

40.016: Analytics Edge

Week 4 Lecture 2

AUTOMOBILE SAFETY

Term 6, 2020



SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

Data, code and R library

- Download and save:
 - `safetydata.csv`
 - `automobiles_class.Rmd`
- `install.packages:`
 - `mlogit`
 - `dfidx`

Automobile safety

- Automobile companies (like GM) conduct customer surveys to understand trade-offs of different attributes of a product or a service.
- The company obtains valuations of attributes by using data on preference of safety features.
- Companies can estimate the effect of new products in the market and how to price them.

Discrete choice models

- Choose one among many (mutually exclusive) alternatives.
- Widespread use in econometrics.
- Data observed are often of two basic types:
 - revealed preferences – observed choice of individuals (Oscars).
 - stated preferences – virtual situation, choose one of a few options (Automobile safety).
- We get: – at individual level choice probabilities,
– at an aggregate level market share.
- Huge literature with variety of models: Multinomial logit, Nested logit, GEV, etc. and applications in econometrics, market science, transportation.

Independence of irrelevant alternatives (IIA)

- Property (and shortcoming) of the Conditional logit model (also called *Multinomial logit (MNL) model*).
- The odds of choosing k versus l is:

$$\frac{\Pr(Y_i = k)}{\Pr(Y_i = l)} = \left(\frac{e^{\beta^\top \mathbf{x}_{ik}}}{\sum_{t=1}^K e^{\beta^\top \mathbf{x}_{it}}} \right) \frac{1}{\frac{e^{\beta^\top \mathbf{x}_{il}}}{\sum_{t=1}^K e^{\beta^\top \mathbf{x}_{it}}}} = e^{\beta^\top (\mathbf{x}_{ik} - \mathbf{x}_{il})}$$

Hence:

$$\ln \frac{\Pr(Y_i = k)}{\Pr(Y_i = l)} = \beta^\top (\mathbf{x}_{ik} - \mathbf{x}_{il}).$$

Example: IIA – Blue bus and red bus

- Suppose we have two alternatives: Car and (Red) Bus.
- Suppose your choice follows: $\Pr(C) = \Pr(R) = 0.5$.
- The city adds a (Blue) bus.
- Commuters don't care about the color of the bus.
- New alternatives: Car, Red Bus, Blue Bus
- Your (reasonable) choice: $\Pr(C) = 0.5, \Pr(R) = \Pr(B) = 0.25$.
- MNL prediction: $\Pr(C) = \Pr(R) = \Pr(B) = 0.33$:

This happens because blue bus and red bus are perfect substitutes here and not captured by MNL model.

Mixed logit model

- STANDARD LOGIT: $U_{ik} = \beta^T \mathbf{x}_{ik} + \epsilon_{ik}$
- MIXED LOGIT: $\tilde{\beta}$ is modeled as a random parameter, $U_{ik} = \tilde{\beta}^T \mathbf{x}_{ik} + \epsilon_{ik}$

STANDARD MULTINOMIAL LOGIT	MIXED LOGIT
$\Pr(Y_i = k) = \frac{e^{\beta^T \mathbf{x}_{ik}}}{\sum_{l=1}^K e^{\beta^T \mathbf{x}_{il}}}$	$\Pr(Y_i = k) = \int \frac{e^{\beta^T \mathbf{x}_{ik}}}{\sum_{l=1}^K e^{\beta^T \mathbf{x}_{il}}} f(\beta) \, d\beta$

- For the mixed logit model, $f(\beta)$ is the probability density function of $\tilde{\beta}$. This leads to an integral over logit probabilities.

Mixed logit model

- Mixed logit is computationally more challenging – simulation optimization methods.
- Not a concave maximization problem. Finding a global optimum might not be easy.
- From an estimation perspective, the goal is to find the parameters θ that define the density function $f(\beta|\theta)$ where the functional form $f(\cdot)$ is given but parameters θ are unknown.

Mixed logit: simulation-optimization to estimate θ

- 1 Make an initial hypothesis about the parameter θ .
- 2 Draw R numbers from this distribution, call them $\beta_r, r = 1, \dots, R$ and compute

$$P_{ik}^r = \frac{e^{\beta_r^\top \mathbf{x}_{ik}}}{\sum_{l=1}^K e^{\beta_r^\top \mathbf{x}_{il}}}.$$

- 3 Now we approximate $\Pr(Y_i = k)$ by the average:

$$\Pr(Y_i = k) \approx \bar{P}_{ik} = \frac{1}{R} \sum_{r=1}^R P_{ik}^r.$$

- 4 Compute the (simulated) log-likelihood for the probabilities:

$$SLL = \sum_{i=1}^n \sum_{l=1}^K z_{ik} \log \bar{P}_{ik}$$

where $z_{ik} = 1$ if i th individual chose k th item.

- 5 Repeat the process with a new choice of θ and maximize (optimize).

Mixed logit model: panel data

- For mixed logit with repeated choices (panel data), where i : individual, k : alternative, t : observation, we have:

$$\Pr(Y_{i1} = k_1, \dots, Y_{iT} = k_T) = \int \prod_{t=1}^T \left(\frac{e^{\beta^\top \mathbf{x}_{ik_t t}}}{\sum_{l=1}^K e^{\beta^\top \mathbf{x}_{il t}}} \right) f(\beta) d\beta.$$

- In panel data, we need to account for the fact that the errors are correlated for the same individual over time.

Willingness to pay

- β_1 is the coefficient for attribute 1;
- β_2 is the coefficient of attribute 2 (Price).
- β_2 will be typically negative. Say β_1 is positive.
- Suppose

$$U = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \text{other terms}$$

- For a one unit increase in attribute 1, how much price are we willing to pay (*wtp*) so that the utility remains the same. Suppose this amount is Δ . Then

$$U = \beta_0 + \beta_1(x_1 + 1) + \beta_2(x_2 + \Delta) + \text{other terms}.$$

- Then for 1 unit increase in attribute 1,

$$wtp = \Delta = -\frac{\beta_1}{\beta_2}.$$

Automobile safety: dataset

- 500 consumers.
- Each consumer performed 19 choice tasks.
- Each attribute has from 2 to 11 levels.
- For each task, three alternatives provide different alternatives on attributes, the fourth option is an outside option or “no choice”.
- Certain customer demographic information such as age, education, income, etc. are also available (not used in our model).
- For training our model we use 12 of these tasks and the rest we use for the test set.

Automobile safety: attributes

Table 1 Attribute and Level Codes

Serial no.	Attribute name	Attribute code	No. of levels	Level codes
1	Cruise control	CC	3	CC1, CC2, CC3
2	Go notifier	GN	2	GN1,GN2
3	Navigation system	NS	5	NS1, NS2, NS3, NS4, NS5
4	Backup aids	BU	6	BU1, BU2, BU3, BU4, BU5, BU6
5	Front park assist	FA	2	FA1, FA2
6	Lane departure	LD	3	LD1, LD2, LD3
7	Blind zone alert	BZ	3	BZ1, BZ2, BZ3
8	Front collision warning	FC	2	FC1, FC2
9	Front collision protection	FP	4	FP1, FP2, FP3, FP4
10	Rear collision protection	RP	2	RP1, RP2
11	Parallel park aids	PP	3	PP1, PP2, PP3
12	Knee air bags	KA	2	KA1, KA2
13	Side air bags	SC	4	SC1, SC2, SC3, SC4
14	Emergency notification	TS	3	TS1, TS2, TS3
15	Night vision system	NV	3	NV1, NV2, NV3
16	Driver assisted adjustments	MA	4	MA1, MA2, MA3, MA4
17	Low speed braking assist	LB	4	LB1, LB2, LB3, LB4
18	Adaptive front lighting	AF	3	AF1, AF2, AF3
19	Head up display	HU	2	HU1, HU2
20	Price	Price	11	\$500, 1,000, 1,500, 2,000, 2,500, 3,000, 4,000, 5,000, 7,500, 10,000, 12,000

Table: Choice alternatives provided

Alternative #1 (Ch1)	Alternative #2 (Ch2)	Alternative #3 (Ch3)	Alternative #4 (Ch4)
GN 1	GN 2	GN 2	0
NS 4	NS 1	NS 4	0
BU 6	BU 2	BU 5	0
FP 3	FP 2	FP 3	0
RP 1	RP 1	RP 3	0
PP 1	PP 1	PP 3	0
TS 1	TS 1	TS 3	0
NV 1	NV 1	NV 1	0
MA 4	MA 1	MA 4	0
Price 2	Price 2	Price 2	0

Automobile safety: survey

Which of the following packages would you prefer the most?

Choose by clicking one of the buttons below

Full speed range adaptive cruise control - Traditional back up aid Lane departure warning Front collision warning Side head air bags Emergency notification with pictures - - Option package price: \$3,000	Adaptive cruise control Navigation system with curve notification & speed advisor Rear vision system - - Side body air bags Emergency notification - Head up display Option package price: \$500	Full speed range adaptive cruise control Traditional navigation system - - Front collision warning Side body & head air bags - Night vision with pedestrian detection Head up display Option package price: \$12,000	None: I wouldn't purchase any of these packages
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>