Module 15
Estimating Customer Preferences from Choice Data

Dr. Jitesh H. Panchal

Decision Making in Engineering Design

School of Mechanical Engineering
Purdue University, West Lafayette, IN

Focus of this Module

Estimating Customers' Preferences

Given: decisions that have already been made.
Objective: to elicit the preference structures which led to the observed decisions.

- Utility theory for making decisions $\Rightarrow$ forward problem.
- Estimating customer preferences $\Rightarrow$ inverse problem.

Module Outline

1. Modeling Framework for Discrete Choice
   - Foundation: Random Utility Theory
   - Estimating Parameters in the Logit Model
   - An Illustrative Example
2. Power and Limitations of Logit
   - 1. Taste Variation
   - 2. Substitution Patterns
3. Other Models
   - 1. Generalized Extreme Value (GEV)
   - 2. Probit
   - 3. Mixed Logit

K. Train (1993). *Discrete Choice Methods with Simulation, 2nd Edition*. New York, NY, Cambridge University Press.

W. Chen, C. Hoyle, and H. J. Wassenaar (2013). *Decision-Based Design: Integrating Consumer Preferences into Engineering Design*. Springer.

Modeling Framework for Discrete Choice

Framing the Problem

1. Assume that the decision maker makes decisions by maximizing his/her utility $u()$.
2. Given two alternatives A_1 and A_2, the decision maker chooses A_1 if

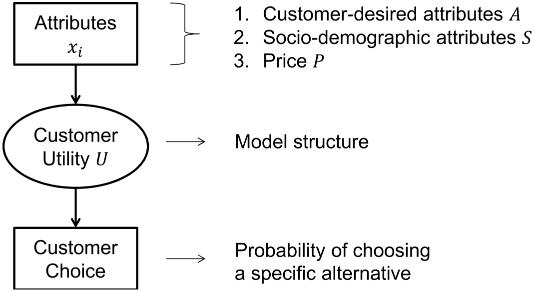
$$E[u(A_1)] > E[u(A_2)]$$

3. Say the alternatives are defined by attributes $X, Y, Z, \ldots$. Then, utility function is given by

$$u(x, y, z)$$

By observing which alternative is chosen (A_1 or A_2), and the values of attributes $(x, y, z, \ldots)$ for the alternatives (chosen or not), determine the function u().

Challenges



Challenges:

- Unable to observe all the attributes
- Unable to know the functional form of the utility function
- Limited number of prior decisions
- Anything else?

Discrete Choice Analysis

What is Discrete Choice?

A dependent variable that is a categorical, unordered variable. The choices/categories are called alternatives, and only one alternative can be selected.

Examples:

- **Mode of travel**: automobile, bus, rail transit, airplane, . . .
- **Class of vehicle owned**: sedan, cross-over, SUV, . . .
- **Brand of laptop**: Dell, Lenovo, HP, . . .

Random Utility Models

Two actors: "Decision maker" and "Researcher"

1. Decision maker:

- U_{nj}: Utility obtained by the decision maker n from choosing alternative j, where $j = 1, \ldots, J$
- Choose alternative i if and only if

$$U_{ni} > U_{nj} \quad \forall j \neq i$$

2. Researcher:

- Observes some attributes of the alternatives as faced by the decision-maker ($x_{nj} \quad \forall j$)
- Also observes some attributes (s_n) of the decision-maker.

Researcher's Perspective

The researcher models a "Representative Utility"

$$V_{nj} = V(x_{nj}, s_n) \quad \forall j$$

Since there are aspects of utility that the researcher does not observe,

$$V_{nj} \neq U_{nj}$$

Instead,

$$U_{nj} = V_{nj} + \epsilon_{nj}$$

where ϵ_{nj} captures the factors that affect utility but are not included in V_{nj}.

Since the researcher does not know ϵ_{nj}, he/she treats it as random. Denote the random vector associated with an individual as $\epsilon_n = \{\epsilon_{n1}, \epsilon_{n2}, \ldots \epsilon_{nJ}\}$.

Researcher's Perspective

Choice Probability

The probability that decision maker n chooses alternative i is

$$\begin{aligned} P_{ni} &= Prob(U_{ni} > U_{nj} \quad \forall j \neq i) \\ &= Prob(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \quad \forall j \neq i) \\ &= Prob(V_{ni} - V_{nj} > \epsilon_{nj} - \epsilon_{ni} \quad \forall j \neq i) \\ &= Prob(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \quad \forall j \neq i) \end{aligned}$$

This is the probability that the random term $\epsilon_{nj} - \epsilon_{ni}$ is below the observed quantity $(V_{ni} - V_{nj})$.

Example

Mode of Transportation

Alternatives:

- Take car to work (*c*)
- Take bus to work (*b*)

Researcher observed:

- Time incurred in travel (T)
- Cost incurred in travel (M)

$$V_{nc} = \alpha T_{nc} + \beta M_{nc}$$
$$V_{nb} = \alpha T_{nb} + \beta M_{nb}$$

Example

Mode of Transportation

The person will choose **bus over car** with

$$Prob(\epsilon_{nc} - \epsilon_{nb} < V_{nb} - V_{nc})$$

The person will choose **car over bus** with

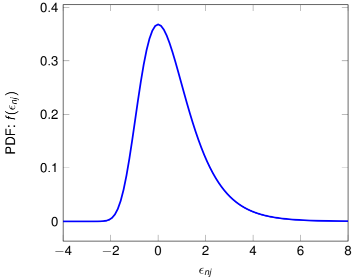
$$Prob(\epsilon_{nb} - \epsilon_{nc} < V_{nc} - V_{nb})$$

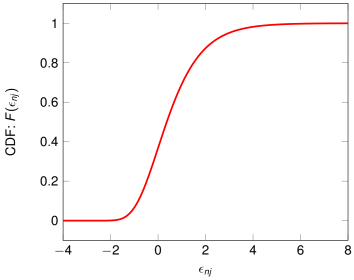
Special Case

$$U_{nj} = V_{nj} + \epsilon_{nj}$$

Assume a specific distribution of unobserved part of the utility ϵ_{nj}: distributed independently, identically extreme value (Gumbel)

- PDF: $f(\epsilon_{nj}) = e^{-\epsilon_{nj}} e^{-e^{-\epsilon_{nj}}}$
- CDF: $F(\epsilon_{nj}) = e^{-e^{-\epsilon_{nj}}}$



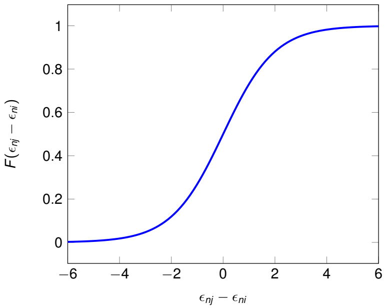


Special Case

If ϵ_{nj} and ϵ_{ni} are iid extreme value, then (derivation skipped)

$$F(\epsilon_{nj} - \epsilon_{ni}) = \frac{e^{(\epsilon_{nj} - \epsilon_{ni})}}{1 + e^{(\epsilon_{nj} - \epsilon_{ni})}}$$

This is the logistic distribution.



Recall: $P_{ni} = Prob(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \quad \forall j \neq i)$

Choice Among Two Alternatives (*i* and *j*)

Probability of choosing alternative *i* over *j*

$$\begin{aligned} P_{ni} &= Prob(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj}) \\ &= \frac{e^{(V_{ni} - V_{nj})}}{1 + e^{(V_{ni} - V_{nj})}} \\ &= \frac{e^{V_{ni}}}{e^{V_{ni}} + e^{V_{nj}}} \end{aligned}$$

Similarly, probability of choosing *j* is

$$P_{nj} = \frac{e^{V_{nj}}}{e^{V_{ni}} + e^{V_{nj}}}$$

In general, for *J* alternatives, the logit choice probability is:

$$\boxed{P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}}}$$

Properties of the Logit Choice Probability

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}}$$

Properties:

- It is always between 0 and 1.
- As V_{ni} rises with V_{nj} held constant, $P_{ni} \rightarrow 1$. Similarly, as V_{ni} is reduced with V_{nj} held constant, $P_{ni} \rightarrow 0$.
- The probability of choosing an alternative is never 0.
- The choice probability of all alternatives sum to 1 (i.e., one of the alternatives WILL be selected).
- If V_{ni} is low, a small improvement in the alternative has little effect on the probability of being chosen.
- The point at which the increase in V_{ni} has maximum impact on the probability when P_{ni} is 0.5.

Properties of the Logit Choice Probability

$$\boxed{P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}}}$$

V_{ni} can be linear or non-linear in the attributes. The simplest linear form is:

$$V_{ni} = \beta_1 x_1(i) + \beta_2 x_2(i) + \ldots$$

Model Estimation

Based on the collected discrete choice data (either *revealed or stated* choice), modeling techniques as introduced can be used to create a choice model that can predict the choices individual customer makes and to forecast the market demand. The preference, β, is readily estimated using maximum likelihood methods.

Existing commercial software that offer logit or probit modeling capabilities

- GENSTAT (www.vsn-intl.com)
- LIMDEP (www.limdep.com)
- SAS (www.sas.com)
- SPSS (www.spss.com)
- STATA (www.stata.com)
- SYSTAT (www.systa.com)

Estimating the Parameters in the Logit Model

A sample of N decision-makers is obtained for the purposes of estimation.

The probability of person n choosing the alternative that he was actually observed to choose can be expressed as

$$\prod_i (P_{ni})^{y_{ni}}$$

where $y_{ni} = 1$ of person n chooses i and zero otherwise.

Estimating the Parameters in the Logit Model

Assuming that each decision-maker's choice is independent of that of other decision-makers, the probability of each person in the sample choosing the alternative that he was observed actually to choose is

$$L(\beta) = \prod_{n=1}^{N} \prod_{i} (P_{ni})^{y_{ni}}$$

where β is a vector containing the parameters of the model.

The log-likelihood function is then

$$LL(\beta) = \sum_{n=1}^{N} \sum_{i} y_{ni} \ln(P_{ni})$$

Find β that maximizes $LL(\beta)$.

An Illustrative Example

Power Saw Design

Demand estimation model for a power saw design

Alternatives	Speed	Maintenance Frequency	Price
Saw 1	High	High	High
Saw 2	Medium	Low	Medium
Saw 3	Low	Medium	Low

Conducting Choice Set

- Different vendors (A, B, . . .) sell the saws.
- Normalized data is used for convenience computation and interpretation.

Vendor	Alternative	Price
A	1	0.97
A	2	0.73
A	3	0.63
B	1	1
B	2	0.72
B	3	0.55
. . .	. . .	. . .

Sample Data

Sample data representing the *revealed preference* of 15 customers who buy these saws from different vendors.

Customer sales data

Customer no.	Income	Vendor	Alternative chosen
1	0.44	A	2
2	0.62	B	3
3	0.81	C	1
4	0.32	D	3
5	0.78	E	2
6	1.00	F	1
7	0.84	G	1
8	0.39	H	2
9	0.55	I	3
10	0.62	J	3
11	0.66	K	1
12	0.50	L	3
13	0.43	M	1
14	0.76	N	1
15	0.32	O	3

Table 3.5 on page 64 (Wei Chen, Christopher Hoyle and Henk J. Wassenaar)

MNL Model

Assuming a linear form of the Representative Utility Function,
For $1 \leq n \leq 15$,

$$
\begin{aligned}
V_{n1} &= \beta_{speed} x_{speed}(1) + \beta_{price} x_{price}(1) + \beta_{maintenance} x_{maintenance}(1) \\
&\quad + \beta_{income(1)} x_{income(n)} \\
V_{n2} &= \beta_{speed} x_{speed}(2) + \beta_{price} x_{price}(2) + \beta_{maintenance} x_{maintenance}(2) \\
&\quad + \beta_{income(2)} x_{income(n)} \\
V_{n3} &= \beta_{speed} x_{speed}(3) + \beta_{price} x_{price}(3) + \beta_{maintenance} x_{maintenance}(3) \\
&\quad + \beta_{income(3)} x_{income(n)}
\end{aligned}
$$

Note: the β-coefficients of the product attributes are identical across all alternatives and all customers. However, the coefficient for the decision-maker specific attribute (income) varies across alternatives.

MNL Model - Output

Results output of the linear model

```
Iteration 0:   log likelihood = -14.359848
Iteration 1:   log likelihood = -8.9144052
Iteration 2:   log likelihood = -7.9954175
Iteration 3:   log likelihood = -7.8154834
Iteration 4:   log likelihood = -7.8035352
Iteration 5:   log likelihood = -7.8034579

Conditional (fixed-effects) logistic regression   Number of obs   =        45
                                                  LR chi2(5)      =     17.35
                                                  Prob > chi2     =    0.0039
Log likelihood = -7.8034579                       Pseudo R2       =    0.5265

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       speed |   47.08898   36.66279     1.28   0.199    -24.76876    118.9467
       price |  -55.95013   24.87006    -2.25   0.024    -104.6945   -7.205713
maintenanc~n |    28.0128   26.03458     1.08   0.282    -23.01405    79.03964
    income_2 |  -13.66506    7.74274    -1.76   0.078    -28.84055    1.510429
    income_3 |  -19.66365   9.744276    -2.02   0.044    -38.76208   -.5652178
------------------------------------------------------------------------------
```

Figure 3.10 on page 70 (Wei Chen, Christopher Hoyle and Henk J. Wassenaar)

MNL Model - Utility Results

For $n = 3$,

$$V_{31} = 47.09 \times 1 - 55.95 \times 0.95 + 28.01 \times 0.64 = 11.86$$
$$V_{32} = 47.09 \times 0.71 - 55.95 \times 0.75 + 28.01 \times 1 - 13.67 \times 0.81 = 8.42$$
$$V_{33} = 47.09 \times 0.67 - 55.95 \times 0.60 + 28.01 \times 0.89 - 19.66 \times 0.81 = 6.98$$

MNL Model – Predicted Choice Probability

For $n = 3$,

$$P_{31} = \frac{e^{V_{31}}}{e^{V_{31}} + e^{V_{32}} + e^{V_{33}}} = \frac{e^{11.86}}{e^{11.86} + e^{8.42} + e^{6.98}} = 0.96$$

$$P_{32} = \frac{e^{V_{32}}}{e^{V_{31}} + e^{V_{32}} + e^{V_{33}}} = \frac{e^{8.42}}{e^{11.86} + e^{8.42} + e^{6.98}} = 0.03$$

$$P_{33} = \frac{e^{V_{33}}}{e^{V_{31}} + e^{V_{32}} + e^{V_{33}}} = \frac{e^{6.98}}{e^{11.86} + e^{8.42} + e^{6.98}} = 0.01$$

Power and Limitations of Logit

Power and Limitations of Logit

The Independence Assumption

ϵ_{nj} and ϵ_{ni} are assumed to be independent. The error for one alternative provides no information about the error for another alternative.

The researcher has specified V_{nj} sufficiently that the remaining unobserved portion of utility is essentially “white noise”.

1. Taste Variation

Logit can represent systematic taste variation, but not random taste variation.

Example: Car buying decision

- Size (S) of the car (large family vs. small family)
- Purchase price (P) of the car (high income vs. low income)

Tastes that vary systematically with **observed** variables can be incorporated in logit models. Tastes that vary with **unobserved** variables cannot be handled.

Taste Variation

Example

Utility of a household in the car-buying decision:

$$U_{nj} = (\alpha_n S_j + \beta_n P_j) + \epsilon_{nj}$$

where α_n and β_n are parameters specific to household n. S_j is the size of alternative j and P_j is the purchase price of alternative j.

Suppose

- the preference for size varies with the number of members in the household (M_n)

$$\alpha_n = \rho M_n$$

- the importance of purchase price is inversely related to income (I_n)

$$\beta_n = \theta / I_n$$

Taste Variation
Example

Utility of a household in the car-buying decision:

$$U_{nj} = \rho(M_n S_j) + \theta(P_j / I_n) + \epsilon_{nj}$$

This utility function accounts for both the vehicle attribute and the household characteristics.

Taste Variation

Example

The limitation comes when the tastes vary with respect to unobserved variables.

Suppose

- the value of size varied with household size (M_n) and some other factors (e.g., size of the people themselves, or frequency with which the household travels together)

$$\alpha_n = \rho M_n + \mu_n$$

  where μ_n is a random variable.

- the importance of purchase price consists of its observed (I_n) and unobserved (η_n) components

$$\beta_n = \theta / I_n + \eta_n$$

Taste Variation

Example

Substituting these in the Utility of a household:

$$U_{nj} = \rho(M_n S_j) + \theta(P_j / I_n) + (\mu_n S_j + \eta_n P_j + \epsilon_{nj})$$

The new error term, $\tilde{\epsilon}_{nj} = (\mu_n S_j + \eta_n P_j + \epsilon_{nj})$ cannot be distributed independently and identically, as required by the logit formulation.

$$Cov(\tilde{\epsilon}_{nj}, \tilde{\epsilon}_{nk}) = Var(\mu_n) S_j S_k + Var(\eta_n) P_j P_k \neq 0$$

2. Substitution Patterns

Logit implies proportional substitution across alternatives.

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}}$$

An increase in the probability of one alternative necessarily means a decrease in probability of other alternatives.

For any two alternatives i and k, the ratio of logit probabilities is:

$$\frac{P_{ni}}{P_{nk}} = \frac{e^{V_{ni}} / \sum_j e^{V_{nj}}}{e^{V_{nk}} / \sum_j e^{V_{nj}}} = \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni} - V_{nk}}$$

The ratio does not depend on any alternatives other than i and k.

Independence of Irrelevant Alternatives (IIA)

The relative odds of choosing one alternative over another are independent of any other alternatives $\Rightarrow$ the logit model exhibits the Independence of Irrelevant Alternatives (IIA) property.

Consider a decision between two transportation alternatives:

- Car
- Blue bus

Say choice probabilities are $P_c = P_{bb} = 1/2$. Therefore, $P_c/P_{bb} = 1$.

Independence of Irrelevant Alternatives (IIA)

Suppose a new “red bus” is introduced.

- For the decision maker, the red bus is exactly the same as the “blue bus”. Therefore, $P_{rb}/P_{bb} = 1$
- Due to the IIA property, $P_c/P_{bb} = 1$ (does not change with the introduction of the irrelevant attribute).

Solving the above two equations, along with $P_c + P_{bb} + P_{rb} = 1$, we get

$$P_c = P_{bb} = P_{rb} = \frac{1}{3}$$

Note: The probability of taking a car has reduced from $\frac{1}{2}$ to $\frac{1}{3}$ just by introducing the red bus.
Instead, we would have expected $P_c = \frac{1}{2}$ and $P_{bb} = P_{rb} = \frac{1}{4}$

Other Models

1. Generalized Extreme Value (GEV)

Assumption

Unobserved portions of the utility, $\epsilon_n = \{\epsilon_{n1}, \epsilon_{n2}, \ldots \epsilon_{nJ}\}$, are jointly distributed as a generalized extreme value distribution.

Allows for correlations over alternatives.
Special case: When all correlations are zero, the model reduces to a standard logit.

Nested Logit: A Type of GEV

Nested Logit: Alternatives are partitioned into subsets (**nests**) such that

1. IIA holds within each nest
   – for two alternatives in the *same* next, the ratio of probabilities is independent of all other alternatives.
2. IIA does not hold for alternatives in different nests
   – for two alternatives within *different* nests, the ratio of probabilities depend on other alternatives in the two nests.

Nested Logit: Red Bus, Blue Bus Example

Nests of alternatives:

1. **Car**
2. **Bus**
   - Red
   - Blue

Nested Logit: Example

Alternatives:

1. auto-alone ($P_a = 0.4$)
2. carpooling ($P_c = 0.1$)
3. taking a bus ($P_b = 0.3$)
4. taking rail ($P_r = 0.2$)

Note:

$$\frac{P_b}{P_r} = \frac{0.3}{0.2} = 1.5 \qquad \frac{P_a}{P_c} = \frac{0.4}{0.1} = 4$$

$$\frac{P_a}{P_b} = \frac{0.4}{0.3} = 1.33 \qquad \frac{P_c}{P_r} = \frac{0.1}{0.2} = 0.5$$

By what proportion would each probability increase when an alternative is removed?

Nested Logit: Example

If "auto-alone" is removed, then say the new probabilities are:

1. ~~auto-alone~~
2. carpooling ($P_c = 0.2$)
3. taking a bus ($P_b = 0.48$)
4. taking rail ($P_r = 0.32$)

Here,

$$\frac{P_b}{P_r} = \frac{0.48}{0.32} = 1.5$$

$$\frac{P_c}{P_r} = \frac{0.2}{0.32} = 0.625$$

Nested Logit: Example

If "bus" is removed, then say the new probabilities are:

1. auto-alone ($P_a = 0.52$)
2. carpooling ($P_c = 0.13$)
3. ~~taking a bus~~
4. taking rail ($P_r = 0.35$)

Here,

$$\frac{P_a}{P_c} = \frac{0.52}{0.13} = 4$$

$$\frac{P_c}{P_r} = \frac{0.13}{0.35} = 0.37$$

Nested Logit: Transportation Example

Nests of alternatives:

1. **Auto**
   - auto-alone
   - carpool
2. **Transit**
   - Bus
   - Rail

Nested Logit: Decomposition into Two Logits

Utility Function

Denote K non-overlapping nests as $B_1, B_2, \ldots, B_K$.

Decompose the Utility of alternative $j \in B_k$ into two parts:

$$U_{nj} = W_{nk} + Y_{nj} + \epsilon_{nj}$$

1. W: constant for all alternatives within a nest
2. Y: varies over alternatives within a nest

Nested Logit: Decomposition into Two Logits

Choice Probability

Probability of choosing alternative $i \in B_k$ can be expressed as:

$$P_{ni} = P_{ni|B_k} P_{nB_k}$$

$$P_{ni} = \left(\frac{e^{Y_{ni}/\lambda_k}}{\sum_{j \in B_k} e^{Y_{nj}/\lambda_k}} \right) \left(\frac{e^{W_{nk} + \lambda_k I_{nk}}}{\sum_{l=1}^{K} e^{W_{nl} + \lambda_l I_{nl}}} \right)$$

where

$$I_{nk} = \ln \sum_{j \in B_k} e^{Y_{nj}/\lambda_k}$$

Parameter λ_k measures the degree of independence in the unobserved utility among alternatives in nest k. If $\lambda_k = 1$ for all nests, all alternatives are independent $\Rightarrow$ standard Logit

Nested Logit Model

Another View

Nested logit is obtained by assuming that $\epsilon_n = \{\epsilon_{n1}, \epsilon_{n2}, \ldots \epsilon_{nJ}\}$ has the following cumulative distribution:

$$\exp\left(-\sum_{k=1}^{K}\left(\sum_{j \in B_k} e^{-\epsilon_{nj}/\lambda_k}\right)^{\lambda_k}\right)$$

Properties:

- ϵ_{nj}'s are correlated within nests but not across nests.
- For alternatives in different nests, $Cov(\epsilon_{nj}, \epsilon_{nm}) = 0$

Compare to CDF in standard Logit: $\exp(-e^{-\epsilon_{nj}})$

Nested Logit Model: Choice Probability

General Results

Choice Probability in the Nested Logit is:

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l}}$$

Ratio of choice probabilities of two alternatives $i \in B_k$ and $m \in B_l$:

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{e^{V_{nm}/\lambda_l}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l - 1}}$$

For alternatives in the same nest, $k = l$. Therefore,

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k}}{e^{V_{nm}/\lambda_l}}$$

2. Probit

Assumption

Unobserved utility components $\epsilon_n = \{\epsilon_{n1}, \epsilon_{n2}, \ldots \epsilon_{nJ}\}$ are jointly normal

Density of ϵ_n is

$$\phi(\epsilon_n) = \frac{1}{(2\pi)^{J/2}|\Omega_n|^{1/2}} e^{-\frac{1}{2}\epsilon_n' \Omega_n^{-1} \epsilon_n}$$

Choice probability:

$$P_{ni} = \int I(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \forall j \neq i) \phi(\epsilon_n) d\epsilon_n$$

Must be evaluated numerically!

2. Probit

Properties

Properties of Probit:

1. It can represent any substitution pattern.
2. It does not exhibit the IIA property.
3. Different covariance matrices provide different substitution patterns. It can be estimated using data.
4. It can handle random taste variation.

3. Mixed Logit

The utility of a person n obtained from alternative j is

$$U_{nj} = \beta_n x_{nj} + \epsilon_{nj}$$

Assumption: The coefficients β vary over decision makers in the population with density $f(\beta)$

The decision maker knows the value of his/her own β and ϵ_{nj}'s for all j. But the researcher only observes x_{nj}.

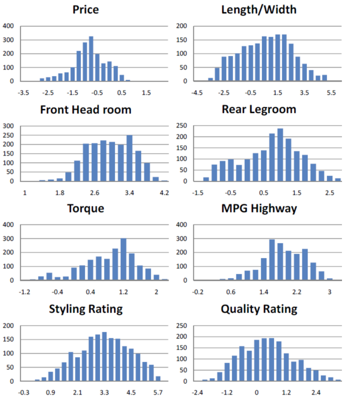
Mixed Logit Example

Example: Vehicle Design Attributes

1. Price
2. Length/Width
3. Front headroom
4. Rear Legroom
5. Torque
6. MPG highway
7. Styling
8. Quality

He, L., Hoyle, C., Chen, W., 2009, A Mixed Logit Choice Modeling Approach Using Customer Satisfaction Surveys to Support Engineering Design, *ASME IDETC Conference*, San Diego CA. doi:10.1115/DETC2009-87052

Mixed Logit Example



Distributions of parameters displaying heterogeneity (Figure 6)

Parameter (β)	Mean	St. Dev.
Price	-0.944	0.251
Length/Width	0.371	0.263
Front headroom	2.855	0.227
Rear legroom	0.590	0.170
Torque	0.751	0.187
MPG highway	1.776	0.154
Styling	3.136	0.485
Quality	0.271	0.126

He, L., Hoyle, C., Chen, W., 2009, A Mixed Logit Choice Modeling Approach Using Customer Satisfaction Surveys to Support Engineering Design, *ASME IDETC Conference*, San Diego CA. doi:10.1115/DETC2009-87052

Mixed Logit

Mixed logit probabilities are the integral of standard logit probabilities over a density of parameters.

$$P_{ni} = \int L_{ni}(\beta) f(\beta) d\beta$$

where,
$f(\beta)$ is a density function (mixing distribution)
$L_{ni}(\beta)$ is the logit probability evaluated at parameters β:

$$L_{ni}(\beta) = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^{J} e^{V_{nj}(\beta)}}$$

In other words, the probability is a weighted average of the logit formula evaluated at different values of β, with weights given by density $f(\beta)$.

Mixed Logit

Standard logit is a special case where the mixing distribution is degenerate at fixed parameters.

Population with m segments:
If β can take discrete values $\beta_1, \beta_2, \ldots, \beta_M$, where β_m has a probability s_m then for a linear V_{ni}

$$P_{ni} = \sum_{m=1}^{M} s_m \left(\frac{e^{\beta_m x_{ni}}}{\sum_j e^{\beta_m x_{nj}}} \right)$$

Mixed Logit

The mixing distribution can be continuous.

For example: Normal with mean μ and covariance W. Then,

$$P_{ni} = \int \left(\frac{e^{\beta x_{ni}}}{\sum_j e^{\beta x_{nj}}} \right) \phi(\beta|\mu, W) d\beta$$

Note: There are two sets of parameters: β and (μ, W).

Summary

1. Modeling Framework for Discrete Choice
   - Foundation: Random Utility Theory
   - Estimating Parameters in the Logit Model
   - An Illustrative Example
2. Power and Limitations of Logit
   - 1. Taste Variation
   - 2. Substitution Patterns
3. Other Models
   - 1. Generalized Extreme Value (GEV)
   - 2. Probit
   - 3. Mixed Logit

References

1. K. Train (1993). *Discrete Choice Methods with Simulation*, 2nd Edition. New York, NY, Cambridge University Press.
2. W. Chen, C. Hoyle, and H. J. Wassenaar (2013). *Decision-Based Design: Integrating Consumer Preferences into Engineering Design*. Springer.