37902 Foundation of Advanced Quantitative Marketing

Session I (Fall 2024)

Objective:

1. Brief Overview of Quantitative Modeling in Marketing
2. Brand Choice Models - the Logit Model

1 Brand Choice Models

1.1 Choice models

- Each brand j has some utility for the consumer, U_j, $j = 1, 2, ..., J$.
- Choice rule: Consumer chooses the brand that maximizes U_j, i.e., $U_j^\star = U_j$ iff $U_j > U_k$, $\forall k \neq j$.

  So if there are 2 brands 1 and 2 with utility U_1 and U_2. Then consumer chooses the brand with the higher utility.

- Now as researchers, if we observe U_1 and U_2, we are done! We can perfectly predict the consumer's choice. Sadly we don't but we observe some of the factors that influence choice.

  Now we can decompose the utilities as follows:

  - $U_1 = V_1 + \epsilon_1$
  - $U_2 = V_2 + \epsilon_2$,
  - where V_1 and V_2 are the "observable" components of utility, and ϵ_1 and ϵ_2 are "unobservable" or "random" components of utility.

  Note: for the consumer both V and ϵ are observable so the decision rule from the consumer perspective is fully deterministic (for now). Later in the quarter, we will relax this assumption. So the RUM is a reference to the researcher and what is random from the researcher's perspective.

- Since the researcher does not observe ϵ, (s)he has to make assumption about it. In particular we need to make some distributional assumption.

  Conditional on the assumption on ϵ, I can at least say something about the probability with which the consumer will choose a brand even if I cannot make a statement with certainty about the choice.

  Typical assumption on ϵ

  - $\epsilon \sim$ Normal (Probit model)
  - $\epsilon \sim$ Type I Extreme Value (Logit model)

1.2 Logit model

- $f(\epsilon) = e^{-\epsilon} e^{-e^{-\epsilon}}$
- $F(\epsilon) = e^{-e^{-\epsilon}}$

  —ϵ's for brands are iid (independent and identically distributed).

- $\begin{cases} U_1 = V_1 + \epsilon_1 \\ U_2 = V_2 + \epsilon_2 \end{cases}$
- As researchers, what is the probability that consumers choose brand 1 assuming we know V_1 and V_2?

$$\begin{aligned} P(Brand\ 1\ is\ chosen) =& P(U_1 > U_2) \\ =& P(V_1 + \epsilon_1 > V_2 + \epsilon_2) \\ =& P(\epsilon_2 < (V_1 - V_2) + \epsilon_1) \end{aligned}$$

To characterize the above

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{(V_1-V_2)+\epsilon_1} f(\epsilon_1, \epsilon_2) d\epsilon_2 d\epsilon_1$$

Given the iid assumption

$$\begin{aligned}
&= \int_{-\infty}^{\infty} f(\epsilon_1) \underbrace{\left[\int_{-\infty}^{(V_1-V_2)+\epsilon_1} f(\epsilon_2) d\epsilon_2 \right]}_{\text{CDF Up to } (V_1-V_2)+\epsilon_1} d\epsilon_1 \\
&= \int_{-\infty}^{\infty} e^{-\epsilon_1} e^{-e^{-\epsilon_1}} e^{-e^{-(V_1-V_2)-\epsilon_1}} d\epsilon_1 \\
&= \int_{-\infty}^{\infty} e^{-\epsilon_1} e^{-e^{-\epsilon_1}} \cdot e^{-e^{-(V_1-V_2)} \cdot e^{-\epsilon_1}} d\epsilon_1
\end{aligned}$$

$$\text{Define } K = e^{-(V_1-V_2)}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} e^{-\epsilon_1} \underbrace{e^{-e^{-\epsilon_1}} e^{-Ke^{-\epsilon_1}}}_{\text{combine terms}} d\epsilon_1 \\
&= \int_{-\infty}^{\infty} e^{-\epsilon_1} e^{-(1+K)e^{-\epsilon_1}} d\epsilon_1 \\
&= \frac{1}{1+K} \int_{-\infty}^{\infty} (1+K) e^{-\epsilon_1} \cdot e^{-(1+K)e^{-\epsilon_1}} d\epsilon_1 \\
&= \frac{1}{1+K} \int_{-\infty}^{\infty} e^{-(\epsilon_1 - ln(1+K))} \cdot e^{-e^{-(\epsilon_1 - ln(1+K))}} d\epsilon_1
\end{aligned}$$

$$\text{Define } \tilde{\epsilon_1} = \epsilon_1 - ln(1+K)$$

$$\begin{aligned}
&= \frac{1}{1+K} \underbrace{\int_{-\infty}^{\infty} e^{-\tilde{\epsilon_1}} e^{-e^{-\tilde{\epsilon_1}}} d\tilde{\epsilon_1}}_{\int_{-\infty}^{\infty} f(\tilde{\epsilon_1}) d\tilde{\epsilon_1} = 1} \\
&= \frac{1}{1+K} \\
&= \frac{1}{1+e^{-(V_1-V_2)}} \\
&= \frac{e^{V_1}}{e^{V_1}+e^{V_2}}
\end{aligned}$$

- In general probability P_{ijt},

$$P_{ijt} = \frac{e^{V_{ijt}}}{\sum_{k=1}^{J} e^{V_{ikt}}} \text{ which is the logit model}$$

So, $0 < P_{ijt} < 1$ and $\sum_{j=1}^{J} P_{ijt} = 1$ satisfy the properties of probabilities.

- To make further progress, we need to characterize V_{ijt}. Usual assumption is:

$$V_{ijt} = \underbrace{\alpha_{ij}}_{\text{Instrinsic preference that consumer } i \text{ has for brand } j} + X_{jt} \cdot \underbrace{\beta_i}_{\text{Consumer } i\text{'s sensitivities to } X_{jt}}$$

X_{jt}: set of characteristics or features of j - prices, promotion, etc.

So,

$$P_{ijt} = \frac{e^{\alpha_{ij} + X_{jt} \cdot \beta_i}}{\sum_{k=1}^{J} e^{\alpha_{ik} + X_{kt} \cdot \beta_i}}$$

Initially, we will assume no heterogeneities in preference and sensitivities:

$$\alpha_{ij} = \alpha_j \qquad \forall\, i$$
$$\beta_i = \beta \qquad \forall\, i$$
$$P_{ijt} = \frac{e^{\alpha_j + X_{jt}\beta}}{\sum_{k=1}^{J} e^{\alpha_k + X_{kt}\beta}}$$

1.2.1 Estimating the logit model

What is observed

ID	Brand 1	Brand 2	Brand 3	Brand 4
1	0	0	0	1
1	0	1	0	0
1	0	1	0	0
2	1	0	0	0
2	1	0	0	0
2	0	1	0	0

What the model predicts

ID	P(Brand 1)	P(Brand 2)	P(Brand 3)	P(Brand 4)
1	P_{11}	P_{21}	P_{31}	P_{41}
1	P_{12}	P_{22}	P_{32}	P_{42}
1	...	...	...	...
2				
2				
2				

- Focusing on the first observation, we see that the consumer bought brand 4.
- What is the estimation task? To figure out $\{\alpha_1,\, \alpha_2,\, \alpha_3,\, \alpha_{4,}\, \beta_f,\, \beta_p\}$. There are two variables, feature and price in the data.
- For identification, we need to set one of the $\alpha's = 0$, Why? Adding or subtracting the same number from $V_{ijt}, \forall\, j$ will leave the choice probability unchanged, so there are 5 parameters to estimate $\{\alpha_1, \alpha_2, \alpha_3, \beta_f, \beta_p\}$ if we set $\alpha_4 = 0$.
- To choose the parameters, we want to do them in such a way that the model's predictions match up as closely as possible to the observed data. In other words, e.g., we would like to make P_{41} as close to 1 as possible, P_{22} as close to 1 as possible, etc.

  IOW, we want to maximize the probability of the chosen brands. To accomplish that, we construct the likelihood function - the collection of the probabilities of the brand chosen on each occasion $P_{41}, P_{22}, P_{23,}$... To construct a simple entity to maximize, we recognize the independence of each observation and create the likelikhood function as:

$$L = \prod_{i=1}^{N} \prod_{t=1}^{T_i} \left[\prod_{j=1}^{J} P_{ijt}^{\delta_{ijt}} \right]$$

$i = 1, 2, ..., N$ households; $j = 1, 2, .., J$ brands; $t = 1, 2, ..., T_i$ choice occasions for household i.

where $\delta_{ijt} = 1$ if brand j is purchased at t by household i; 0 otherwise.

Practically this is equivalent to maximizing the log of the above function

$$LL = ln(L) = \sum_{i=1}^{N} \sum_{t=1}^{Ti} \sum_{j=1}^{J} \delta_{ijt} ln(P_{ijt})$$

- So to estimate the model parameters, we use a maximization algorithm to pick values of $\Theta = \{\alpha_1, \alpha_2, \alpha_3, \beta_f, \beta_p\}$ that maximizes $LL(\Theta)$.