Overview: singleton vs. subset choice

Given a set of alternatives to choose from, how do people choose?

- If choosing just one thing (buying a car, picking a restaurant, etc.), there are lots of good ML techniques (logistic regression, your favorite deep net, etc.)
- If choosing a subset of the alternatives (what to buy after browsing Amazon, constructing a playlist on Spotify, etc.), there aren’t as many tools.

We provide an interpretable and computationally feasible model for subset selection based on random utility maximization.

Basic concept of the model

You are throwing a small party and want to provide some snacks. Large set of snack options and want to choose a couple.

- tortilla chips, potato chips, cookies, pretzels, guacamole, celery, nut mix, hummus, meatballs, cupcakes, pigs in blankets, cupcakes, potato skins, chicken wings, taquitos, pineapple, …

Model 1 (Separable model).
- Independent choices.
- Easy computation, but not realistic.

Model 2 (Full Model).
- All subsets as options.
- Harder computation, but more realistic.

Discrete choice model for subset selection

A person makes a selection based on random utility \(U_i\) of sets \(i, j\).

\[
U_i = \{V_i + V_j + e_i \mid \{i, j\} \notin H, \{i, j\} \in H^i\}
\]

Base item utilities

Corrective utility of subset

The \(e_i\) are i.i.d. errors (per person, per choice) sampled from a Gumbel distribution.

A “rational agent” that chooses the set with largest utility chooses \(\{i, j\}\) from a set of alternatives \(C\) containing \(i\) and \(j\) with probability

\[
\prod_{\{i,j\} \subseteq C} P_{ij} = \begin{cases} \prod_p \prod_i p_i & \{i,j\} \notin H, \\ \prod_i & \{i,j\} \in H, \end{cases}
\]

\[
\sum_{\{i,j\} \subseteq C} P_{ij} = 1, \quad p_i \geq 0, \quad q_i \geq 0, \quad \sum_i p_i = 1
\]

Generalizing to larger sets.

A person makes a selection based on random utility \(U_k\) of sets \(i, j, k\).

Key concept.
- Base item utilities (the \(V_i\) are the same regardless of size of set.

\[
U_k = \{V_i + V_j + V_k + e_{ijk} \mid \{i, j, k\} \notin H, \{i, j, k\} \in H\}
\]

Suppose that \(H = \{i, j, k\}\), then

\[
\text{Pr} (\text{choose } \{i,j\} \text{ size-2 choice}) = \sum_k \Pr (\text{select } S \text{ | size-2 choice}) \times \gamma_j \times p_i \times q_j
\]

Putting everything together.

Use a mixture model and condition on size of selected set.

\[
\text{Pr} (\text{select } S \text{ | alternatives } C) = \sum_{k=1}^n \Pr (\text{select } S \text{ | size-k selection})
\]

\[
2k \geq 0, \quad k = 1, \ldots, n, \quad \sum_{k=1}^n 2k = 1, \quad n = \text{size of largest choice set}
\]

Observation. Likelihood of \(z_{ijk}\) is concave with a linear constraints \(\rightarrow\) easy to learn.

Acknowledgements.
This research was partially completed while ARB was visiting Google.
ARB also supported in part by a Simons Investigator Award.

Findings with “universal choice datasets”

Universal choice datasets: the set of available alternatives is always the same.

- Bakery. Sets of items purchased at a bakery.
- WalmartItems. Sets of items bought at Walmart.
- WalmartDepts. Sets of departments from which items were purchased at Walmart.
- Kosarak. Sets of hyperlinks visited during a session on a Hungarian news portal.
- Instacart. Sets of items in ln.
- LastfmGenres. Sets of genres of music listened to in a listening session on Last.fm.

Learning model parameters.

Theorem. Given a budget constraint on the number of special subsets (size of \(H\)), it is NP-hard to find the set which will maximize likelihood (and it is also not a submodular optimization problem).

Theorem. Given \(H\), there is a closed form for the model parameters that maximize likelihood.

A person makes a selection based on random utility

\[
\text{if choosing a subset of the alternatives}
\]

\[
\text{alternatives}
\]

\[
\text{a rational agent that chooses the set with largest utility}
\]

\[
\text{chooses }
\]

\[
\text{easy computation, but}
\]

\[
\text{based on random utility maximization.}
\]

\[
\frac{1}{\gamma_j}
\]

\[
\text{a person makes a selection based on random utility}
\]

\[
\text{if choosing a subset of the alternatives}
\]

\[
\text{alternatives}
\]

\[
\text{a rational agent that chooses the set with largest utility}
\]

\[
\text{chooses }
\]

\[
\text{easy computation, but}
\]

\[
\text{based on random utility maximization.}
\]

\[
\frac{1}{\gamma_j}
\]

\[
\text{a person makes a selection based on random utility}
\]

\[
\text{if choosing a subset of the alternatives}
\]

\[
\text{alternatives}
\]

\[
\text{a rational agent that chooses the set with largest utility}
\]

\[
\text{chooses }
\]

\[
\text{easy computation, but}
\]

\[
\text{based on random utility maximization.}
\]

\[
\frac{1}{\gamma_j}
\]

\[
\text{a person makes a selection based on random utility}
\]

\[
\text{if choosing a subset of the alternatives}
\]

\[
\text{alternatives}
\]

\[
\text{a rational agent that chooses the set with largest utility}
\]

\[
\text{chooses }
\]

\[
\text{easy computation, but}
\]

\[
\text{based on random utility maximization.}
\]

\[
\frac{1}{\gamma_j}
\]

\[
\text{a person makes a selection based on random utility}
\]

\[
\text{if choosing a subset of the alternatives}
\]

\[
\text{alternatives}
\]

\[
\text{a rational agent that chooses the set with largest utility}
\]

\[
\text{chooses }
\]

\[
\text{easy computation, but}
\]

\[
\text{based on random utility maximization.}
\]

\[
\frac{1}{\gamma_j}
\]

\[
\text{a person makes a selection based on random utility}
\]

\[
\text{if choosing a subset of the alternatives}
\]

\[
\text{alternatives}
\]

\[
\text{a rational agent that chooses the set with largest utility}
\]

\[
\text{chooses }
\]

\[
\text{easy computation, but}
\]

\[
\text{based on random utility maximization.}
\]

\[
\frac{1}{\gamma_j}
\]

\[
\text{a person makes a selection based on random utility}
\]

\[
\text{if choosing a subset of the alternatives}
\]

\[
\text{alternatives}
\]

\[
\text{a rational agent that chooses the set with largest utility}
\]

\[
\text{chooses }
\]

\[
\text{easy computation, but}
\]

\[
\text{based on random utility maximization.}
\]

\[
\frac{1}{\gamma_j}
\]