Computing Parametric Ranking Models via Rank-Breaking

Hossein Azari Soufiani
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138, USA
azari@fas.harvard.edu

David C. Parkes
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138, USA
parkes@eecs.harvard.edu

Lirong Xia
Computer Science Department
Rensselaer Polytechnic Institute
Troy, NY 12180, USA
xial@cs.rpi.edu

Abstract

We propose a general algorithm for computing parametric ranking models where the data consist of full rankings. Our main technique is breaking, that is, we first break full rankings in the data to pairwise comparisons. We apply a *generalized method of moments (GMM)* algorithm to compute the parameters. This approach of breaking and then applying GMM often provides a significant computational advantage, and with the right breakings, retains statistical consistency.

We focus on the location family of Random Utility Models (RUMs). We obtain the following results: 1) A complete characterizations of consistent breakings for the Plackett-Luce model and RUMs with flipped Gumbel distributions; 2) For any model in the location family, we prove that if the density functions are all symmetric, then the only consistent breaking is the full breaking; 3) We prove a trichotomy theorem for single-edge breakings to be consistent. Experimental results show that our algorithm with full breaking runs much faster than the state-of-the-art algorithm for MLE inference for models in the location family, while achieving comparable or even better statistical efficiency.

1 Introduction

Suppose there are m alternatives and n agents. Each agent gives a full ranking over alternatives to represent her preferences. We want to find a good understanding of the data in order to make useful inferences. One popular approach is to use parametric models. That is, let Ω denote the set of all possible parameters, given any $\vec{\gamma} \in \Omega$, each ranking is generated i.i.d. according to $\Pr(\cdot|\vec{\gamma})$. Given the agents' rankings D, the maximum likelihood estimator (MLE) computes the parameters $\vec{\gamma}^*$ that maximizes $\Pr(D|\gamma^*)$. Then, we can make inference based on $\vec{\gamma}^*$, for example computing an aggregate ranking. Applications of this approach date back to the Condorcet Jury theorem in the 18th century [7]. The approach has been widely studied in recent years in econometrics [4], social choice [8], and learning to rank [12].

However, for many parametric ranking models the MLE is hard to compute. For example, computing MLE for Condorcet's models is $\mathsf{P}_{||}^{\mathsf{NP}}$-complete [10]. Among all *Random Utility Models (RUMs)* [18], only the *Plackett-Luce (PL)* model [17, 14] is known to have an analytical solution to the likelihood function. Recently, we [1] proposed an Monte-Carlo Expectation-Maximization (MC-EM) algorithm to compute MLE for a general class of RUMs. While this extends the computational reach to more expressive RUMs beyond PL, the running time may still be high for large data sets. On the other hand, parametric models that generate pairwise comparisons are usually easy to compute. Examples include the *Bradley-Terry-Luce (BTL)* model [5, 14] and the *Thurstone-Mosteller (TM)* model [18, 15] for pairwise comparisons.

Our Contributions. First, we propose a general *Generalized Method of Moments (GMM)* algorithm for computing parametric ranking models. Our key technique is *rank-breaking*, that is, each ranking in the data D_r is decomposed into pairwise comparisons and a subset of these pairwise comparisons are selected to constitute the new data D_p. Given this, we apply GMM to compute the parameters w.r.t. D_p. Our algorithm often provides significant computational advantage over the MLE on D_r, and if we do breakings in the right way that will be specified soon, our algorithm will retain the important property of statistical *consistency*.

The comparison of our approach and MLE on D_r is illustrated in Figure 1. Taking a statistical point of view, we first convert the original data D_r to D_p via rank-breaking, then compute parameters of $\mathcal{M}_p$ that generate pairwise comparisons whose parameter space and the ground truth are the same as in the original parametric ranking model $\mathcal{M}_r$. This also provides a new angle of understanding relationships between some classical models, for example PL and BTL. See Section 4.1 for more discussions.

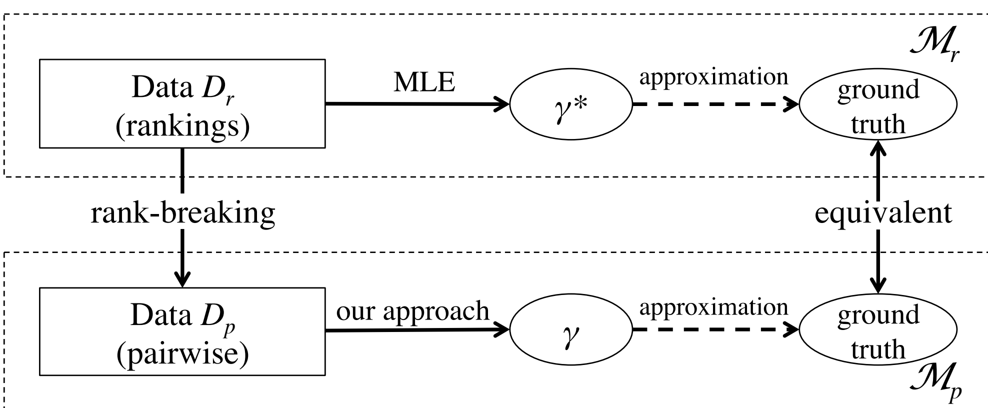


Figure 1: Our approach vs. MLE.

We show that our rank-breaking approach provides a good estimate with large data, by giving several characterizations of consistency w.r.t. the original parametric ranking model $\mathcal{M}_r$. Consistency is a desired statistical property, providing the property that for data generated according to $\mathcal{M}_r$, as the size of data grows without bound, the output of the estimator converges to the ground truth. Focusing on the location family of RUMs, our main theoretical results:

(1) Completely characterize consistent breakings for the Plackett-Luce model and RUMs with a flipped Gumbel distribution.

(2) Prove that if the density function of each distribution is symmetric, then the only consistent breaking is the full breaking.

(3) Provide a trichotomy theorem that characterizes what is required for single-edge breakings to be consistent.

We conduct experimental studies to compare our algorithm to the MC-EM algorithm for RUMs [1]. We consider RUMs with normal distributions and study running time, mean squared error (MSE), and Kendall correlation. Experimental results show that our algorithm achieves comparable, or even better MSE and Kendall correlation, but runs much faster than the MC-EM algorithm, which, to the best of our knowledge, was the only known algorithm for computing MLE of the RUM with normal distributions.

Related Work and Comparisons. Most previous work has focused on computing specific parametric ranking models. Hunter [11] proposes an Minorize-Maximization (MM) algorithm for MLE w.r.t. the PL model. Negahban et al. [16] proposes a rank aggregation algorithm to handle cases where the data are composed of pairwise comparisons and shows that it is consistent for PL applied to pairwise comparisons. Lu and Boutilier [13] propose an Expectation-Maximization (EM) algorithm for MLE inference w.r.t. the Condorcet (Mallows) model. Caragiannis et al. [6] study the consistency of various common voting rules w.r.t. the Condorcet (Mallows) model.

The closest related work is our own [2], where we propose an algorithm for inference in the PL model based on rank-breaking and GMM, and give several sufficient conditions for breakings to be consistent. In this paper, we significantly generalize the idea of GMM and breaking to handle any parametric ranking model. We also obtain more general characterizations on consistency, including complete characterizations of consistent breakings for PL, the RUM with flipped Gumbel distributions, and RUMs with symmetric pdfs. These results help us design computationally efficient and consistent GMM algorithms via rank-breaking.

2 Preliminaries

Let $\mathcal{A} = \{a_1, \ldots, a_m\}$ denote the set of alternatives. Let $D_r = (d_1, \ldots, d_n)$ denote the data, where each d_j is a full ranking over $\mathcal{A}$. Let $\mathcal{L}(\mathcal{A})$ denote the set of all full rankings over $\mathcal{A}$. For any $d \in \mathcal{L}(\mathcal{A})$ and any pair of alternatives a, a', we use $a \succ a'$ to denote (a, a') and write $a \succ_d a'$ if and only if $(a, a') \in d$. Given a *parametric ranking model* $\mathcal{M}_r$, we let $\Omega \subseteq \mathbb{R}^s$ denote the parameter space and let $\Pr_{\mathcal{M}_r}(\cdot|\vec{\gamma})$ denote a distribution over $\mathcal{L}(\mathcal{A})$. Sometimes the subscript in $\Pr_{\mathcal{M}_r}$ is omitted when it does not cause confusion.

Random Utility Models (RUMs)

In a RUM, each alternative a_j is characterized by a utility distribution μ_j, parameterized by a vector $\vec{\gamma}_j$. Given any ground truth $\vec{\gamma} = (\vec{\gamma}_1, \ldots, \vec{\gamma}_m)$, an agent generates a full ranking over $\mathcal{A}$ in the following way: she independently samples a random utility U_j for each alternative a_j with conditional distribution $\Pr_j(\cdot|\theta_j)$, then ranks the alternatives according to their respective perceived utilities, such that she prefers a_{j_1} to a_{j_2} if and only if $U_{j_1} > U_{j_2}$.[1] The probability for a ranking d is the following, where $d(j)$ is the index of the alternative ranked in the jth position:

$$\Pr(d|\vec{\gamma}) = \Pr(U_{d(1)} > U_{d(2)} > \ldots > U_{d(m)})$$

In this paper, the *location family* refers to the class of RUMs where each distribution is only parameterized by its mean. In other words, the shapes of utility distributions are fixed, though they are not necessarily

[1] We ignore the case of ties where $U_{j_1} = U_{j_2}$ since this happens with negligible probability for popular utility distributions.

identical. A *homogeneous location family* is a location family where the shapes of the distributions are identical. In this paper, we will study homogeneous location families with the following distributions on perceived utilities (and thus rankings):

- Gumbel distribution with $\lambda = 1$: the corresponding homogeneous location family is PL.
- Flipped Gumbel distribution: the pdf is $\Pr_G(-x)$, where $\Pr_G = e^{-x}e^{-e^{-x}}$ is the pdf of the Gumbel distribution with $\lambda = 1$.
- Normal distribution: no analytic solution to the likelihood function is known.

Generalized Method-of-Moments

The *Generalized Method-of-Moments (GMM)* provides a wide class of algorithms for parameter estimation. In GMM, we are given a parametric model whose parametric space is $\Omega \subseteq \mathbb{R}^s$, an infinite series of $q \times q$ matrices $\mathcal{W} = \{W_t : t \geq 1\}$, and a column-vector-valued function $g(d, \vec{\gamma}) \in \mathbb{R}^q$. For any vector $\vec{h} \in \mathbb{R}^q$ and any $q \times q$ matrix W, we let $\|\vec{h}\|_W = (\vec{h})^T W \vec{h}$. For any data D, let $g(D, \vec{\gamma}) = \frac{1}{n}\sum_{d \in D} g(d, \vec{\gamma})$, and the GMM method computes parameters $\vec{\gamma}' \in \Omega$ that minimize $\|g(D, \vec{\gamma}')\|_{W_n}$, formally defined as follows:

$$\text{GMM}_g(D, \mathcal{W}) = \{\vec{\gamma}' \in \Omega : \|g(D, \vec{\gamma}')\|_{W_n} = \inf_{\vec{\gamma} \in \Omega} \|g(D, \vec{\gamma})\|_{W_n}\} \tag{1}$$

Since Ω may not be compact (as is the case for PL), the set of parameters $\text{GMM}_g(D, \mathcal{W})$ can be empty. A GMM is *consistent* if and only if for any $\vec{\gamma}^* \in \Omega$, $\text{GMM}_g(D, \mathcal{W})$ converges in probability to $\vec{\gamma}^*$ as $n \to \infty$ and the data is drawn i.i.d. given $\vec{\gamma}^*$.

3 Breakings

In this paper, a *rank-breaking* (*breaking* for short) B_G is defined as a function $\mathcal{L}(\mathcal{A}) \to 2^{\{a \succ a' : a, a' \in \mathcal{A}\}}$ that is represented by an undirected graph G, whose vertices are $\{1, \ldots, m\}$. For any full ranking $d = [a_{i_1} \succ a_{i_2} \succ \cdots \succ a_{i_m}]$, $B_G(d) = \{a_i \succ a_{i'} : a_i \succ_d a_{i'} \text{ and } \{i, i'\} \in G\}$. That is, B_G breaks d into pairwise comparisons for all pairs of alternatives at position i and i' such that $\{i, i'\}$ is an edge in G. If G only contains a single edge, then B_G it is called a *single-edge breaking*. For any data composed of full rankings D_r, we let $B_G(D_r) = \bigcup_{d \in D} B_G(d)$.

We are interested in the following breakings, illustrated in Figure 2:

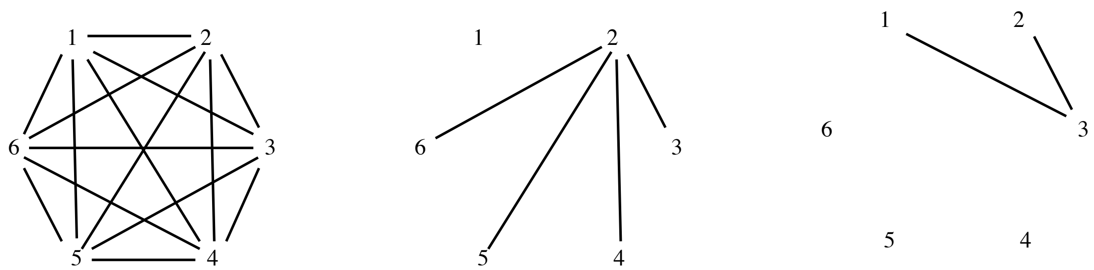
- **Full breaking**: G_F is the complete graph.
- **Position-k breaking**: for any $k \leq m-1$, $G_P^k = \{\{k, i\} : i > k\}$.
- **Position*-k** breaking: for any $k \geq 2$, $G_{P*}^k\{\{k, i\} : i < k\}$.

These breakings are of interest because they are easy to be characterized analytically and we can generate other breakings using union of them.

Example 1. *Let $D_r = \{[a_1 \succ a_2 \succ a_3], [a_2 \succ a_1 \succ a_3]\}$. We have:*
$B_{G_F}(D_r) = \{a_1 \succ a_2, a_1 \succ a_3, a_2 \succ a_3, a_2 \succ a_1, a_2 \succ a_3, a_1 \succ a_3\}$.
$B_{G_P^1}(D_r) = \{a_1 \succ a_2, a_1 \succ a_3, a_2 \succ a_1, a_1 \succ a_3\}$.
$B_{G_{P*}^3}(D_r) = \{a_1 \succ a_3, a_2 \succ a_3, a_2 \succ a_3, a_1 \succ a_3\}$.

4 A General GMM Algorithm

Given a parametric ranking model $\mathcal{M}_r$, for any $i_1 \neq i_2 \leq m$, any $\vec{\gamma} \in \Omega$, and any breaking B_G, we let $f_G^{i_1 i_2}(\vec{\gamma})$ denote the probability that given $\vec{\gamma}$, $a_{i_1} \succ a_{i_2}$ in $G(d)$. That is, $f_G^{i_1 i_2}(\vec{\gamma}) = \Pr_{\mathcal{M}_r}(a_{i_1} \succ a_{i_2} \in B_G(d)|\vec{\gamma})$.



(a) Full breaking. (b) Position-2 breaking. (c) Position*-3 breaking.

Figure 2: Some breakings for $m = 6$.

Definition 1. *Given any breaking B_G, any $d \in \mathcal{L}(\mathcal{A})$, and any $a_{i_1}, a_{i_2} \in \mathcal{A}$, we let:*

- $X_G^{a_{i_1} \succ a_{i_2}}(d) = \begin{cases} 1 & a_{i_1} \succ a_{i_2} \in B_G(d) \\ 0 & \textit{otherwise} \end{cases}$,
- $X_G^{a_{i_1} \succ a_{i_2}}(D_r) = \frac{1}{n} \sum_{d \in D_r} X_G^{a_{i_1} \succ a_{i_2}}(d)$, *and*

$$g_G^{i_1 i_2}(d, \vec{\gamma}) = X_G^{a_{i_1} \succ a_{i_2}}(d) - f_G^{i_1 i_2}(\vec{\gamma}) \tag{2}$$

Where, $X_G^{a_{i_1} \succ a_{i_2}}(D_r)$ is the normalized frequency of times that alternative i_1 is preferred to alternative i_2 ($a_{i_1} \succ a_{i_2}$).

For all $i_1 \neq i_2$, $g_G^{i_1 i_2}(d, \vec{\gamma}) = 0$ are the moment conditions that the ground truth should satisfy in expectation. Let $g_G(d, \vec{\gamma})$ denote the row vector composed of all moment conditions indexed by (i_1, i_2). By definition, for any $\vec{\gamma}^* \in \Omega$, $E_{d|\vec{\gamma}^*}(g_G(d, \vec{\gamma}^*)) = 0$. The GMM algorithm is presented as Algorithm 1.

Algorithm 1: GMM$_G(D_r)$

Input: A breaking B_G and data $D_r = \{d_1, \ldots, d_n\}$ composed of full rankings.
Output: Estimation GMM$_G(D_r)$ of parameters.
For all $1 \leq i_1, i_2 \leq m$, compute $X_G^{a_{i_1} \succ a_{i_2}}(D_r)$.
Compute GMM$_G(D_r)$ according to (1).
return GMM$_G(D_r)$.

Theorem 1. *For any parametric ranking model $\mathcal{M}_r$ and any breaking B_G, Algorithm 1 is consistent if the following conditions are satisfied.*

- *$\vec{\gamma}^*$ is the only solution to $E_{d|\vec{\gamma}^*}(g_G(d, \vec{\gamma})) = 0$.*
- *Ω is compact.*
- *For any $i_1 \neq i_2$, $f_G^{i_1 i_2}$ is continuous in $\vec{\gamma}$.*

All omitted proofs are in Appendix A. The conditions in Theorem 1 are usually satisfied; e.g., we will show that they hold for the location family of RUMs. But the main challenge of Algorithm 1 is that the computation of $\vec{\gamma}$ in Equation (1) on rank data is often intractable, since $f_G^{i_1 i_2}$ might not have an analytic solution, or because the gradient of Equation (1) is too hard to compute. In Section 5, we will see that this computation is tractable for the location family and a large class of breakings.

4.1 A Statistical Viewpoint on the Algorithm

From a statistical viewpoint, a parametric ranking model $\mathcal{M}_r$ that generates rank data can be used to define a parametric model $\mathcal{M}_p$ that generates pairwise comparisons, by retaining the same parameters and rank distribution combined with a second breaking step. Given ranking data D_r, we can use breaking to generate $D_p = G(D_r)$, and then fit $\mathcal{M}_p$ on D_p. Algorithm 1 uses GMM in the fitting process, but we can use any sensible estimator.

The relationship between $\mathcal{M}_r$ and $\mathcal{M}_p$ is illustrated in Figure 3, and recall that the fitting process is illustrated in Figure 1.

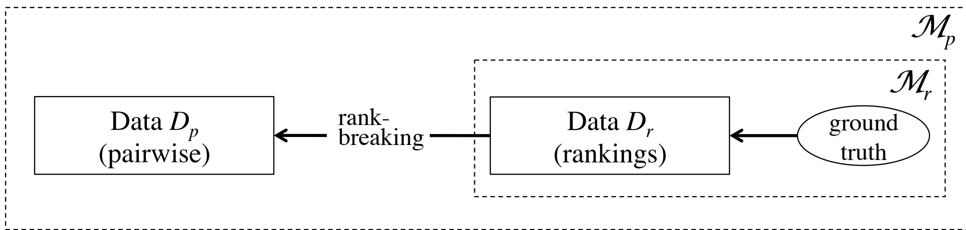


Figure 3: The relationship between $\mathcal{M}_r$ and $\mathcal{M}_p$.

Example 2. *When $\mathcal{M}_r$ is PL, the corresponding $\mathcal{M}_p$ is the BTL model. When $\mathcal{M}_r$ is RUM with normal distributions, the corresponding $\mathcal{M}_p$ is TM for pairwise comparisons.*

5 Application to the Location Family of RUM

In this section, we apply Algorithm 1 to the location family of RUMs.

We recall that in the location family, each utility distribution only has one parameter (its mean). Therefore, we can write $\vec{\gamma} = (\gamma_1, \ldots, \gamma_m)$, where for any $i \leq m$, γ_i is the mean parameter of the utility distribution for a_i. W.l.o.g. let $\gamma_m = 0$.

Definition 2. *A breaking G is* consistent*, if for any $\vec{\gamma}$ and any $i_1 \neq i_2$, $f_G^{i_1 i_2}(\vec{\gamma}) = \Pr_{\mathcal{M}_r}(a_{i_1} \succ a_{i_2}|\vec{\gamma})$.*[2]

By definition, the full breaking is consistent. Let CDF_i denote the CDF of $\Pr_i(\cdot|0)$. For the location family and any consistent breaking G, we have:

$$f_G^{i_1 i_2}(\vec{\gamma}) = \int_{-\infty}^{\infty} \Pr_{i_2}(y)(1 - \text{CDF}_{i_1}(y - \gamma_{i_1} + \gamma_{i_2}))dy$$

Therefore, for the location family we can rewrite $f_G^{i_1 i_2}$ as a function of $\gamma_{i_1} - \gamma_{i_2}$. We have the following proposition.

Proposition 1. *For any model in the location family and any consistent breaking G, $f_G^{i_1 i_2}(\cdot)$ is monotonic increasing on $(-\infty, \infty)$ with $\lim_{x \to -\infty} f_G^{i_1 i_2}(x) = 0$ and $\lim_{x \to \infty} f_G^{i_1 i_2}(x) = 1$. Moreover, if $\Pr_{i_1}$ and $\Pr_{i_2}$ are continuous then $f_G^{i_1 i_2}$ is continuously differentiable with $f_G^{i_1 i_2}(x)' = \int_{-\infty}^{\infty} \Pr_{i_2}(y) \Pr_{i_1}(y - x)dy$.*

Theorem 2. *For any $\mathcal{M}_r$ in the location family and any consistent breaking B_G, if the pdf of every utility distribution is continuous, then Algorithm 1 is consistent.*

[2] The definition of consistent breakings is more general than the definition in [2], which was defined only for PL.

Proof. We use Theorem 1 to prove the consistency of GMM_G. Specifically:

• $\vec{\gamma}^*$ is the only solution to $E_{d|\vec{\gamma}^*}(g_G(d,\vec{\gamma})) = 0$. This follows after the assumption that $f_G^{i_1 i_2}$ is strictly monotonic (so that $g_G^{i_1 i_2}$ is strictly monotonic), and we already have that $E_{d|\vec{\gamma}^*}(g_G(d,\vec{\gamma}^*)) = 0$.

• Ω is compact. This is by definition since $\Omega = \mathbb{R}^{m-1} \times \{0\}$.

• $f_G^{i_1 i_2}$ is continuous in $\vec{\gamma}$. This follows from Proposition 1. □

Corollary 1. *For any consistent breaking B_G, Algorithm 1 is consistent for PL, RUM with flipped Gumbel distributions, and RUM with normal distributions.*

Compared to the MC-EM algorithm [1], Algorithm 1 runs faster since optimizing Equation 1 is much easier; e.g., by using gradient descent or Newton-Raphson. This is because $f_G^{i_1 i_2}(x)'$ is usually easy to compute, and sometimes an analytic solution exists, as shown in the following example. Breaking is particularly helpful here since it enables analytic expression for gradient.

Example 3. *Consider RUM with normal distributions whose variances are* 1. *For any consistent breaking B_G and any $i_1 \neq i_2$ we have:*

$$f_G^{i_1 i_2}(x)' = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} e^{-\frac{(y-x)^2}{2}} dy = \frac{1}{2\sqrt{\pi}} e^{-\frac{x^2}{4}}$$

A similar formula exists for location families with normal distributions whose variances are not identical.

6 Which Breakings are Consistent?

We first present some general lemmas on consistent breakings, then use them as building blocks to prove the main results.

This section provides conditions for the consistency of lower complexity breakings (breakings which take only part of the available ranks). This provides us with a possible tradeoff between time complexity and sample complexity of algorithms we design. We also characterize a condition that only full-breaking is going to be consistent and hence the only applicable approach is full-breaking.

6.1 The Four Core Lemmas

For any model $\mathcal{M}_r$ in the location family, let $\mathcal{M}_r^*$ denote the model in the location family where the pdf of each distribution (whose mean is 0) is flipped around the y-axis. That is, for any $i \leq m$ and any x, $(\Pr_{\mathcal{M}_r})_i(x|0) = (\Pr_{\mathcal{M}_r^*})_i(-x|0)$. For any breaking B_G, we let B_{G^*} denote the breaking such that $\{i,j\} \in G^*$ if and only if $\{m+1-i, m+1-j\} \in G$.

Example 4. *PL* is the RUM with flipped Gumbel distribution. Let $\mathcal{M}_N$ denote the RUM with normal distributions. we have $\mathcal{M}_N = \mathcal{M}_N^*$. For any $k \geq 2$, we have $(G_P^k)^* = G_{P^*}^{m-k}$.*

Lemma 1. *For any $\mathcal{M}_r$ in the location family, if B_G is consistent for $\mathcal{M}_r$, then B_{G^*} is consistent for $\mathcal{M}_r^*$.*

For any graph G and any $1 \leq k_1 < k_2 \leq m$, we let $G_{[k_1,k_2]}$ denote the subgraph of G where the vertices $1,\ldots,k_1$ and $k_2+1,\ldots,m$ are removed, and the vertices are renamed to $1,\ldots,k_2+1-k_1$ by subtracting k_1-1 from all vertices.

Example 5. *For $m = 6$, a breaking B_G and its restriction to $[2,4]$ are shown in Figure 4.*

Lemma 2. *For any model $\mathcal{M}_r$ in the location family, if B_G is consistent then for any $1 \leq k_1 < k_2 \leq m$, either $G_{[k_1,k_2]} = \emptyset$, or $B_{G_{[k_1,k_2]}}$ is consistent for any location family for k_2-k_1+1 alternatives where the utility distributions can be any combination of k_2-k_1+1 utility distributions in $\mathcal{M}_r$.*

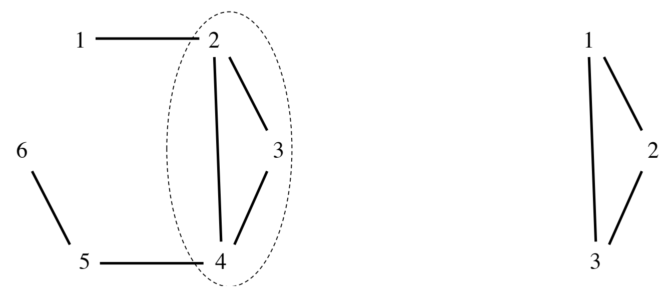


Figure 4: A graph G and $G_{[2,4]}$ for $m = 6$.

Lemma 3. *For any location family where each utility distribution has support* $(-\infty, \infty)$*, the single-edge breaking* $B_{\{\{1,m\}\}}$ *is not consistent.*

The last lemma (specifically, part (3), (4), (5)) is a natural extension of Theorem 4 in [2].

Lemma 4. *Let* B_{G_1}, B_{G_2} *be a pair of breakings.*

- *Suppose both* B_{G_1}*and* B_{G_2} *are consistent,*
  - *(1) if* $G_1 \cap G_2 = \emptyset$*, then* $B_{G_1 \cup G_2}$ *is also consistent;*
  - *(2) if* $G_1 \subsetneq G_2$*, then* $B_{G_2 \setminus G_1}$ *is also consistent.*
- *Suppose* B_{G_1} *is consistent but* B_{G_2} *is not consistent,*
  - *(3) if* $G_1 \cap G_2 = \emptyset$*, then* $B_{G_1 \cup G_2}$ *is not consistent;*
  - *(4) if* $G_1 \subsetneq G_2$*, then* $B_{G_2 \setminus G_1}$ *is not consistent.*
  - *(5) if* $G_2 \subsetneq G_1$*, then* $B_{G_1 \setminus G_2}$ *is not consistent.*

Proof. The proof is based on the following two observations. 1) If $G_1 \cap G_2 = \emptyset$, then $f^{i_1 i_2}_{G_1 \cup G_2}(d) = f^{i_1 i_2}_{G_1}(d) + f^{i_1 i_2}_{G_2}(d)$ and $X^{a_{i_1} \succ a_{i_2}}_{G_1 \cup G_2}(d) = X^{a_{i_1} \succ a_{i_2}}_{G_1}(d) + X^{a_{i_1} \succ a_{i_2}}_{G_2}(d)$. 2) If $G_1 \subsetneq G_2$, then $f^{i_1 i_2}_{G_1 \setminus G_2}(d) = f^{i_1 i_2}_{G_1}(d) - f^{i_1 i_2}_{G_2}(d)$ and and $X^{a_{i_1} \succ a_{i_2}}_{G_1 \setminus G_2}(d) = X^{a_{i_1} \succ a_{i_2}}_{G_1}(d) - X^{a_{i_1} \succ a_{i_2}}_{G_2}(d)$. □

6.2 Results for Specific Location Families

Theorem 3. *For PL, a breaking* B_G *is consistent if and only if G is the union of multiple position-*k *breakings.*

Proof. The "if" direction was proved in [2]. We now prove the "only if" part by induction on m. When $m = 3$, the theorem obviously holds. Suppose the theorem holds for l. When $m = l + 1$, we first apply Lemma 2 to $G_{[2,m]}$. By induction hypothesis, $G_{[2,m]}$ must be the union of position-k breakings for some $k \geq 2$. Now apply Lemma 2 to $G_{[1,m-1]}$. There are two cases.

- **Case 1:** for all $i \leq m - 1$, $\{1, i\} \in G$. We claim that $\{1, m\} \in G$. This is because $B_{\{1,m\} \cup G}$ is consistent, and $B_{\{1,m\}}$ is not consistent due to Lemma 3. Hence $B_{G \setminus \{1,m\}}$ is not consistent.
- **Case 2:** for all $i \leq m - 1$, $\{1, i\} \notin G$. In this case $\{1, m\} \notin G$ following a similar argument as in Case 1.

This means that the theorem holds for $m = l + 1$, which proves the theorem. □

Theorem 3 and Lemma 1 immediately imply the following characterization for RUMs with flipped Gumbel distributions.

Theorem 4. *For the RUM with flipped Gumbel distributions (PL*), B_G is consistent if and only if G is the union of multiple position*-k breakings.*

Theorem 5. *Let $\mathcal{M}_r$ be a model in the location family where each utility distribution has support $(-\infty, \infty)$. If the pdf of each utility distribution in $\mathcal{M}_r$ is symmetric (around the y-axis), then the only consistent breaking is the full breaking.*

Proof. Let B_G denote a consistent breaking. We prove the theorem by induction on m. When $m = 3$, the full breaking is consistent and by Lemma 3, the single edge-breaking $B_{\{\{1,3\}\}}$ is not consistent. By Lemma 4 part (5), $B_{\{\{1,2\},\{2,3\}\}}$ is not consistent.

We now prove that the single-edge breaking $B_{\{\{1,2\}\}}$ is not consistent. For the sake of contradiction suppose it is. By Lemma 1, $B_{\{\{1,2\}\}^*} = B_{\{\{2,3\}\}}$ is consistent for $\mathcal{M}_r^*$. Since all utility distributions in $\mathcal{M}_r$ are symmetric, $\mathcal{M}_r^* = \mathcal{M}_r$. Therefore, $B_{\{\{2,3\}\}}$ is consistent for $\mathcal{M}_r$. By Lemma 4 part (1), $B_{\{\{1,2\},\{1,3\}\}}$ is consistent, which is a contradiction.

Similarly the single-edge breaking $B_{\{\{2,3\}\}}$ is not consistent. It follows from Lemma 4 part (5) that $B_{\{\{1,2\},\{1,3\}\}}$ and $B_{\{\{1,3\},\{2,3\}\}}$ are not consistent. Therefore, the only consistent breaking for $m = 3$ is the full breaking.

Suppose the theorem holds for $m = l$. When $m = l+1$, we first apply Lemma 2 to $G_{[2,m]}$ and $G_{[1,m-1]}$. By induction hypothesis, $G_{[2,m]}$ ($G_{[1,m-1]}$) is either empty or the full graph. We have the following two cases.

Since $m > 3$, if $G_{[2,m]}$ is empty, then $G_{[1,m-1]}$ is empty as well. Since G is non-empty, $G = \{\{1, m\}\}$, which contradicts Lemma 3.

If $G_{[2,m]}$ is full, then $G_{[1,m-1]}$ is full as well. Hence G can be either the full graph G_F, or $G_F \setminus \{1, m\}$. By Lemma 3, $B_{\{\{1,m\}\}}$ is inconsistent, which means that $B_{G_F \setminus \{1,m\}}$ is not consistent (Lemma 4 part (5)).

Therefore, the only remaining case is that G is the full breaking, which means that the theorem holds for $m = l + 1$, which proves the theorem. $\square$

Since the normal distribution is symmetric, we immediately have the following corollary of Theorem 5.

Corollary 2. *For any m and any model in the location family with normal distributions (the variances are not necessary identical), the only consistent breaking is the full breaking.*

Theorem 6. *For any homogeneous location family where each utility distribution has support $(-\infty, \infty)$, if the full breaking is the only consistent breaking for $m = 3$, then the full breaking is the only consistent breaking for any m.*

Proof. The proof is similar to the proof of theorem 5. We prove the theorem by induction on m. $m = 3$ is the assumption. Suppose the theorem holds for l. When $m = l + 1$, we first apply Lemma 2 to $G_{[2,m]}$. By induction hypothesis, $G_{[2,m]}$ is either empty or full.

If $G_{[2,m]}$ is empty, then $G_{[1,m-1]}$ is empty as well. Hence if G is non-empty, then $G = \{\{1, m\}\}$, which contradicts Lemma 3.

If $G_{[2,m]}$ is full, then $G_{[1,m-1]}$ is full as well. Hence G can be either the full graph G_F, or $G_F \setminus \{1, m\}$. By Lemma 3, $B_{\{\{1,m\}\}}$ is inconsistent, which means that $B_{G_F \setminus \{1,m\}}$ is inconsistent (since G_F is always consistent by definition).

Therefore, the theorem holds for $m = l + 1$, which proves the theorem. $\square$

The last result of this section is a trichotomy theorem for single-edge breakings to be consistent for the homogeneous location family.

Theorem 7. *For any m and model $\mathcal{M}_r$ in the homogeneous location family (with support $(-\infty, \infty)$), one and exactly one of the following holds.*

1. No single-edge breaking is consistent.

2. Among all single-edge breakings, only $\{1, 2\}$ is consistent.

3. Among all single-edge breakings, only $\{m-1, m\}$ is consistent.

Proof. For any $k_2 > k_1 + 1$, let us first consider $G_{[k_1,k_2]}$. By Lemma 3, $B_{\{\{1,k_2-k_1+1\}\}}$ is not consistent. Therefore by Lemma 2, any non-adjacent single-edge breaking is not consistent.

Now for an adjacent single-edge graph $\{\{k_1, k_1 + 1\}\}$ that is different from $\{\{1, 2\}\}$ and $\{\{m-1, m\}\}$, by applying Lemma 2 on $G_{[k_1-1,k_1+1]}$ and $G_{[k_1,k_1+2]}$, we have that both $B_{\{\{1,2\}\}}$ and $B_{\{\{2,3\}\}}$ are consistent for the model in the location family with $m = 3$ and any combination of 3 utility distributions in $\mathcal{M}_r$. By Lemma 4 part (1), $\{\{1, 2\}, \{2, 3\}\}$ is consistent, which contradicts Lemma 4 part (5) applied to Lemma 3.

Now, we only need to prove that it is impossible for both $B_{\{\{1,2\}\}}$ and $B_{\{\{m-1,m\}\}}$ to be consistent. If on the contrary both are consistent, then we apply Lemma 2 on $G_{[1,3]}$ and $G_{[m-2,m]}$. Following a similar argument as in the previous paragraph, we can show a contradiction. This proves the theorem. □

This theorems is corresponding to a symmetry notion in the specific location family. We conjecture that case (1) corresponds to the symmetric location families and case (2) and (3) correspond to negative and positive skewness in the locations families distribution.

The next example shows that each of the three cases in Theorem 7 (but not any two of them) holds for some natural location family.

Example 6. *By Corollary 2, the location family with normal distributions belongs to Case 1 in Theorem 7; by Theorem 3, PL belongs to Case 2 in Theorem 7; by Theorem 4, PL* belongs to Case 3 in Theorem 7.*

Remark: Theorem 3 and Theorem 4 are positive, since they tell us that there are consistent breakings other than the full breaking, and some of them are more tractable (compared to the full breaking). On the other hand, Theorem 5 tells us that for certain natural location families, the only breaking that provides a good estimate to the ground truth is the full breaking. This will be demonstrated in the experiments in the next section. Theorem 6 provide a quick check if the full-break is the only consistent breaking by just checking $m = 3$ case. The cases 2 and 3 in Theorem 7 can be falsified by simple testing and can provide correctness of case 1.

7 Experiments

We implemented the MC-EM algorithm [1], Algorithm 1 with the full breaking, and Algorithm 1 with top-3 breaking (whose graph is $G_P^1 \cup G_P^2 \cup G_P^3$) for the homogeneous location family with normal distributions. We evaluate the three algorithms according to run-time and the following two representative criteria. For this, let $\vec{\gamma}^*$ denote the ground truth, and $\vec{\gamma}$ denote the output of the algorithm.

- *Mean Squared Error*: MSE $= E(\|\vec{\gamma} - \vec{\gamma}^*\|_2^2)$.
- *Kendall Rank Correlation Coefficient*: Let $K(\vec{\gamma}, \vec{\gamma}^*)$ denote the *Kendall tau distance* between the ranking over components in $\vec{\gamma}$ and the ranking over components in $\vec{\gamma}^*$. The Kendall correlation is $1 - 2\frac{K(\vec{\gamma},\vec{\gamma}^*)}{m(m-1)/2}$.

The synthetic datasets are generated as follows. Let $m = 5$. The ground truth $\vec{\gamma}^*$ is generated from the Dirichlet distribution Dir$(\vec{1})$. Then, for any given $\vec{\gamma}^*$ we generate up to $n = 200$ full rankings from the location family with normal distributions. All experiments are run on a 2.4 Ghz, Intel Core 2 duo 32 bit laptop.

Table 1 (a) shows the paired t-test on running time for the three methods for $n = 5, 50, 100, 150, 200$, where F, T, M represents values for full breaking, top-3 breaking, and MC-EM, respectively. We clearly

n	F−T	M−T	M−F
5	-10^{-4} (10^{-3})	**17 (.05)**	**17 (.05)**
50	.004 (.005)	**198 (1.3)**	**198 (1.3)**
100	**.008 (.0005)**	**359 (11)**	**359 (11)**
150	**.035 (.004)**	**970 (10)**	**970 (10)**
200	**.017 (.0015)**	**1021 (31)**	**1021 (31)**

(a) Run time.

F−T	M−T	M−F
.001 (.006)	-3×10^{-5} (.005)	**-.001 (.0002)**
-.022 (.008)	**-.017 (.008)**	**.005 (.0001)**
-.018 (.0018)	**-.013 (.0015)**	**.0044 (7×10^{-5})**
-.072 (.0057)	**-.064 (.0062)**	**.008 (.0001)**
-.063 (.001)	**-.05 (.002)**	**.014 (.0004)**

(b) MSE.

n	F−T	M−T	M−F
5	.09 (.55)	.08 (.57)	**-.01 (.001)**
50	.27 (.4)	.26 (.37)	**-.01 (.001)**
100	.08 (.08)	.04 (.08)	**-.04 (.004)**
150	**.34 (.1)**	**.33 (.11)**	**-.01 (.001)**
200	**.29 (.027)**	**.27 (.022)**	**-.02 (.0057)**

(c) Kendall correlation.

Table 1: Paired t-tests for the three algorithms. F, T, M represents values for full breaking, top-3 breaking, and MC-EM, respectively. Mean (std) are shown. Significance results with 95% confidence are in bold.

observe that the running time of Algorithm 1 with full breaking and Algorithm 1 with top-3 breaking are significantly lower than the running time of MC-EM.

Table 1 (b) and (c) show paired t-tests for the three methods, for MSE and Kendall correlation respectively. We note that a lower MSE means that the estimation is more accurate, while a higher Kendall correlation means that the estimation is more accurate. Surprisingly, for both MSE and Kendall correlation, Algorithm 1 with full breaking outperforms MC-EM with 95% confidence for almost all n in our experiments (except MSE $n = 5$). Both algorithms are significantly better than Algorithm 1 with top-3 breaking with 95% confidence when n is not too small. The latter observation is because Algorithm 1 with top-3 breaking is not consistent for the location family with normal distributions.

8 Future Work

We plan to extend the algorithms and analysis to partial orders, non-location families such as RUMs parameterized by mean and variance, and to GRUMs [3].

References

[1] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Random utility theory for social choice. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, pages 126–134, Lake Tahoe, NV, USA, 2012.

[2] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Generalized method-of-moments for rank aggregation. In *submission*, 2013.

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

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

[5] Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[6] Ioannis Caragiannis, Ariel Procaccia, and Nisarg Shah. When do noisy votes reveal the truth? In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, Philadelphia, PA, 2013.

[7] Marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: L'Imprimerie Royale, 1785.

[8] Vincent Conitzer and Tuomas Sandholm. Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 145–152, Edinburgh, UK, 2005.

[9] Lars Peter Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054, 1982.

[10] Edith Hemaspaandra, Holger Spakowski, and Jörg Vogel. The complexity of Kemeny elections. *Theoretical Computer Science*, 349(3):382–391, December 2005.

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

[12] Tie-Yan Liu. *Learning to Rank for Information Retrieval*. Springer, 2011.

[13] Tyler Lu and Craig Boutilier. Learning mallows models with pairwise preferences. In *Proceedings of the Twenty-Eighth International Conference on Machine Learning (ICML 2011)*, pages 145–152, Bellevue, WA, USA, 2011.

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

[15] Frederick Mosteller. Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. II. The Effect of an Aberrant Standard Deviation When Equal Standard Deviations and Equal Correlations Are Assumed. III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, 16(2):3–9, 203–206, 207–218, 1951.

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

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

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

A Proofs

In the appendix, we will sometimes abuse the notation and use G to denote the breaking B_G.

Theorem 1. Algorithm 1 is consistent if the following conditions are satisfied.

- $\vec{\gamma}^*$ is the only solution to $E_{d|\vec{\gamma}^*}(g_G(d,\vec{\gamma}))=0$.
- Ω is compact.
- f_G^{ij} is continuous in $\vec{\gamma}$.

Proof. We prove the theorem by verifying conditions in Theorem 2.1 in [9].

Assumption 2.1: D is stationary and ergodic. This holds because in PL, data in D are generated i.i.d.

Assumption 2.2: Ω is a separable metric space. Since $\mathbb{R}^m$ is a metric separable space and Ω is an subset of $\mathbb{R}^m$, Ω is also separable.

Assumption 2.3: $g_G(\cdot,\vec{\gamma})$ is Borel measurable for each $\vec{\gamma}\in\Omega$ and $g_G(d,\cdot)$ is continuous on Ω for each d. Since the domain of $g_G(\cdot,\vec{\gamma})$ discrete, $g_G(\cdot,\vec{\gamma})$ is continues, which means that $g_G(\cdot,\vec{\gamma})$ is Borel measurable. We note that $g_G(d,\cdot)$ is linear in $f_G^{ij}(\vec{\gamma})$ and by assumption f_G^{ij} is linear in $\vec{\gamma}$, which means that it is continuous.

Assumption 2.4: $E_{d|\vec{\gamma}^*}[g_G(d,\vec{\gamma})]$ exists and is finite for all $\vec{\gamma}\in\Omega$, and $E_{d|\vec{\gamma}^*}[g_G(d,\vec{\gamma}^*)]=0$. The former is because $E_{d|\vec{\gamma}^*}[g_G(d,\vec{\gamma})]$ is linear in $f_G(\vec{\gamma})$ and $f_G(\Omega)$ is bounded. The latter is follows the definition of f_G^{ij}.

Assumption 2.5: The sequence $\mathcal{W}$ converges almost surely to a positive semi-definite matrix. This holds since $W_n=I$ for all t.

Premise (1): $g_G(d,\vec{\gamma})$ is first moment continuous. Since $|g_G(d,\vec{\gamma})|\leq 2$, by Lemma 2.1 of [9], we have that $g_G(d,\vec{\gamma})$ is first moment continuous.

Premise (2): Ω is compact, which is the assumption of our theorem.

Premise (3): $E_{d|\vec{\gamma}^*}[g_G(d,\vec{\gamma})]$ has a unique zero at $\vec{\gamma}^*$. This follows from the assumption of our theorem and the observation that $E_{d|\vec{\gamma}^*}[g_G(d,\vec{\gamma}^*)]=0$. □

Proposition 1. For any model in the location family and any consistent breaking G, $f_G^{i_1i_2}(\cdot)$ is monotonic increasing on $(-\infty,\infty)$ with $\lim_{x\to-\infty}f_G^{i_1i_2}(x)=0$ and $\lim_{x\to\infty}f_G^{i_1i_2}(x)=1$. Moreover, if $\Pr_{i_1}$ and $\Pr_{i_2}$ are continuous then $f_G^{i_1i_2}$ is continuously differentiable with $f_G^{i_1i_2}(x)'=\int_{-\infty}^{\infty}\Pr_{i_2}(y)\Pr_{i_1}(y-x)dy$.

Proof. The first three claims are easy to verify. By Leibniz's rule we have the following calculation.

$$\begin{aligned}f_G^{i_1i_2}(x)' &= \int_{-\infty}^{\infty}\Pr_{i_2}(y)(1-\text{CDF}_{i_1}(y-x))'dy_{i_2}\\ &= \int_{-\infty}^{\infty}\Pr_{i_2}(y)\Pr_{i_1}(y-x)dy\end{aligned}$$

□

Lemma 1. For any $\mathcal{M}_r$ in the location family, if B_G is consistent for $\mathcal{M}_r$, then B_{G^*} is consistent for $\mathcal{M}_r^*$.

Proof. Suppose B_G is consistent for $\mathcal{M}_r$. This means that for any $\vec{\gamma}\in\Omega$,

$$\Pr_{\mathcal{M}_r^*}(x_1,\ldots,x_m|\vec{\gamma})=\Pr_{\mathcal{M}_r}(-x_1,\ldots,-x_m|-\vec{\gamma})$$

Therefore, for any $d \in \mathcal{L}(\mathcal{A})$, we have $\Pr_{\mathcal{M}_r^*}(d|\vec{\gamma}) = \Pr_{\mathcal{M}_r}(rev(d)|-\vec{\gamma})$, where $rev(d)$ is the reverse of d. Meanwhile, for any $d \in \mathcal{L}(\mathcal{A})$ and any breaking B_G, $a_{i_1} \succ a_{i_2} \in B_{G^*}(d)$ if and only if $a_{i_2} \succ a_{i_1} \in B_G(rev(d))$. Therefore, for any $\vec{\gamma}$ and any $i_1 \neq i_2$, we have the following calculation.

$$\begin{aligned}
&\Pr_{\mathcal{M}_r^*}(a_{i_1} \succ a_{i_2} \in B_{G^*}(d)|\vec{\gamma}) \\
=&\Pr_{\mathcal{M}_r}(a_{i_2} \succ a_{i_1} \in B_G(d)|-\vec{\gamma}) \\
=&\Pr_{\mathcal{M}_r}(a_{i_2} \succ a_{i_1}|-\vec{\gamma}) && \text{(Consistency of } B_G \text{ for } \mathcal{M}_r\text{)} \\
=&\Pr_{\mathcal{M}_r^*}(a_{i_1} \succ a_{i_2}|\vec{\gamma})
\end{aligned}$$

It follows that B_{G^*} is consistent for $\mathcal{M}_r^*$. □

Lemma 2. For any model $\mathcal{M}_r$ in the location family, if G is consistent then for any $1 \leq k_1 < k_2 \leq m$, $G_{[k_1,k_2]}$ is either empty or consistent for any location family for $k_2 - k_1 + 1$ alternatives where the utility distributions can be any combination of $k_2 - k_1 + 1$ utility distributions in $\mathcal{M}_r$.

Proof. We prove that if $G_{[k_1,k_2]}$ is not consistent then G is also not consistent. Suppose $G_{[k_1,k_2]}$ is not consistent for a model $\mathcal{M}'_r$ in the location family whose utility distributions are a subset of the utility distributions of $\mathcal{M}_r$. W.l.o.g. suppose the utility distributions in $\mathcal{M}'_r$ are those for $a_{k_1}, \dots, a_{k_2}$ in $\mathcal{M}_r$. Then, there exists $\gamma_{k_1}, \dots, \gamma_{k_2}$ and $1 \leq i, j \leq k_2 - k_1 + 1$ such that:

$$f_G^{ij}(\gamma_{k_1}, \dots, \gamma_{k_2}) \neq \Pr_{\mathcal{M}'_r}(a_i \succ a_j|\gamma_i, \gamma_j) = \Pr_{\mathcal{M}'_r}(U_i > U_j|\gamma_i, \gamma_j)$$

We now construct other components in $\vec{\gamma}$ to show that G is not consistent for $\mathcal{M}_r$. Let $\gamma_1 = \ldots = \gamma_{k_1-1}$ go to ∞ and let $\gamma_{k_2+1} = \ldots = \gamma_m$ go to $-\infty$. Then, with probability that goes to 1, $a_1, \dots, a_{k_1-1}$ are ranked in the top $k_1 - 1$ positions and $a_{k_2+1}, \dots, a_m$ are ranked in the bottom $m - k_2 + 1$ positions. Hence we have

$$\lim_{\gamma_1 \to \infty, \gamma_m \to -\infty} |f_G^{(i+k_1-1)(j+k_1-1)}(\vec{\gamma}) - f_{G_{[k_1,k_2]}}^{ij}(\gamma_{k_1}, \dots, \gamma_{k_2})| = 0$$

We note that for any $1 \leq i, j \leq k_2 - k_1 + 1$,

$$\Pr_{\mathcal{M}'_r}(U_i > U_j|\gamma_i, \gamma_j) = \Pr_{\mathcal{M}_r}(U_{i+k_1-1} > U_{j+k_1-1}|\gamma_{i+k_1-1}, \gamma_{j+k_1-1})$$

Therefore, there exists $\vec{\gamma}$ so that

$$|f_G^{(i+k_1-1)(j+k_1-1)}(\vec{\gamma}) - f_{G_{[k_1,k_2]}}^{ij}(\gamma_{k_1}, \dots, \gamma_{k_2})| < |f_G^{ij}(\gamma_{k_1}, \dots, \gamma_{k_2}) - \Pr_{\mathcal{M}'_r}(U_i > U_j|\gamma_i, \gamma_j)|$$

This proves that G is not consistent. □

Lemma 3. For any location family where each utility distribution has support $(-\infty, \infty)$, the single-edge breaking $\{\{1, m\}\}$ is not consistent.

Proof. We will first give a high-level description of the proof. Let $G = \{\{1, m\}\}$. Let p_i denote the pdf of u_i and F_i denote the cdf of u_i. W.l.o.g. let $\gamma_m = 0$. Depending on the shape of the distribution, we will prove the inconsistency using the following two models.

1. $\mathcal{M}_1$: $\gamma_1 > \gamma_2 = \cdots = \gamma_{m-1} \gg \gamma_m$.
2. $\mathcal{M}_1$: $\gamma_1 \gg \gamma_2 = \cdots = \gamma_{m-1} > \gamma_m$.

For the sake of contradiction, suppose G is consistent. We will show that in either $\mathcal{M}_1$ or $\mathcal{M}_2$ we have

$$\frac{\Pr(a_1 \succ a_m|\vec{\gamma})}{\Pr(a_m \succ a_1|\vec{\gamma})} \neq \frac{\Pr(a_1 \text{ at top and } a_m \text{ at bottom}|\vec{\gamma})}{\Pr(a_1 \text{ at bottom and } a_m \text{ at top}|\vec{\gamma})}$$

which is equivalent to the following according to Bayes' rule.

$$\frac{\Pr(a_1 \text{ at top and } a_m \text{ at bottom}|a_1 \succ a_m, \vec{\gamma})}{\Pr(a_1 \text{ at bottom and } a_m \text{ at top}|a_m \succ a_1, \vec{\gamma})} \neq 1 \quad (3)$$

The intuition behind the proof is that for $\mathcal{M}_1$, given $a_1 \succ a_m$, it is very likely that a_1 is ranked in the top place and a_m is ranked in the bottom place, which means

$$\Pr(a_1 \text{ at top and } a_m \text{ at bottom}|a_1 \succ a_m, \vec{\gamma}_{\mathcal{M}_1}) \approx 1$$

Meanwhile, if we can construct $\vec{\gamma}_{\mathcal{M}_1}$ such that given $a_m \succ a_1$, it is very likely that U_1 is much larger than γ_1, which means that it is very unlikely that a_1 is ranked in the top position, then we will have

$$\Pr(a_1 \text{ at bottom and } a_m \text{ at top}|a_m \succ a_1, \vec{\gamma}) \approx 0$$

Subsequently we will have (3). If we cannot find $\vec{\gamma}_{\mathcal{M}_1}$ such that given $a_m \succ a_1$, U_1 is much smaller than γ_1 with high probability (note that this probability only depends on γ_1 and γ_m), then we will show that given $a_m \succ a_1$, U_m is much smaller than $\gamma_m = 0$ with high probability, which implies (3).

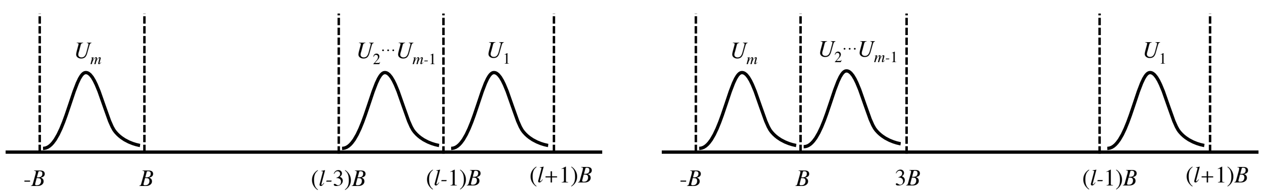
Formally, let $B \in \mathbb{R}_{\geq 0}$ denote an arbitrary number such that for all $1 \leq i \leq m$,

$$\Pr[|U_i - \gamma_i| < B] > 1 - 1/(10m)$$

For a natural number l that will be specified later, we define $\mathcal{M}_1$ and $\mathcal{M}_2$, illustrated in Figure 5.

$\mathcal{M}_1$: $\gamma_1 = lB$, $\gamma_2 = \cdots = \gamma_{m-1} = (l-2)B$, $\gamma_m = 0$.

$\mathcal{M}_2$: $\gamma_1 = lB$, $\gamma_2 = \cdots = \gamma_{m-1} = 2B$, $\gamma_m = 0$.



(a) $\mathcal{M}_1$. (b) $\mathcal{M}_2$

Figure 5: Illustration of $\mathcal{M}_1$ and $\mathcal{M}_2$.

We will show that for a large enough l, either $\Pr(U_m < (l-3)B|U_m > U_1) > 1/3$ or $\Pr(U_1 > 3B|U_m > U_1) > 1/3$. It suffices to show that $\Pr(U_m < (l-3)B \text{ or } U_1 > 3B|U_m > U_1) > 2/3$. We will prove the following stronger claim:

$$\lim_{l \to \infty} \Pr(U_m \geq (l-3)B \text{ and } U_1 \leq 3B|U_m > U_1) = 0$$

We have

$$
\begin{aligned}
&\lim_{l\to\infty} \frac{\Pr(U_m \geq (l-3)B \text{ and } U_1 \leq 3B | U_m > U_1)}{\Pr(3B < U_m < (l-3)B \text{ and } U_1 \leq 3B | U_m > U_1)} \\
=&\lim_{l\to\infty} \frac{\Pr(U_m \geq (l-3)B \text{ and } U_1 \leq 3B)}{\Pr(3B < U_m < (l-3)B \text{ and } U_1 \leq 3B)} \\
=&\lim_{l\to\infty} \frac{\Pr(U_m \geq (l-3)B)\Pr(U_1 \leq 3B)}{\Pr(3B < U_m < (l-3)B)\Pr(U_1 \leq 3B)} \\
=&\lim_{l\to\infty} \frac{\Pr(U_m \geq (l-3)B)}{\Pr(3B < U_m < (l-3)B)} \\
=&\frac{0}{1-F_m(3B)} = 0
\end{aligned}
$$

Hence, for large enough l, either $\Pr(U_m < (l-3)B | U_m > U_1) > 1/3$ or $\Pr(U_1 > 3B | U_m > U_1) > 1/3$.

If $\Pr(U_m < (l-3)B | U_m > U_1) > 1/3$, then in $\mathcal{M}_1$, given $a_m \succ a_1$, the probability that U_m is smaller than any of $U_2, \ldots, U_{m-1}$ is at least $1/3 * 0.9$, which means that

$$\Pr_{\mathcal{M}_1}(a_1 \text{ at bottom and } a_m \text{ at top} | a_m \succ a_1, \vec{\gamma}) < 1 - 1/3 * 0.9 < 0.9$$

Meanwhile,

$$\Pr_{\mathcal{M}_1}(a_1 \text{ at top and } a_m \text{ at bottom} | a_1 \succ a_m, \vec{\gamma}) > 1 - \frac{1}{10m} m = 0.9$$

This implies (3), which means that $\{\{1, m\}\}$ is not consistent.

If $\Pr(U_1 > 3B | U_m > U_1) > 1/3$, then in M_2, given $a_m \succ a_1$, the probability that U_1 is larger than any of $U_2, \ldots, U_{m-1}$ is at least $1/3 * 0.9$, which means that

$$\Pr_{\mathcal{M}_2}(a_1 \text{ at bottom and } a_m \text{ at top} | a_m \succ a_1, \vec{\gamma}) < 1 - 1/3 * 0.9$$

Similarly we have

$$\Pr_{\mathcal{M}_2}(a_1 \text{ at top and } a_m \text{ at bottom} | a_1 \succ a_m, \vec{\gamma}) > 1 - \frac{1}{10m} m = 0.9$$

Again (3) holds, which means that $\{\{1, m\}\}$ is not consistent. $\square$