The Limit of Rationality in Choice Modeling

Srikanth Jagabathula
NYU Stern School of Business, New York

Joint with:
Paat Rusmevichientong
USC Marshall School of Business
Random Utility Max. (RUM) class: most popular discrete-choice model class used for demand predictions

**Goal:** predict customer’s choice from a menu of items

**Model:** discrete-choice model to capture substitution

**Assumes**
- randomly sample product utilities
- choose max utility product

**Models**
- multinomial logit (MNL),
- nested logit (NL), mixed logit (ML), rank-based,…
How to select a good model?

Models to fit to transaction-level data

- Collection of offer sets $S_1, S_2, \ldots, S_m$
- Sales fractions $f_{jS} = \text{fraction of times } j \text{ purchased when } S \text{ was offered}$
RUM model fit: are revealed customer preferences transitive?

Question: are the choice observations (stochastically) rationalizable?
RUM model fit: are revealed customer preferences transitive?

**Question:** are the choice observations (stochastically) rationalizable?

sales fractions

<table>
<thead>
<tr>
<th>Offer Set 1</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Offer Set 2</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
RUM model fit: are revealed customer preferences transitive?

**Question:** are the choice observations (stochastically) rationalizable?

<table>
<thead>
<tr>
<th>Offer Set 1</th>
<th>Offer Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales Fractions</td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>50%</td>
<td>60%</td>
</tr>
<tr>
<td>30%</td>
<td>40%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type 1A</th>
<th>Type 1B</th>
<th>Type 2</th>
<th>Type 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>10%</td>
<td>50%</td>
<td>30%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top Choices</td>
<td>Second Choices</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
P(2 | 2 3) = 10\% + 50\% = 60\%\]
Question: are the choice observations (stochastically) rationalizable?

Yes, choices consistent w/ distribution over rankings.
RUM model fit: are revealed customer preferences transitive?

**Question:** are the choice observations *(stochastically)* rationalizable?

<table>
<thead>
<tr>
<th>Offer Set 1</th>
<th>Offer Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>Type 1</td>
</tr>
<tr>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>30%</td>
<td>60%</td>
</tr>
<tr>
<td>Type 2</td>
<td>Type 2</td>
</tr>
<tr>
<td>40%</td>
<td>50%</td>
</tr>
<tr>
<td>60%</td>
<td>30%</td>
</tr>
<tr>
<td>Type 3</td>
<td>Type 3</td>
</tr>
<tr>
<td>20%</td>
<td>50%</td>
</tr>
<tr>
<td>30%</td>
<td>30%</td>
</tr>
</tbody>
</table>

sales of 2 cannot decrease if preferences are rational

Choices not consistent with ANY distribution over rankings

**CHOICE DATA NOT STOCHASTICALLY RATIONALIZABLE**
The LIMIT OF RATIONALITY: the smallest possible loss from assuming rationality

\[ \text{Limit of Rationality} \quad \text{LoR} \ (S_1, \ldots, S_m) \]

\[ \min_{\mathbf{x}} \quad \sum_S \text{loss} \left( \left( x_{j,S} : j \in S \right), \left( f_{j,S} : j \in S' \right) \right) \]

- **\( n \)** mutually substitutable products
- **observations**: choice/sales fractions for sets \( S_1, S_2, \ldots, S_m \)
- \( f_{j,S} = \) fraction of times \( j \) purchased when \( S \) offered

**Convex loss fn.**

**Obs. choice fractions**

**Rational choice fractions consistent with a distribution over rankings**
The LIMIT OF RATIONALITY:
the smallest possible loss from assuming rationality

$n$ mutually substitutable products

observations: choice/sales fractions for sets $S_1, S_2, \ldots, S_m$

$f_{j,S} =$ fraction of times $j$ purchased when $S$ offered

Limit of Rationality
LoR ($S_1, \ldots, S_m$)

$$\min_x -\frac{1}{m} \sum_{S} \sum_{j \in S} f_{j,S} \log \left( \frac{x_{j,S}}{f_{j,S}} \right)$$

KL-divergence:
between obs. and pred. sales fractions

s.t. $x_{j,S} = \sum_{\sigma} \lambda_\sigma \cdot \mathbb{I} [j \text{ top ranked among items in } S \text{ under } \sigma] \quad \forall \ j, S$

$\sum_{\sigma} \lambda_\sigma = 1, \lambda_\sigma \geq 0 \quad \forall \ \sigma$
The LIMIT OF RATIONALITY: the smallest possible loss from assuming rationality

Limit of Rationality
\[ \text{LoR } (S_1, \ldots, S_m) \]

\[ \min_{x} - \frac{1}{m} \sum_{S} \sum_{j \in S} f_{j,S} \log (x_{j,S} / f_{j,S}) \]

KL-divergence:
between obs. and pred. sales fractions

\[ s.t. \quad x_{j,S} = \sum_{\sigma} \lambda_{\sigma} \cdot \mathbb{I}[j \text{ top ranked among items in } S \text{ under } \sigma] \quad \forall j, S \]

\[ \sum_{\sigma} \lambda_{\sigma} = 1, \lambda_{\sigma} \geq 0 \quad \forall \sigma \]

vectors of length \( L \)
\[ |S_1| + |S_2| + \cdots + |S_m| = L \]
\[ x \in \text{conv} \left( \{ e_{\sigma} : \sigma \text{ is a ranking} \} \right) \]
convex hull of permutation representations

\[ (e_{\sigma})_{j,S} = \mathbb{I}[j \text{ top ranked among items in } S \text{ under } \sigma] \]
The LIMIT OF RATIONALITY: 
the smallest possible loss from assuming rationality

\[
\text{Limit of Rationality} \quad \text{LoR } (S_1, \ldots, S_m) \\
\min_{\mathbf{x}} \quad -\frac{1}{m} \sum_{S} \sum_{j \in S} f_{j,S} \log \left( \frac{x_{j,S}}{f_{j,S}} \right)
\]

\[\text{s.t.} \quad \mathbf{x} \in \text{conv} \left( \left\{ e_\sigma : \sigma \text{ is a ranking} \right\} \right)\]

KL-divergence: \nbetween obs. and pred. sales fractions
The LIMIT OF RATIONALITY:
the smallest possible loss from assuming rationality

Limit of Rationality
LoR \((S_1, \ldots, S_m)\)

\[
\min_{\mathbf{x}} \ -\frac{1}{m} \sum_{S} \sum_{j \in S} f_{j,S} \log \left( \frac{x_{j,S}}{f_{j,S}} \right)
\]

KL-divergence:
between obs. and pred. sales fractions

\[
\text{s.t.} \quad \mathbf{x} \in \text{conv} \left( \{ \mathbf{e}_\sigma : \sigma \text{ is a ranking} \} \right)
\]

Key Contributions:

1. characterization of the complexity of the above polytope
   in terms of structure of the set collection \(S_1, \ldots, S_m\)
2. model-selection diagnostic tool for practitioners
Existing work doesn’t address computation

Economics

Focus is on the binary question:
are the observed sales fractions stochastically rationalizable?

Key result: necessary & sufficient conditions for \( \text{LoR}(S_1, \ldots, S_m) = 0 \)

[Barbera, Pattanaik ’86][McFadden, Richter ‘90][McFadden ‘05]

Does not address computability

Computer Science

Focus on rank aggregation for PAIRS = \( \{ \{i, j\} \mid i \neq j \} \)

Key results: NP-hardness [Dwork 01] and approx. algorithms

[Dwork, Kumar, Naor, Sivakumar ’01][Kenyon-Mathieu, Schudy ’07][Ali, Melia ’12]

Does not consider general set collections

Computational hardness depends on the structure of collection of subsets
The Limit of Rationality in Choice Modeling

1. Case study with IRI Academic Dataset
   what is the ‘right’ choice model?

2. Complexity of computing LoR
   rational separation, complexity in terms of graph properties

3. Summary/Conclusions
   takeaway messages
1. Case study with IRI Academic Dataset
what is the ‘right’ choice model?

2. Complexity of computing LoR
rational separation, complexity in terms of graph properties

3. Summary/Conclusions
takeaway messages
LoR provides guidance on which model to fit: more complex RUM or outside RUM?

Distance measured in terms of KL-div

\[ D_{KL} (\text{obs. choice prob.} \parallel \text{pred. choice prob.}) \]

Total loss = Rationality loss + Parametric loss

To reduce:

- Parametric loss: fit more complex RUM models
- Rationality loss: go outside RUM class
Rationality loss is the largest component of loss.

More complex RUM model only reduces parametric loss.

Provides a bound on the number of mixture components.
Region 1: MNL acceptable

Region 2: fit more complex RUM models

Region 3: RUM unacceptable
Region 1: MNL acceptable

Region 2: fit more complex RUM models

Region 3: RUM unacceptable
Factors influencing rationality loss: variety-seeking and hedonic purchasing

\[ \text{loss}_j = \beta_0 + \beta_1 \cdot (\text{market concentration})_j + \beta_2 \cdot (\text{hedonic indicator})_j + \varepsilon_j \]

- negative of market share entropy
- food/snack (excl. condiments)
- cigarettes
- drinks
- variety seeking \( \Rightarrow \) low concentration
  (e.g., yogurt w/ new product intros.)
- hedonic \( \Rightarrow \) complex choice behavior
- non-transitive preferences
- non-transitive preferences
1. Case study with IRI Academic Dataset
what is the ‘right’ choice model?

2. Complexity of computing LoR
rational separation, complexity in terms of graph properties

3. Summary/Conclusions
takeaway messages
Computing the Limit of Rationality: solve a large-dimensional constrained convex program

\[ \min_x -\frac{1}{m} \sum_{S} \sum_{j \in S} f_{j,S} \log \left( \frac{x_{j,S}}{f_{j,S}} \right) \]

s.t. \( x \in \text{conv} \left( \{ e_{\sigma} : \sigma \text{ is a ranking} \} \right) \)

Constrained convex program in a large dimensional space

Frank-Wolfe algorithm

\( \min_{x \in \mathcal{D}} h(x) \) in iteration \( k+1 \), optimize linear approx. \( \min_{y \in \mathcal{D}} \langle y, \nabla h(x^{(k)}) \rangle \)

\( x^{(k+1)} = \text{convex combination of } x^{(k)} \) and \( \arg \min_{y \in \mathcal{D}} \langle y, \nabla h(x^{(k)}) \rangle \)
Computing the Limit of Rationality: solve a large-dimensional constrained convex program

\[
\text{LoR } (S_1, \ldots, S_m) \quad \min_x \left\{ -\frac{1}{m} \sum_S \sum_{j \in S} f_{j,S} \log \left( \frac{x_{j,S}}{f_{j,S}} \right) \right\} \\
\text{subject to } x \in \text{conv}\left( \{e_\sigma : \sigma \text{ is a ranking}\} \right)
\]

\[
\min_{y \in \mathcal{D}} \langle y, \nabla h(x^{(k)}) \rangle \quad \text{in our case} \quad \min_y \left\{ -\frac{1}{m} \sum_S \sum_{j \in S} (f_{j,S}/x_{j,S}^{(k)}) \cdot y_{j,S} \right\} \\
\text{subject to } y \in \text{conv}\left( \{e_\sigma : \sigma \text{ is a ranking}\} \right)
\]

because LP, optimal at extreme point

\[
\min_{\sigma} \left\{ -\frac{1}{m} \sum_S \sum_{j \in S} (f_{j,S}/x_{j,S}^{(k)}) \cdot (e_\sigma)_{j,S} \right\}
\]
Computing the Limit of Rationality: solve a rank-aggregation problem

FW-subproblem
optimize linear approx.

\[
\min_{\sigma} \quad \frac{-1}{m} \sum_S \sum_{j \in S} (f_{j,S}/x_{j,S}^{(k)}) \cdot (e_{\sigma})_{j,S}
\]

\[
\min_{\sigma} \sum_S \sum_{j \in S} c_{j,S} \cdot \mathbb{I}[j \text{ is top-ranked among items in } S \text{ under } \sigma]
\]

\[\text{cost coef.} \quad -f_{j,S}/(m \cdot x_{j,S}^{(k)})\]

Also called the RANK AGGREGATION problem

LoR choosing best ranking \equiv choosing best item from each subset

computational complexity \equiv computational complexity
The challenge in solving rank-aggregation: choices from different sets are not independent.

\[
\begin{align*}
\min_{x,y,z} & \quad c_{1x} + c_{2y} + c_{3z} \\
& \quad x \text{ chosen from set A} \\
& \quad y \text{ chosen from set B} \\
& \quad z \text{ chosen from set C} \\
& \quad \text{such that the choices are transitive}
\end{align*}
\]
The challenge in solving rank-aggregation: choices from different sets are not independent

\[
\begin{align*}
\min_{x,y,z} & \quad c_{1x} + c_{2y} + c_{3z} \\
\text{s.t.} & \quad x \text{ chosen from set A} \\
& \quad y \text{ chosen from set B} \\
& \quad z \text{ chosen from set C} \\
& \quad 1 \succ 2 \succ 3
\end{align*}
\]
The challenge in solving rank-aggregation: choices from different sets are not independent

\[
\begin{align*}
\min_{x,y,z} c_{1x} + c_{2y} + c_{3z} \\
&x \text{ chosen from set } A \\
&y \text{ chosen from set } B \\
&z \text{ chosen from set } C \\
&\text{such that the choices are transitive}
\end{align*}
\]

choices NOT transitive
2 > 3 contradicts 3 > 2
Solution: “conditional” choices can be independent

<table>
<thead>
<tr>
<th></th>
<th>(c_{11})</th>
<th>(c_{12})</th>
<th>(c_{13})</th>
<th>(c_{14})</th>
</tr>
</thead>
<tbody>
<tr>
<td>set A</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

choice \(x\)

<table>
<thead>
<tr>
<th></th>
<th>(c_{22})</th>
<th>(c_{23})</th>
<th>(c_{24})</th>
</tr>
</thead>
<tbody>
<tr>
<td>set B</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

choice \(y\)

<table>
<thead>
<tr>
<th></th>
<th>(c_{33})</th>
<th>(c_{34})</th>
</tr>
</thead>
<tbody>
<tr>
<td>set C</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

choice \(z\)
Solution: “conditional” choices can be independent

\[ \begin{array}{cccc}
1 & 2 & 3 & 4 \\
\text{set A} & c_{11} & c_{12} & c_{13} & c_{14}
\end{array} \]

\[ \begin{array}{ccc}
2 & 3 & 4 \\
\text{set B} & c_{22} & c_{23} & c_{24}
\end{array} \]

\[ \begin{array}{cc}
3 & 4 \\
\text{set C} & c_{33} & c_{34}
\end{array} \]

choice \( x \)

\[ \begin{array}{cc}
1 & 2 \\
\text{choice x}
\end{array} \]

choice \( y \)

\[ \begin{array}{c}
2 \\
\text{choice y}
\end{array} \]

choice \( z \)

\[ \begin{array}{cc}
3 & 4 \\
\text{choice z}
\end{array} \]

ALL POSSIBLE CHOICE COMBINATIONS ARE CONSISTENT

\( (x, y, z) \)

consistent with

\( 1 > 2 > 3 > 4 \)

\( 1 > 2 > 4 > 3 \)

\( 2 > 1 > 3 > 4 \)

\( 2 > 1 > 4 > 3 \)
Solution: “conditional” choices can be independent

Set A: 
- Choice x
  - 1
  - 2
  - 3
  - 4

Set B: 
- Choice y
  - c_{11} = 2
  - c_{12} = 3
  - c_{13} = 4

Set C: 
- Choice z
  - c_{33} = 3
  - c_{34} = 4
Solution: “conditional” choices can be independent

Given choice \( y \) in set B,

- choice \( x \) consistent with \( y \)
- choice \( z \) consistent with \( y \)

Therefore, choices \( x, y, z \) mutually consistent.
The concept of “rational separation”

Given choice $y$ in set $B$,

choice $x$ consistent w/ $y$ and choice $z$ consistent w/ $y$

choices $x,y,z$ mutually consistent for all choices $x,y,z$

$B$ rationally separates $A$ and $C$  

$A \perp C \mid B$

Definition extends to collections of subsets: $\mathcal{A} \ B \ C$  

$A \perp C \mid B$
Choice graph: graphical representation of rational separation

$G$ is a choice-graph with subsets as nodes if whenever $B$ separates $A$ and $C$ in $G$, $B$ must rationally separate $A$ and $C$, i.e., $A \nparallel C \mid B$ for all collections of nodes $A$, $B$, and $C$.

B separates $A$ and $C$ in $G$: if set $B$ is removed, then $A$ and $C$ are disconnected or every path from set $A$ to set $C$ goes through set $B$.
Rational separation reduces computational complexity of rank-aggregation

Reduces search space from $O(n^3)$ to $O(n^2)$
[for each choice from $C$, search over $O(n)$ choices]

for every element in $B$, find best element in $A$
$O(n^2)$

for every element in $B$, find best element in $C$
$O(n^2)$

Exploit rational separation to join the two solutions $O(n^2)$
Exploiting rational separation:
DP for finding the best set of choices

\[ V_A(x) = c_{1x} + \min_y \left\{ V_B(y) \middle| x \& y \text{ consistent} \right\} \]

Theorem:
If the underlying choice graph is a line graph, then rank aggregation LP can be solved via a DP in \( O(n|M|) \) operations

Through pre-processing, computed in \( O(1) \)
Exploiting rational separation: DP for finding the best set of choices for a tree

Theorem:
If the underlying choice graph is a tree, then rank aggregation LP can be solved via a DP in $O(n|M|)$ operations

$$V_A(x) = c_1x + \sum_{i=1}^{m} \min_{y_i} \left\{ V_{B_i}(y_i) \mid x \ & \ y_i \ \text{consistent} \right\}$$
Exploiting rational separation: rank aggregation for a tree can also be solved with an LP

\[
\min_{x} \sum_{S \in M} \sum_{i \in S} c_{i,S} x_{i,S} \quad \rightarrow x_{i,S} = 1 \text{ if } i \text{ is chosen from } S; \ 0 \text{ o.w.}
\]

if \( r \) chosen from \( A \), then
\( r \) must be chosen from \( A \cap B \)

\[
z_{r,A,B} \geq x_{r,A} \ \forall \ r \in A \cap B
\]

\[
\sum_{i \in A} x_{i,A} = 1
\]

set \( A \)

\[
\sum_{i \in B} x_{i,B} = 1
\]

set \( B \)

if \( r \) chosen from \( B \), then
\( r \) must be chosen from \( A \cap B \)

\[
z_{r,A,B} \geq x_{r,B} \ \forall \ r \in A \cap B
\]

\[
\sum_{r \in A \cap B} z_{r,A,B} = 1
\]
Exploiting rational separation: rank aggregation for a tree can also be solved with an LP

\[
\min_x \sum_{S \in M} \sum_{i \in S} c_{i,S} x_{i,S} \quad \text{subject to } x_{i,S} = 1 \text{ if } i \text{ is chosen from } S; \ 0 \text{ o.w.}
\]

# variables \( O(n|E|) \)

# constraints \( O(n|E|) \)

\[
\begin{align*}
z_{r,A,B} &\geq x_{r,A} \quad \forall r \in A \cap B \quad \forall \{A, B\} \in E \\
z_{r,A,B} &\geq x_{r,B} \quad \forall r \in A \cap B \quad \forall \{A, B\} \in E \\
\sum_{r \in A \cap B} z_{r,A,B} & = 1 \quad \forall \{A, B\} \in E \\
\sum_{i \in A} x_{i,A} & = 1 \quad \forall A \\
x, z &\geq 0
\end{align*}
\]

Normalization and non-negativity

Theorem:
If the underlying choice graph is a tree, then the above LP is integral
Set collections with tree choice graph: nested, laminar, and differentiated collections

Nested collection: arises in revenue management  choice graph: line graph
Set collections with tree choice graph: nested, laminar, and differentiated collections

Laminar collection: choice graph: tree

arises through elimination-by-aspects (EBA)


- Morning (AM) Flights
  - AM Non-Stop Flights
    - United UA8328
    - Air Canada AC786
  - AM Flights with 1 Stop
    - Delta DL1154 /4956
    - Air Canada AC782 /421

- Afternoon (PM) Flights
  - PM Non-Stop Flights
    - United UA8022
    - Air Canada AC790
  - PM Flights with 1 Stop
    - Delta DL1833 /3706
    - Air Canada AC774 /405

Features

- L1: Departure Time
- L2: Number of Connections
- L3: Airline

Itinerary #1  Itinerary #2  Itinerary #3  Itinerary #4  Itinerary #5  Itinerary #6  Itinerary #7  Itinerary #8
Set collections with tree choice graph: nested, laminar, and differentiated collections

- **Differentiated collection:**
  - core brands offered in all stores
  - unique brands catering to local tastes

- **Star-shaped choice graph**

soup brands carried at different store locations

- Detroit
- Healthy-choice Campbell Progresso
- Indianapolis
- Private-label
- Detroit
- St. Louis
Computational complexity for general choice graphs: scales exponentially in \( \min(\text{tree-width}, \text{choice-depth}) \)

**Theorem:**
Rank aggregation solved (using DP and LP) in \( O(n^{\min(CD, \text{tw})}) \)

**tree-width (tw)** [equivalent definitions]
- size of largest vertex set in tree-decomposition of graph
- size of largest clique in the chordal completion of graph

**choice-depth (CD):** “spread” of subsets in the same bag of tree-decomposition

\[
\max_b \left\{ \left| \bigcup_{S \in \mathcal{X}_b} S \right| - \min_{A \in \mathcal{X}_b} |A| \right\}
\]

**Theorem:**
Rank aggregation solved (using DP and LP) in \( O(n^{\min(CD, \text{tw})}) \)
**Theorem:**
choice-depth $\leq k \leq n \leq$ tree-width for k-del choice graph. Complexity = $O(n^k)$
1. Case study with IRI Academic Dataset
   what is the ‘right’ choice model?

2. Complexity of computing LoR
   rational separation, complexity in terms of graph properties

3. Summary/Conclusions
   takeaway messages
Summary and Key Findings

KEY CONTRIBUTIONS

‣ characterization of complexity of computing LoR
‣ approx. techniques to compute LoR efficiently

MAIN TAKEAWAYS

Computational complexity of rank-aggregation

exponential in tree-width/choice-depth of the choice graph

Rationality loss

‣ larger than parametric loss
‣ suggests the need to go beyond rationality

FUTURE DIRECTIONS

‣ relate complexity lower bounds to tree-width/choice-depth
‣ general constructions of choice graphs
‣ general model classes that go beyond rationality