta}, R)]$ and let $\nabla\text{ELL}_{\mathcal{M}}(\vec{\theta})$ be the gradient of $\text{ELL}_{\mathcal{M}}(\vec{\theta})$, whose ith element is $\nabla_i\text{ELL}_{\mathcal{M}}(\vec{\theta}) = \sum_{i' \neq i}(\frac{\bar{\kappa}_{ii'}w_{ii'}}{p_{ii'}(\vec{\theta})}\frac{\partial p_{ii'}(\vec{\theta})}{\partial \theta_i} + \frac{\bar{\kappa}_{i'i}w_{i'i}}{p_{i'i}(\vec{\theta})}\frac{\partial p_{i'i}(\vec{\theta})}{\partial \theta_i})$. Let $H(\vec{\theta}, P)$ be the Hessian matrix evaluated at $\vec{\theta}$. And let $H_0(\vec{\theta}_0)$ denote the expected Hessian of $\text{CLL}_{\mathcal{M}}(\vec{\theta}, P)$ at $\vec{\theta}_0$, where $\vec{\theta}_0$ is the ground truth parameter.

Theorem 5 (Consistency and asymptotic normality) *Given any RUM $\mathcal{M}$, any $\vec{\theta}_0$ and any profile P with n rankings. Let $\vec{\theta}^*$ be the output of RBCML$(\mathcal{G}, \mathcal{W})$. When $n \to \infty$, we have $\vec{\theta}^* \xrightarrow{p} \vec{\theta}_0$ and*

$$\sqrt{n}(\vec{\theta}^* - \vec{\theta}_0) \xrightarrow{d} N(0, H_0^{-1}(\vec{\theta}_0)Var[\nabla CLL_{\mathcal{M}}(\vec{\theta}_0, R)]H_0^{-1}(\vec{\theta}_0))$$

if and only if $\vec{\theta}_0$ is the only solution to

$$\nabla ELL_{\mathcal{M}}(\vec{\theta}) = \vec{0}, \quad (7)$$

Proof: The "only if" direction is straightforward. The solution to (7) is unique because $\text{CLL}_{\mathcal{M}}(\vec{\theta}, P)$ is strictly concave. Suppose $\vec{\theta}_1$, other than $\vec{\theta}_0$, is the solution to (7), then when $n \to \infty$, $\vec{\theta}_1$ will be the estimate of RBCML$(\mathcal{G}, \mathcal{W})$, which means RBCML$(\mathcal{G}, \mathcal{W})$ is not consistent.

Now we prove the "if" direction. First we prove consistency. It is required by Xu & Reid (2011) that for different parameters, the probabilities for any composite likelihood event are different, which is not true in our case. A simple counterexample is $\theta_1^{(1)} = 1, \theta_1^{(2)} = 2, \theta_2^{(1)} = \theta_3^{(1)} = \theta_2^{(2)} = \theta_3^{(2)} = 0$. Then $\Pr(a_2 \succ a_3 | \vec{\theta}^{(1)}) = \Pr(a_2 \succ a_3 | \vec{\theta}^{(2)})$.

By the law of large numbers, we have for any ϵ, $\Pr(|\text{CLL}_{\mathcal{M}}(\vec{\theta}, P) - \text{ELL}_{\mathcal{M}}(\vec{\theta})| \leq \epsilon/2) \to 1$ as $n \to \infty$. This implies $\lim_{n\to\infty} \Pr(\text{CLL}_{\mathcal{M}}(\vec{\theta}^*, P) \leq \text{ELL}_{\mathcal{M}}(\vec{\theta}^*) + \epsilon/2) = 1$. Similarly we have $\lim_{n\to\infty} \Pr(\text{ELL}_{\mathcal{M}}(\vec{\theta}_0) \leq \text{CLL}_{\mathcal{M}}(\vec{\theta}_0, P) + \epsilon/2) = 1$. Since $\vec{\theta}^*$ maximize $\text{CLL}_{\mathcal{M}}(\vec{\theta}, P)$, we have $\Pr(\text{CLL}_{\mathcal{M}}(\vec{\theta}_0, P) \leq \text{CLL}_{\mathcal{M}}(\vec{\theta}^*, P)) = 1$. The above three equations imply that $\lim_{n\to\infty} \Pr(\text{ELL}_{\mathcal{M}}(\vec{\theta}_0) - \text{ELL}_{\mathcal{M}}(\vec{\theta}^*) \leq \epsilon) = 1$.

Let Θ_ϵ be the subset of parameter space s.t. $\forall \vec{\theta} \in \Theta_\epsilon$, $\text{ELL}_{\mathcal{M}}(\vec{\theta}_0) - \text{ELL}_{\mathcal{M}}(\vec{\theta}) \leq \epsilon$. Because $\text{ELL}_{\mathcal{M}}(\vec{\theta})$ is strictly concave, Θ_ϵ is compact and has a unique maximum at $\vec{\theta}_0$. Thus for any $\epsilon > 0$, $\lim_{n\to\infty} \Pr(\vec{\theta}^* \in \Theta_\epsilon) = 1$. This implies consistency, i.e., $\vec{\theta}^* \xrightarrow{p} \vec{\theta}_0$.

Now we prove asymptotic normality. By mean value theorem, we have $0 = \nabla\text{CLL}_{\mathcal{M}}(\vec{\theta}^*, P) = \nabla\text{CLL}_{\mathcal{M}}(\vec{\theta}_0, P) + H(\alpha\vec{\theta}^* + (1-\alpha)\vec{\theta}_0, P)(\vec{\theta}^* - \vec{\theta}_0)$, where $0 \leq \alpha \leq 1$. Therefore, we have $\sqrt{n}(\vec{\theta}^* - \vec{\theta}) = -H^{-1}(\alpha\vec{\theta}^* + (1-\alpha)\vec{\theta}_0, P)(\sqrt{n}\nabla\text{CLL}_{\mathcal{M}}(\vec{\theta}_0, P))$. Since $\nabla\text{CLL}_{\mathcal{M}}(\vec{\theta}_0, P) = \frac{1}{n}\sum_{j=1}^n \nabla\text{CLL}_{\mathcal{M}}(\vec{\theta}_0, R_j)$, by the central limit theorem, we have

$$\sqrt{n}\nabla\text{CLL}_{\mathcal{M}}(\vec{\theta}_0, P) \xrightarrow{d} N(0, \text{Var}[\nabla\text{CLL}_{\mathcal{M}}(\vec{\theta}_0, R)])$$

Because $\vec{\theta}^* \xrightarrow{p} \vec{\theta}_0$ and H is continuous, we have $H(\alpha\vec{\theta}^* + (1-\alpha)\vec{\theta}_0, P) \xrightarrow{p} H(\vec{\theta}_0, P)$. Since $H(\vec{\theta}, P) = \frac{1}{n}\sum_{j=1}^n H(\vec{\theta}, R_j)$, by law of large numbers, we have $H(\vec{\theta}, P) \xrightarrow{p} H_0(\vec{\theta}_0)$. Therefore, we have

$$\sqrt{n}(\vec{\theta}^* - \vec{\theta}) = -H_0^{-1}(\vec{\theta}_0)(\sqrt{n}\nabla\text{CLL}_{\mathcal{M}}(\vec{\theta}_0, P)),$$

which implies that $\text{Var}[\sqrt{n}(\vec{\theta}^* - \vec{\theta})] = H_0^{-1}(\vec{\theta}_0)\text{Var}[\nabla\text{CLL}_{\mathcal{M}}(\vec{\theta}_0, R)]H_0^{-1}(\vec{\theta}_0)$. ■

7. Consistency of RBCML

Formal proofs of theorems in this section depends on a series of lemmas, which can be found in the appendix. The full proofs can also be found in the appendix.

Theorem 6 *RBCML$(\mathcal{G}, \mathcal{W}_u)$ is consistent for Plackett-Luce if and only if the breaking is weighted union of position-k breakings.*

Proof sketch: The "if" direction is proved in (Khetan & Oh, 2016b). We only prove the "only if" direction by induction on m. When $m = 2$, the only breaking is the comparison between the two alternatives. The conclusion holds.

Suppose it holds for $m = l$, then when $m = l + 1$, we first prove a lemma which says that by restricting $\mathcal{G}$ to any set of continuous positions, the theorem must hold for the subgraph. Then, we focus on $\mathcal{G}_{[2,m]}$, which is the subgraph of $\mathcal{G}$ on {2,...,m}. $\mathcal{G}_{[2,m]}$ must be a weighted union of position-k breakings. Then we focus on $\mathcal{G}_{[1,m-1]}$. The only remaining case is to prove that the weight on edge $\{1, m\}$ is the same as the weight on edges $\{1, i\}$ for all $i \leq m - 1$.

Suppose for the sake of contradiction this is not true, then we can subtract a weighted union of position-k breakings from the graph, so that the remaining graph has a single edge $\{1, m\}$. We then prove that such an single-edge breaking is inconsistent by proving that (7) is not satisfied, which leads to a contradiction. ■

Theorem 7 *Let $\pi_1, \pi_2, \dots, \pi_m$ denote the utility distributions for a symmetric RUM. Suppose there exists π_i s.t. (1) $(\ln \pi_i(x))'$ is monotonically decreasing, and (2) $\lim_{x\to-\infty}(\ln \pi_i(x))' \to \infty$. Then, RBCML$(\mathcal{G}, \mathcal{W}_u)$ is consistent if and only if $\mathcal{G}$ is uniform.*

Proof sketch: Define the single-edge breaking $\mathcal{G}_1 = \{g_{1m} = 1\}$. We first prove RBCML$(\mathcal{G}_1, \mathcal{W}_{\text{u}})$ is not consistent. Then we prove the theorem by induction on m. $m = 2$ is trivial because the only breaking is uniform. For $m = 3$, we first prove that the single-edge breaking $\mathcal{G}_1 = \{g_{13} = 1\}$ is not consistent. Suppose the breaking is $\mathcal{G} = \{g_{12} = x, g_{23} = y, g_{13} = z\}$. Let $\mathcal{G}^* = \{g_{12} = y, g_{23} = x, g_{13} = z\}$. We prove that RBCML$(\mathcal{G}^*, \mathcal{W}_{\text{u}})$ is consistent for $\mathcal{M}^*$, which is the RUM obtained from $\mathcal{M}$ by flipping the shapes of the utility distributions. Because $\mathcal{M}$ is symmetric, we have $\mathcal{M}^* = \mathcal{M}$. Then we prove that RBCML$(\mathcal{G} + \mathcal{G}^*, \mathcal{W}_{\text{u}})$ is consistent. If $x + y < 2z$, We subtract $(x + y)\mathcal{G}_{\text{u}}$ from $\mathcal{G} + \mathcal{G}^*$ and get a consistent breaking $(2z - (x + y))\mathcal{G}_1$, which is a contradiction. For the case where $x + y = 2z$ we use the premise in the theorem statement to directly prove that the breaking is inconsistent.

Suppose the theorem holds for $m = k$. When $m = k + 1$, W.l.o.g. we let π_2 satisfy the conditions that $(\ln \pi_i(x))'$ is monotonically decreasing and $\lim_{x\to-\infty}(\ln \pi_i(x))' \to \infty$. Let $\theta_1 = L$, $\theta_m = -L$, and $\theta_2 = \ldots = \theta_{m-1} = 0$. So when $L \to \infty$, with probability goes to 1, a_1 is ranked at the top and a_m is ranked at the bottom. We then focus on $\mathcal{G}_{[2,m]}$ and $\mathcal{G}_{[1,m-1]}$. By induction hypothesis, $\mathcal{G}_{[2,m]}$ (respectively, $\mathcal{G}_{[1,m-1]}$) is either uniform or empty. If $\mathcal{G}_{[2,m]}$ is empty, then $\mathcal{G}_{[1,m-1]}$ is also empty. Because $\mathcal{G}$ is nonempty, we must have $\mathcal{G} = C\mathcal{G}_1$, where $C > 0$. This is a contradiction. If $\mathcal{G}_{[2,m]}$ is uniform but $\mathcal{G}$ is not uniform, then the single edge breaking $\mathcal{G}_1$ must be consistent, which is a contradiction. ■

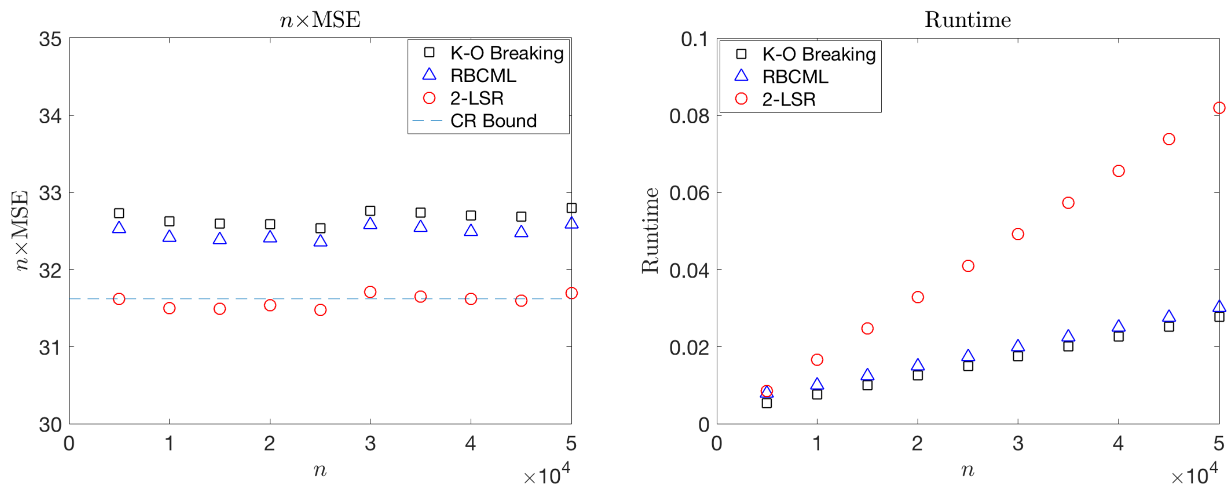


Figure 2. The $n\times$MSE and running time for the Plackett-Luce model. Values are calculated over 50000 trials. "K-O Breaking" denotes the algorithm by Khetan & Oh (2016b), "RBCML" denotes the proposed RBCML with heuristic $\mathcal{W}$, "2-LSR" denotes the 2-iteration I-LSR algorithm by Maystre & Grossglauser (2015). "CR Bound" line is the lower bound of $n\times$MSE for any unbiased estimator.

Corollary 1 *Theorem 7 holds for any RUM with symmetric distributions where any single distribution is Gaussian.*

The following two theorems give stronger characterizations by leveraging Theorems 6 and 7.

Theorem 8 *RBCML($\mathcal{G}, \mathcal{W}$) for Plackett-Luce is consistent if and only if $\mathcal{G}$ is the weighted union of position-k breakings and $\mathcal{W}$ is connected and symmetric.*

Theorem 9 *Let π be any symmetric distribution that satisfies the condition in Theorem 7. Then RBCML($\mathcal{G}, \mathcal{W}$) is consistent for RUM(π) if and only if $\mathcal{G}$ is uniform and $\mathcal{W}$ is connected and symmetric.*

The proofs for Theorems 8 and 9 are similar. The "if" direction can be proved by verifying that the ground truth parameter is the solution to (7). For the "only if" direction, we first prove that consistency of RBCML($\mathcal{G}, \mathcal{W}$) implies consistency of RBCML($\mathcal{G}, \mathcal{W}_{\text{u}}$), which further implies $\mathcal{G}$ is the weighted union of position-k breakings for PLs (Theorem 6) or uniform breaking for RUMs (Theorem 7). Given this condition on $\mathcal{G}$, we prove that $\mathcal{W}$ must be connected and symmetric.

8. The RBCML Framework

The asymptotic covariance of RBCML depends on $\mathcal{G}$ and $\mathcal{W}$. The optimal $\mathcal{G}$ and $\mathcal{W}$ depend on the ground truth parameter $\vec{\theta}_0$[2], which is exactly what we want. To tackle this problem, we propose the adaptive RBCML framework, guided by our Theorems 8 and 9 and shown as Algorithm 1. In this algorithm, $\mathcal{G}$ and $\mathcal{W}$ are iteratively updated given the estimate of $\vec{\theta}$ from the previous iteration.

Algorithm 1 Adaptive RBCML

Input: Profile P of n rankings, the number of iterations T, the heuristics of breaking $\mathcal{G}(\vec{\theta})$ and the weights $\mathcal{W}(\vec{\theta})$.
Output: Estimated parameter $\vec{\theta}^*$.
Initialize $\vec{\theta}^{(0)} = \vec{0}$

```
for t = 1 to T do
    Compute G(θ^(t-1)) and W(θ^(t-1)).
    Estimate θ^(t) using G(θ^(t-1)) and W(θ^(t-1)) by
    maximizing (1) (or (2) for Plackett-Luce)
end for
```

No efficient way of computing the optimal $\mathcal{G}(\vec{\theta})$ and $\mathcal{W}(\vec{\theta})$ is known since the asymptotic covariance is generally hard to compute, where an expectation is taken over $m!$ rankings. How to efficiently compute the optimal $\mathcal{G}$ and $\mathcal{W}$ is a promising future direction. In the experiments of this paper, we use $\mathcal{G}_{\text{u}}$ and $\mathcal{W}_{\text{u}}$ for Gaussian RUMs since $\mathcal{G}_{\text{u}}$ is the only consistent breaking. For the Plackett-Luce model, we use the $\mathcal{G}$ proposed by Khetan & Oh (2016b) and a heuristic $\mathcal{W}(\vec{\theta})$ (See Section 9).

9. Experiments

We compare RBCML with state-of-the-art algorithms for both Gaussian RUMs (GMM algorithm by Azari Soufiani et al. (2014)) and the Plackett-Luce model (the I-LSR algorithm by Maystre & Grossglauser (2015) and the consistent rank-breaking algorithm by Khetan & Oh (2016b)). In both experiments, we generate synthetic datasets of full rankings over $m = 10$ alternatives. The ground truth parameter is generated uniformly at random between 0 and 5 and shifted

[2] Khetan & Oh (2016b) proposed a breaking $\mathcal{G}$, which is not a function of $\vec{\theta}_0$.

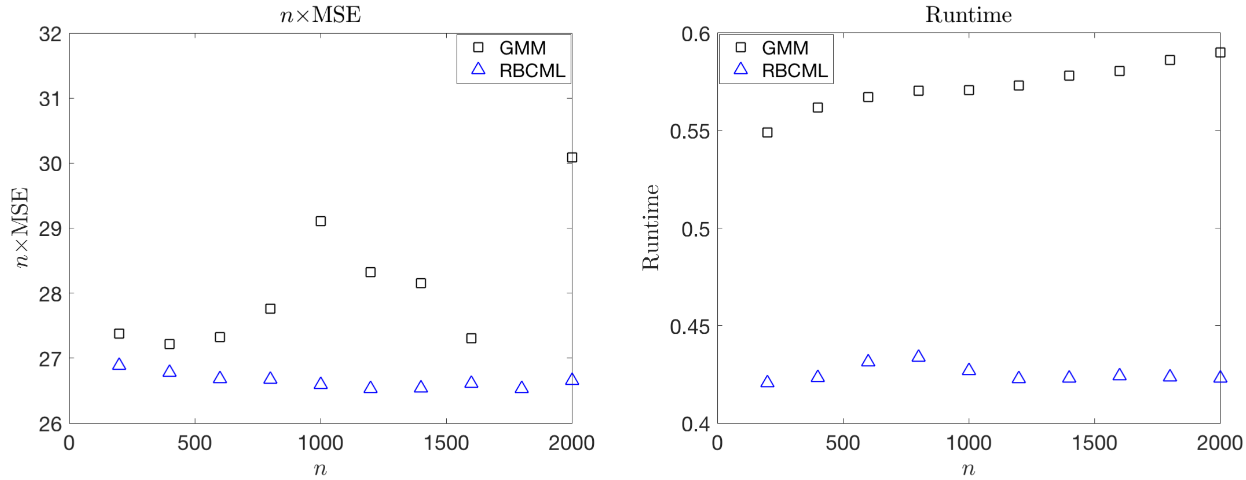


Figure 3. $n\times$MSE and runtime of GMM and RBCML for Gaussian RUMs over 10 alternatives. Values are averaged over 50000 trials.

s.t. $\theta_{10} = 0$. For Gaussian RUMs, the utility distribution of a_i is $N(\theta_i, 1)$. The results are averaged over 50000 trials.

Metrics. We measure statistical efficiency by $n \times$ MSE, where n is the number of rankings in the dataset. **We use $n\times$MSE rather than the standard MSE, because it is easier to see the difference between algorithms w.r.t. the former.** The reason is that $n\times$MSE approaches a positive constant as $n \to \infty$, due to asymptotic normality of RBCML. We use running time to measure computational efficiency of each algorithm.

Gaussian RUMs. We use a one-step ($T = 1$ in Algorithm 1) RBCML($\mathcal{G}_{\text{u}}, \mathcal{W}_{\text{u}}$) for Gaussian RUMs and the results are shown in Figure 3. We use uniform breaking rather than other breakings because it is the only consistent breaking according to our theoretical results.

We observe that our RBCML outperforms the GMM algorithm by Azari Soufiani et al. (2014) w.r.t. both statistical efficiency and computational efficiency.

The Plackett-Luce Model. We use a two-step ($T = 2$ in Algorithm 1) RBCML, where the first step is exactly the algorithm by Khetan & Oh (2016b) (denoted by K-O Breaking). In the second step, we still use the breaking by Khetan & Oh (2016b) but propose a heuristic $\mathcal{W}(\vec{\theta})$. For any pair of alternatives a_{i_1} and a_{i_2}, we let $w_{i_1 i_2} = w_{i_2 i_1} = \frac{1}{|\theta_{i_1} - \theta_{i_2}| + 4}$. The intuition is that we should put a higher weight on the pair of alternatives that are closer to each other. Moreover, we use the output of the first step as the starting point of the second step optimization to improve computational efficiency.

The results are shown in Figure 2. We use 2-LSR to denote the two-iteration I-LSR algorithms by Maystre & Grossglauser (2015). LSR (one-iteration I-LSR) results are not shown because of the high $n\times$MSE and runtime for large n. The "CR bound" line is n times the trace of Cramér-Rao bound (Cramér, 1946; Rao, 1945), which is the lower bound of the covariance matrix of any unbiased estimator. Because Cramér-Rao bound decreases at the rate of $1/n$, the CR bound line is horizontal. Since RBCML is not necessarily unbiased, the Cramér-Rao bound is not a lower bound for RBCML.

We observe that on datasets with large numbers of rankings ("$\succ$" means "is better than"):

- Statistical efficiency: 2-LSR $\succ$ RBCML $\succ$ K-O Breaking.
- Runtime: K-O Breaking $\succ$ RBCML $\succ$ 2-LSR.

Beyond the experiments. We have only shown the RBCML with simple $\mathcal{G}$ and $\mathcal{W}$. Other configurations of $\mathcal{G}$ and $\mathcal{W}$ can potentially have better performances or achieve other tradeoffs. Exploring RBCMLs for Gaussian RUMs, the Plackett-Luce model, as well as other RUMs is an interesting direction for future work.

10. Summary and Future Work

We propose a flexible rank-breaking-then-composite-marginal-likelihood (RBCML) framework for learning RUMs. We characterize conditions for the objective function to be strictly log-concave, and for RBCML to be consistent and asymptotically normal. Experiments show that RBCML for Gaussian RUMs improve both statistical efficiency and computational efficiency, and the proposed RBCML for the Plackett-Luce model is competitive against state-of-the-art algorithms in that it provides a tradeoff between statistical efficiency and computational efficiency. For future work we plan to find efficient ways to compute optimal choices of $\mathcal{G}$ and $\mathcal{W}$, and to extend the algorithm to partial orders.

Acknowledgments

We thank all anonymous reviewers for helpful comments and suggestions. This work is supported by NSF #1453542 and ONR #N00014-17-1-2621.

References

Azari Soufiani, Hossein, Parkes, David C., and Xia, Lirong. Random utility theory for social choice. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pp. 126–134, Lake Tahoe, NV, USA, 2012.

Azari Soufiani, Hossein, Chen, William, Parkes, David C., and Xia, Lirong. Generalized method-of-moments for rank aggregation. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, Lake Tahoe, NV, USA, 2013a.

Azari Soufiani, Hossein, Parkes, David C., and Xia, Lirong. Preference Elicitation For General Random Utility Models. In *Proceedings of Uncertainty in Artificial Intelligence (UAI)*, Bellevue, Washington, USA, 2013b.

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

Ball, Keith and Böröczky, Károly. Stability of the prékopa-leindler inequality. *Mathematika*, 56(2):339–356, 2010.

Berry, Steven, Levinsohn, James, and Pakes, Ariel. Automobile prices in market equilibrium. *Econometrica*, 63 (4):841–890, 1995.

Bhat, Chandra R., Eluru, Naveen, and Copperman, Rachel B. Flexible model structures for discrete choice analysis. In Hensher, David A. and Button, Kenneth J. (eds.), *Handbook of Transport Modelling*, volume 1, pp. 75–104. Emerald Group Publishing Limited, 2nd edition, 2007.

Bockstael, Nancy E. The Use of Random Utility in Modeling Rural Health Care Demand: Discussion. *American Journal of Agricultural Economics*, 81(3):692–695, 1999.

Chen, Yuxin and Suh, Changho. Spectral MLE: Top-k rank aggregation from pairwise comparisons. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, 2015.

Cramér, Harald. Mathematical methods of statistics. 1946.

Dubuc, Serge. Critère De Convexité Et Inégalités Intégrales. *Annales de l'institut Fourier*, 27(1):135–165, 1977.

Gormley, Isobel Claire and Murphy, Thomas Brendan. Exploring voting blocs within the irish exploring voting blocs within the irish electorate: A mixture modeling approach. *Journal of the American Statistical Association*, 103(483):1014–1027, 2008.

Hunter, David R. MM algorithms for generalized Bradley-Terry models. In *The Annals of Statistics*, volume 32, pp. 384–406, 2004.

Khetan, Ashish and Oh, Sewoong. Computational and statistical tradeoffs in learning to rank. In *Advances in Neural Information Processing Systems (NIPS)*, 2016a.

Khetan, Ashish and Oh, Sewoong. Data-driven rank breaking for efficient rank aggregation. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, 2016b.

Lindsay, Bruce G. Composite likelihood methods. *Contemporary Mathematics*, 80:220–239, 1988.

Luce, Robert Duncan. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

Maystre, Lucas and Grossglauser, Matthias. Fast and accurate inference of plackett–luce models. In *Advances in Neural Information Processing Systems*, pp. 172–180, 2015.

McFadden, Daniel L. Economic Choice. Nobel Prize Lecture, 2000.

Negahban, Sahand, Oh, Sewoong, and Shah, Devavrat. Iterative ranking from pair-wise comparisons. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, pp. 2483–2491, Lake Tahoe, NV, USA, 2012.

Pfeiffer, Thomas, Gao, Xi Alice, Mao, Andrew, Chen, Yiling, and Rand, David G. Adaptive Polling and Information Aggregation. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pp. 122–128, Toronto, Canada, 2012.

Plackett, Robin L. The analysis of permutations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 24(2):193–202, 1975.

Rao, C. Radhakrishna. Information and accuracy attainable in the estimation of statistical parameters. *Bull Calcutta. Math. Soc.*, 37:81–91, 1945.

Saumard, Adrien and Wellner, Jon A. Log-concavity and strong log-concavity: A review. *Statistics Surveys*, 8 (45–114), 2014.

Thurstone, Louis Leon. A law of comparative judgement. *Psychological Review*, 34(4):273–286, 1927.

Varin, Cristiano. On composite marginal likelihoods. *Advances in Statistical Analysis*, 92(1):1–28, 2008.

Wang, Shuaiqiang, Huang, Shanshan, Liu, Tie-Yan, Ma, Jun, Chen, Zhumin, and Veijalainen, Jari. Ranking-Oriented Collaborative Filtering: A Listwise Approach. *ACM Transactions on Information Systems*, 35(2):Article No. 10, 2016.

Xu, Ximing and Reid, N. On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference*, 141(9):3047–3054, 2011.

Yang, Rong, Kiekintveld, Christopher, Ordóñez, Fernando, Tambe, Milind, and John, Richard. Improving resource allocation strategy against human adversaries in security games. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, Barcelona, Catalonia, Spain, 2011.

Zhao, Zhibing, Li, Haoming, Wang, Junming, Kephart, Jeffrey, Mattei, Nicholas, Su, Hui, and Xia, Lirong. A cost-effective framework for preference elicitation and aggregation. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI-18)*, 2018a.

Zhao, Zhibing, Villamil, Tristan, and Xia, Lirong. Learning mixtures of random utility models. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, 2018b.