Applying Decision Theory to Discrete Choice Model

Kota Saito

Caltech

January 16, 2022

Application: Who will Use a New Train Line?

Bay Area Rapid Transit opened in 1973

https://www.youtube.com/watch?v=O_eCgv6j3-s

BART was the epicenter of modern structural modeling

Application: Who will Use a New Train Line?

Before BART opened, D. McFadden (then a prof at UC Berkeley) was asked to predict BART ridership.

Q: How to predict BART ridership with data prior to BART opening?

Requirements for a Counterfactual Prediction Method

McFadden'd like to develop a general method that

1. predicts choice behavior in counterfactual settings
   $\rightarrow$ needs a "theory" of how choices change in counterfactuals

2. can be used only with limited real-world data
   - each choice-maker makes only a small number of choices
   - characteristics of alternatives & individuals are only incompletely known

3. is computationally practical (even in 1970s)

Prediction by Revealed Preference over Characteristics 1/2

The Data:

- Information about characteristic of each transportation method such as price, speed, etc.
  - Let K be the number of characteristic variables.
  - Each transportation method corresponds to an element of $\Re^K$.
  - For example, $x_{car} = (\text{price of car}, \text{speed of car}, etc) \in \Re^K$.
  - Let $X \subset \Re^K$ be the set of all alternatives (represented as characteristic vectors)
- Consumer's choice data before BART is open:
  - How many percentage of people use each of transportation methods

Random utility: Assumption of individual's behavior

Model:

- Assume utility of alternative i is

$$u(i) = \beta \cdot x(i) + \varepsilon(i),$$

  where β captures individual's preference unknown to researcher; ε is random component of utilities unknown to researcher.
- This is called *random utility model*.
- Each individual chooses an alternative i from choice set C if $u(i) > u(j)$ for all $j \in C \setminus \{i\}$
- Market share of alternative i:

$$\rho(C, i) = \text{Prob}(\beta \cdot x(i) + \epsilon(i) > \beta \cdot x(j) + \epsilon(j) \text{ for all } j \in C \setminus \{i\})$$

Canonical Example: Logit

Definition 1 (Type I Extreme Value (Gumbel) Distribution)

$$\text{Cumulative Density Function } F(\epsilon) \equiv e^{-e^{-\epsilon}}$$

Logit Assumption

$$\epsilon(i) \sim_{iid} \text{ Type I Extreme Value (Gumbel) Distribution } F(\epsilon)$$

Theorem 2 (Logit Choice Probabilities)

$$\textit{Market share of alternative } i{:}\rho(C,i) = \frac{\exp(\beta \cdot x(i))}{\sum_{j \in C} \exp(\beta \cdot x(j))}$$

Estimation

- Estimate β by using maximum likelihood
- Add Bart to choice set C by specifying

$$x(bart) = (\text{price of bart}, \text{speed of bart}, etc) \in \Re^K.$$

- Calculate

$$\text{Market share of bart} : \rho(C, \text{bart}) = \frac{\exp(\beta \cdot x(i))}{\sum_{j \in C \cup \{\text{bart}\}} \exp(\beta \cdot x(j))}$$

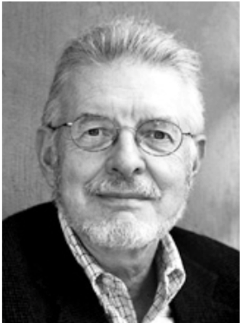
Result

Table 3.2. *Predictions for after BART opened*

	Actual Share	Predicted Share
Auto alone	59.90	55.84
Bus with walk access	10.78	12.51
Bus with auto access	1.426	2.411
BART with bus access	0.951	1.053
BART with auto access	5.230	5.286
Carpool	21.71	22.89

Ex ante predictions match ex post reality well!

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2000



Result

- The same methods has been now used by policymakers & researchers.
- Researchers use parametric discrete choice models to describe choice behavior by
  - consumers over products
  - students over schools
  - patients over health providers

Underling theoretical model: Random Utility Model

- X: a finite subset of $\mathbb{R}^K$ (Set of all products)
  - K: the number of explanatory variables.
- $\mathcal{D} \subset 2^X \setminus \{\emptyset\}$ (Set of choice sets)
- A *stochastic choice* ρ is a mapping on $\mathcal{D}$ such that for any menu $D \in \mathcal{D}$, $\rho(D, \cdot)$ is a probability distribution over D.
  - $\rho(D, x)$ is the probability that x is chosen from D. (market share)

Random Utility Model

- Π: the set of rankings (i.e., strict preference relations) on X.
  - $|\Pi| = |X|!$
- If $\pi(x) > \pi(y)$, then x is better than y with respect to π.
- For any ranking $\pi \in \Pi$, define a stochastic choice function ρ^π by

$$\rho^\pi(D, x) = \begin{cases} 1 & \text{if } \pi(x) \geq \pi(y) \text{ for all } y \in D, \\ 0 & \text{otherwise.} \end{cases}$$

Random Utility Model

Definition 3

A stochastic choice function ρ is called a *random utility function* if there exists a probability measure $\nu \in \Delta(\Pi)$ s.t.

$$\rho(D, x) = \nu(\pi \in \Pi | \pi(x) \geq \pi(y) \text{ for all } y \in D).$$

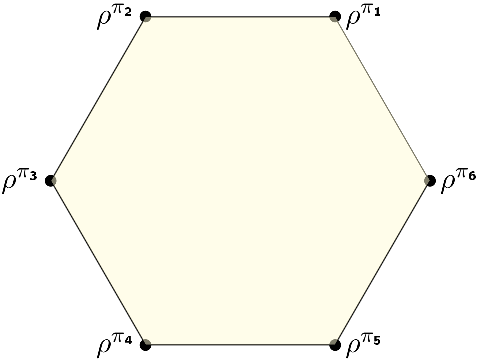
The set of random utility functions is denoted by $\mathcal{P}_r$.

- Notice that for any random utility model $\rho \in \mathcal{P}_r$,

$$\rho = \sum_{\pi \in \Pi} \nu(\pi)\rho^{\pi} \in \text{co.}\{\rho^{\pi} | \pi \in \Pi\}.$$

Geometry

- The set $\mathcal{P}_r$ is a polytope co.$\{\rho^{\pi} | \pi \in \Pi\}$!



Question (to you)

- What is the geometric structure of $\mathcal{P}_r$?
- The question is mathematically interesting and important to understand human behavior.

Why this geometric insight is useful?

- Random utility model is standard model to describe true human behavior but researchers often use more parametric model such as logit model.
- When can we represent random utility model by using a generalized version of logit model?

Logit Model

Definition 4

- A stochastic choice function ρ is called a *logit function* if there exists a vector β and $\eta \in \Re^X$ s.t.

$$\rho(D, x) = \frac{\exp(\beta \cdot x + \eta_x)}{\sum_{y \in D} \exp(\beta \cdot y + \eta_y)}.$$

- The set of logit functions of degree d with fixed effects η is denoted by $\mathcal{P}_l(\eta)$.

Mixed Logit Model

Definition 5

- A stochastic choice function ρ is called a *mixed logit function* if there exists a probability measure m and $\eta \in \Re^X$ s.t.

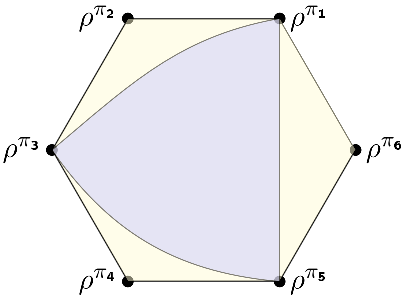
$$\rho(D, x) = \int \frac{\exp(\beta \cdot x + \eta_x)}{\sum_{y \in D} \exp(\beta \cdot y + \eta_y)} dm(\beta).$$

  The mixture is over β not over η.

- The set of mixed logit models with fixed effects η is denoted by $\mathcal{P}_{ml}(\eta)$.

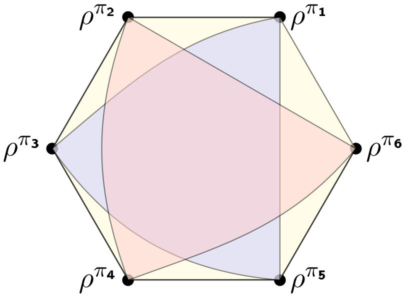
Geometry

- For each fixed effect η, the set $\mathcal{P}_{ml}(\eta)$ of mixed logit model with fixed effect η is a convex subset of the polytope $\mathcal{P}_r$.
  - A logit model is a random utility model
  - Integral $\approx$ convex combination



Geometry

- For different fixed effect η, $\mathcal{P}_{ml}(\eta)$ can be different.
- Q1: Under what condition does $\bigcup_{\eta\in\Re^{x}} \mathcal{P}_{ml}(\eta)$ cover $\mathcal{P}_r$?
  - Notice that $\bigcup_{\eta\in\Re^{x}} \mathcal{P}_{ml}(\eta)$ may not be convex.
- Q2: When the condition is not satisfied, how large can the approximation error be?



Result

X is affinely independent if and only if any random utility model can be approximated by a mixed logit model.

Coauthors

- Yusuke Narita (Yale)
- Haoge Chang (Yale)
- Kono Haruki (MIT)
- Jay Lu (UCLA)

Question (to you)

What is the geometric structure of $\mathcal{P}_r$?

- Adjacency of vertices
- Adjacency of facets