April 30, 2019

Discrete choice

behavioral modelling of how people choose things

Have a choice set
↳ 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 = $\{A, B\}$

$\{A, B, \text{A and B}\}$

$\{A, B, \text{A and B, neither}\}$

Examples:

- Transportation to campus

  {walk, drive, bus, Uber, ...}

- Purchase at Gimme

  {latte, espresso, muffin, ...}

Large class of models for Discrete Choice:

<u>Random Utility Models</u>

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

chooser #k

- has choice set $C_k \subseteq I$
- observe utilities

  U_{kj}, $j \in C_k$

  ↳ random variables
- select $\arg\max_j U_{kj}$

$$U_{kj} = \underbrace{V_{kj}}_{\text{constant}} + \underbrace{\varepsilon_{kj}}_{\text{random errors}}$$

Prob(chooser k selects $j \in C_k$)

$= \text{Prob}(U_{kj} > U_{kl},\ j \neq l,\ l \in C_k)$

$= \text{Prob}(V_{kj} + \varepsilon_{kj} > V_{kl} + \varepsilon_{kl},\ j \neq l)$

$= \text{Prob}(V_{kj} - V_{kl} > \varepsilon_{kl} - \varepsilon_{kj},\ j \neq l)$

Get different choice models with different error distrib.

Usually, assume functional form

for V_{kj}

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

features of chooser

features of item j

Typical setup

Data: $(C_1, s_1), \ldots, (C_N, s_N)$

$$s_i \in C_i$$

Numerics: we would like to learn θ

Simple model: logit

$$U_{kj} = V_{kj} + \varepsilon_{kj}$$

$$\varepsilon_{kj} \sim \text{Gumble}(0,1)$$
i.i.d.

$$p(x) = e^{-(x+e^{-x})}$$

Theorem:

$$\text{Prob}(k \text{ chooses } j)$$
$$= \text{Prob}(U_{kj} > U_{kl},\ l \neq j)$$
$$= \frac{\exp(V_{kj})}{\sum_{l \in C_k} \exp(V_{kl})}$$

Examples of logit

① Bradley-Terry-Luce model for pairwise comparisons

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

$$V_i = \log(p_i)$$

$$V_j = \log(p_j)$$

$$\text{Prob}(i > j) = \frac{\exp(V_i)}{\exp(V_i) + \exp(V_j)}$$

choice sets always size 2

no parameterization of V_i

② ELO Ratings

(chess, football, video games...)

Players A, B

Current ratings R_A, R_B

$\left.\begin{array}{l} \text{1 point for win} \\ \text{1/2 point for draw} \\ \text{0 point for loss} \end{array}\right\}$ 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)

(3) Multinomial Logistic Regression

$$V_{kj} = \beta^T x_j$$

latent coeffs

features of item j

$$C_k = C$$

$$\text{Prob}(\text{class } j) = \frac{\exp(\beta^T x_j)}{\sum_{l \in C} \exp(\beta^T x_l)}$$

Nice property: log likelihood function is concave

Observe $(C_1, s_1), \ldots, (C_N, s_N) = D$

$$LL(\beta; D) = \sum_{j=1}^{N} \log(p(C_j, s_j))$$

$$= \sum_{j=1}^{N} \underbrace{\beta^T x_{s_j}}_{\text{linear}} - \underbrace{\log\left(\sum_{\ell \in C_j} \exp(\beta^T x_\ell)\right)}_{\substack{\text{log-sum-exp} \\ \text{convex}}}$$

Can also do things like negative sampling when choice sets are large

Independence of Irrelevant Alternatives (IIA)

choice set C

$$\frac{\text{Prob(choose } j \in C)}{\text{Prob(choose } i \in C)}$$

$$= \frac{\exp(V_j)}{\cancel{\sum_{\ell \in C} \exp(V_\ell)}} \Big/ \frac{\exp(V_i)}{\cancel{\sum_{\ell} \exp(V_\ell)}}$$

$$= \frac{\exp(V_j)}{\exp(V_i)} = \exp(V_j - V_i)$$

independent of C

if $C' = C \cup \{x\}$
then relative probs remain same

Example: Gimme

1/3 black coffee
1/3 latte
1/3 espresso

introduce new maple latte
(really good)

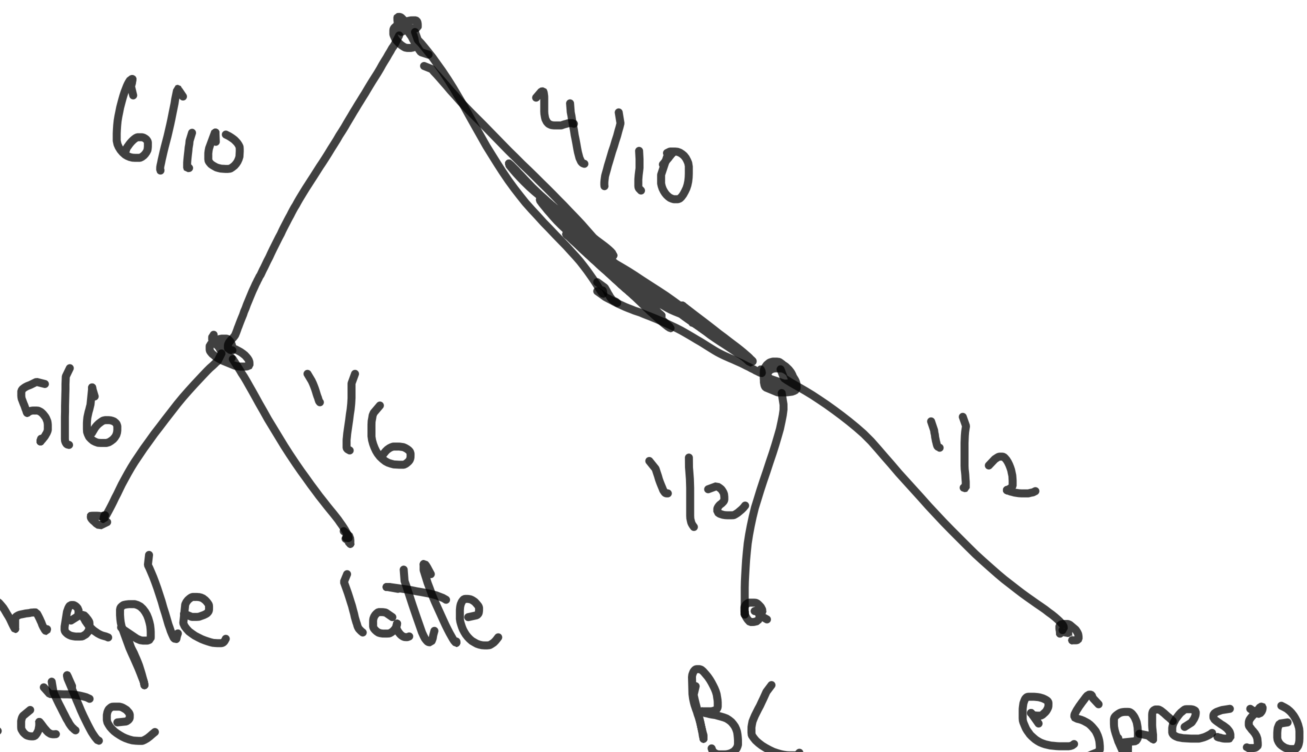
1/2 maple latte
if IIA holds...

$\left.\begin{array}{l}\text{1/6 black coffee} \\ \text{1/6 latte} \\ \text{1/6 espresso}\end{array}\right\}$ relative probs. same

Maybe real probabilities

1/2 maple latte
1/10 latte
2/10 black coffee
2/10 espresso

violates IIA



"mixed logit"

r latent mixture components

choose component i w.p. π_i

$$\text{Prob(chooser } k \text{ selects } j \in C_k) = \sum_{i=1}^{r} \pi_i \underbrace{\left[\frac{\exp(\beta_i^T x_j)}{\sum_{l \in C_j} \exp(\beta_i^T x_l)}\right]}_{\text{single logit}}$$

RUM interpretation

$$U_{kj} = \beta_k^T x_j + \varepsilon_{kj}$$

ε_{kj} iid Gumble

Chooser k sampled from distribution $\{\pi_1, \dots, \pi_r\}$

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

$\Rightarrow$ estimate numerically

errors can be correlated

$\Rightarrow$ IIA not necessary

Discrete choice

choosers $1, 2, \ldots, N$

Chooser # k

- has choice set $C_k \subseteq I$
- Observe utilities

  $U_{kj}, \; j \in C_k$

  (random variables)

- select $\arg\max_j U_{kj}$

Nex time: frame network growth as discrete choice