April 30, 2019

Discrete choice

behavioral modelling of how people choose things

Have a choice set

\( \Rightarrow \) options that are available

1. mutually exclusive
can only choose one thing

2. exhaustive
something will be selected
(3) finite

1+2 not restrictive

choice set = \{ A, B \}
\{ A, B, A and B \}
\{ A, B, A and B, neither \}

Examples:
- Transportation to campus
  \{ walk, drive, bus, Uber, ... \}
- Purchase at Gimme
  \{ latte, espresso, muffin, ... \}
Large class of models for Discrete Choice:

Random Utility Models

Universe of items $I$, $|I|<\infty$

chooser $\neq k$

- has choice set $C_k \subseteq I$
- observe utilities
  $$U_{kj}, \ j \in C_k$$
  $\rightarrow$ random variables
- select $\text{arg max}_j U_{kj}$
\[ U_{kj} = V_{kj} + \epsilon_{kj} \]

constant \hspace{1cm} random errors

\[
\text{Prob(chooser } k \text{ selects } j \in (k)) = \text{Prob}(U_{kj} > U_{kl}, j \neq l, l \in (k)) = \text{Prob}(V_{kj} + \epsilon_{kj} > V_{kl} + \epsilon_{kl}, j \neq l) = \text{Prob}(V_{kj} - V_{kl} > \epsilon_{kl} - \epsilon_{kj}, j \neq l)
\]

Get different choice models with different error distributions.

Usually, assume functional form
for $V_{kj}$

$$V_{kj} = V(x_k, y_j) = V(\Theta)$$

Typical setup

Data: $(C_i, s_i), \ldots, (C_N, s_N)$

$s_i \in C_i$

Numerics: we would like to learn $\Theta$

Simple model: logit

$$u_{kj} = V_{kj} + \varepsilon_{kj}$$
\[ e_{kj} \sim \text{Gumble}(0, 1) \]
\[ \text{i.i.d.} \]
\[ p(x) = e^{-xe^x} \]

**Theorem:**

\[ \text{Prob(} k \text{ chooses } j) \]
\[ = \text{Prob(} U_{k,j} > U_{k,l}, \ l \neq j) \]
\[ = \frac{\exp(V_{k,j})}{\sum_{l \in \mathcal{R}} \exp(V_{k,l})} \]

Examples of logit
Bradley-Terry-Luce model for pairwise comparisons

\[ \text{Prob}(i > j) = \frac{p_i}{p_i + p_j} \]

\[ V_i = \log(p_i) \]
\[ V_j = \log(p_j) \]

\[ \text{Prob}(i \geq j) = \frac{\exp(V_i)}{\exp(V_i) + \exp(V_j)} \]

choice sets always size 2
no parameterization of \( V_i \)
ELO Ratings
(chess, football, video games...)

Players A, B

Current ratings $R_A$, $R_B$

1 point for win
$rac{1}{2}$ point for draw
0 point for loss

Score

$S_A, S_B$

$$E_A = \frac{10^{R_A/400}}{10^{R_A/400} + 10^{R_B/400}}$$

$$E_B = 1 - E_A$$
\[ 10^{R_A/400} = \exp(V_A) \]
\[ 10^{R_B/400} = \exp(V_B) \]

Choice set \( \{A, B, \text{draw}\} \)

selection determines score after match, update scores:

\[ R_A \leftarrow R_A + K(S_A - E_A) \]

constant (16 or 32)
Multinomial Logistic Regression

\[ V_{kj} = \beta^T x_j \]

\[ C_k = C \]

\[ \text{Prob(class } j) = \frac{\exp(\beta^T x_j)}{\sum_{\ell \in C} \exp(\beta^T x_\ell)} \]

Nice property: log likelihood function is concave
Observe \((C_1, s_1), \ldots, (C_N, s_N) = D\)

\[
\mathbb{L}(\beta; D) = \sum_{j=1}^{N} \log \left( p(C_j | s_j) \right)
\]

\[
= \sum_{j=1}^{N} \beta^{T} x_{s_j} - \log \left( \sum_{\text{lec}_j} \exp (\beta^{T} x_{s_j}) \right)
\]

linear

log-sum-exp

convex

Can also do things like negative sampling when choice sets are large
Independence of Irrelevant Alternatives (IIA)

choice set $C$

\[
\frac{\Pr(\text{choose } j \in C)}{\Pr(\text{choose } i \in C)} = \frac{\exp(V_{ij})}{\sum_{c \in C} \exp(V_{ic})} = \frac{\exp(V_{ij})}{\exp(V_{ij})} = \exp(V_{ij} - V_{ij})
\]

independent of $C$
if \( C' = C \cup \{x\} \)

then relative probs remain same

Example: Gimme

\( \frac{1}{3} \) black coffee

\( \frac{1}{3} \) latte

\( \frac{1}{3} \) espresso

introduce new maple latte (really good)

\( \frac{1}{2} \) maple latte

if IIA holds...

\( \frac{1}{6} \) black coffee

\( \frac{1}{6} \) latte

\( \frac{1}{6} \) espresso

\} \quad \text{relative probs same}
Maybe real probabilities

\[
\begin{align*}
&\frac{1}{2} \text{ maple latte} \\
&\frac{1}{10} \text{ latte} \\
&\frac{2}{10} \text{ black coffee} \\
&\frac{2}{10} \text{ espresso}
\end{align*}
\]

\{ \text{ violates IIA} \}

\[
\begin{array}{c}
6/10 \\
5/6 \\
\text{maple latte} \\
1/6 \\
\text{latte} \\
4/10 \\
\text{bl} \\
1/2 \\
\text{BC} \\
1/2 \\
\text{espresso}
\end{array}
\]

"mixed logit"
r latent mixture components
choose component i w.p. \( \pi_i \)

\[
\text{Prob(chooser k selects } j \in C_k) = \sum_{i=1}^r \pi_i \left[ \frac{\exp(\beta_i^T x_j)}{\sum_{k \in C_j} \exp(\beta_k^T x_j)} \right]
\]

**single logit**

RUM interpretation

\( U_{kj} = \beta_k^T x_j + \epsilon_{kj} \)
\( \epsilon_{kj} \text{ iid } \text{Gumbel} \)
Chooser is sampled from distribution \( \{ \pi_1, \ldots, \pi_r \} \)

Universality of mixed logit RUM

\[ U_{kj} = V_{kj} + \varepsilon_{kj} \]

Theorem:

Any RUM can be approx. to any degree of accuracy by a mixed logit (possibly large # of mixture components)
Probit model

\[ U_{kj} = V_{kj} + \varepsilon_{kj} \]

\[ C_k = C \]

\[ \varepsilon_k = \begin{bmatrix} \varepsilon_{k1} \\ \vdots \\ \varepsilon_{km} \end{bmatrix}, \quad m = |C| \]

\[ \varepsilon_k \sim N(0, \Sigma) \]

No closed form solutions for item selection probability \[ \Rightarrow \text{estimate numerically} \]

errors can be correlated \[ \Rightarrow \text{IIA not necessary} \]
Discrete choice
choosers 1, 2, ..., N
Chooser # k
- has choice set \( C_k \subseteq I \)
- Observe utilities
  \( U_{kj}, j \in C_k \)
  (random variables)
- select arg max \( U_{kj} \)

Next time: frame network growth as discrete choice