Discrete Choice Analysis II

Moshe Ben-Akiva

1.201 / 11.545 / ESD.210
Transportation Systems Analysis: Demand & Economics

Fall 2008
Review – Last Lecture

- Introduction to Discrete Choice Analysis
- A simple example – route choice
- The Random Utility Model
  - Systematic utility
  - Random components
- Derivation of the Probit and Logit models
  - Binary Probit
  - Binary Logit
  - Multinomial Logit
Outline – This Lecture

- Model specification and estimation
- Aggregation and forecasting
- Independence from Irrelevant Alternatives (IIA) property – Motivation for Nested Logit
- Nested Logit - specification and an example
- Appendix:
  - Nested Logit model specification
  - Advanced Choice Models
Specification of Systematic Components

● Types of Variables
  – Attributes of alternatives: $Z_{in}$, e.g., travel time, travel cost
  – Characteristics of decision-makers: $S_{n}$, e.g., age, gender, income, occupation
  – Therefore: $X_{in} = h(Z_{in}, S_{n})$

● Examples:
  – $X_{in1} = Z_{in1} = \text{travel cost}$
  – $X_{in2} = \log(Z_{in2}) = \log(\text{travel time})$
  – $X_{in3} = Z_{in1}/S_{n1} = \text{travel cost} / \text{income}$

● Functional Form: Linear in the Parameters

\[
V_{in} = \beta_1 X_{in1} + \beta_2 X_{in2} + \ldots + \beta_k X_{inK}
\]
\[
V_{jn} = \beta_1 X_{jn1} + \beta_2 X_{jn2} + \ldots + \beta_k X_{jnK}
\]
Data Collection

- Data collection for each individual in the sample:
  - *Choice set: available alternatives*
  - *Socio-economic characteristics*
  - *Attributes of available alternatives*
  - *Actual choice*

<table>
<thead>
<tr>
<th>n</th>
<th>Income</th>
<th>Auto Time</th>
<th>Transit Time</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35</td>
<td>15.4</td>
<td>58.2</td>
<td>Auto</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>14.2</td>
<td>31.0</td>
<td>Transit</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>19.6</td>
<td>43.6</td>
<td>Auto</td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td>50.8</td>
<td>59.9</td>
<td>Auto</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>55.5</td>
<td>33.8</td>
<td>Transit</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>N/A</td>
<td>48.4</td>
<td>Transit</td>
</tr>
</tbody>
</table>
Model Specification Example

\[ V_{\text{auto}} = \beta_0 + \beta_1 TT_{\text{auto}} + \beta_2 \ln(\text{Income}) \]

\[ V_{\text{transit}} = \beta_1 TT_{\text{transit}} \]

<table>
<thead>
<tr>
<th></th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto</td>
<td>1</td>
<td>( TT_{\text{auto}} )</td>
<td>( \ln(\text{Income}) )</td>
</tr>
<tr>
<td>Transit</td>
<td>0</td>
<td>( TT_{\text{transit}} )</td>
<td>0</td>
</tr>
</tbody>
</table>
Probabilities of Observed Choices

- **Individual 1:**
  
  \[
  V_{\text{auto}} = \beta_0 + \beta_1 \times 15.4 + \beta_2 \ln(35)
  
  V_{\text{transit}} = \beta_1 \times 58.2
  \]

  \[
P(\text{Auto}) = \frac{e^{\beta_0 + 15.4\beta_1 + \ln(35)\beta_2}}{e^{\beta_0 + 15.4\beta_1 + \ln(35)\beta_2} + e^{58.2\beta_1}}
  \]

- **Individual 2:**

  \[
  V_{\text{auto}} = \beta_0 + \beta_1 \times 14.2 + \beta_2 \ln(45)
  
  V_{\text{transit}} = \beta_1 \times 31.0
  \]

  \[
P(\text{Transit}) = \frac{e^{31.0\beta_1}}{e^{\beta_0 + 14.2\beta_1 + \ln(45)\beta_2} + e^{31.0\beta_1}}
  \]
Maximum Likelihood Estimation

- Find the values of $\beta$ that are most likely to result in the choices observed in the sample:
  \[ \max L^*(\beta) = P_1(\text{Auto})P_2(\text{Transit})\ldots P_6(\text{Transit}) \]

- If \( y_{in} = \begin{cases} 1, \text{ if person } n \text{ chose alternative } i \\ 0, \text{ if person } n \text{ chose alternative } j \end{cases} \)

- Then we maximize, over choices of \( \{\beta_1, \beta_2, \ldots, \beta_k\} \), the following expression:
  \[ L^*(\beta_1, \beta_2, \ldots, \beta_k) = \prod_{n=1}^{N} P_n(i)^{y_{in}} P_n(j)^{y_{jn}} \]

- $\beta^* = \arg \max_{\beta} L^*(\beta_1, \beta_2, \ldots, \beta_k)$
  \[ = \arg \max_{\beta} \log L^*(\beta_1, \beta_2, \ldots, \beta_k) \]
Sources of Data on User Behavior

- Revealed Preferences Data
  - Travel Diaries
  - Field Tests
- Stated Preferences Data
  - Surveys
  - Simulators
Stated Preferences / Conjoint Experiments

● Used for product design and pricing
  – For products with significantly different attributes
  – When attributes are strongly correlated in real markets
  – Where market tests are expensive or infeasible

● Uses data from survey “trade-off” experiments in which attributes of the product are systematically varied

● Applied in transportation studies since the early 1980s
Aggregation and Forecasting

- Objective is to make aggregate predictions from
  - A disaggregate model, $P(i / X_n)$
  - Which is based on individual attributes and characteristics, $X_n$
  - Having only limited information about the explanatory variables
The Aggregate Forecasting Problem

- The fraction of population $T$ choosing alt. $i$ is:

\[
W(i) = \int P(i|X)p(X)dX, \quad p(X) \text{ is the density function of } X
\]

\[
= \frac{1}{N_T} \sum_{n=1}^{N_T} P(i|X_n), \quad N_T \text{ is the # in the population of interest}
\]

- Not feasible to calculate because:
  - We never know each individual’s complete vector of relevant attributes
  - $p(X)$ is generally unknown

- The problem is to reduce the required data
Sample Enumeration

- Use a sample to represent the entire population
- For a random sample:
  \[
  \hat{W}(i) = \frac{1}{N_s} \sum_{n=1}^{N_s} \hat{P}(i \mid x_n)
  \]
  where \( N_s \) is the # of obs. in sample

- For a weighted sample:
  \[
  \hat{W}(i) = \sum_{n=1}^{N_s} \frac{w_n}{\sum_n w_n} \hat{P}(i \mid x_n)
  \]
  where \( \frac{1}{w_n} \) is \( x_n \)'s selection prob.

- No aggregations bias, but there is sampling error
Disaggregate Prediction

1. Generate a representative population

2. Apply demand model
   - Calculate probabilities or simulate decision for each decision maker
   - Translate into trips
   - Aggregate trips to OD matrices

3. Assign traffic to a network

4. Predict system performance
Generating Disaggregate Populations

- Household surveys
- Exogenous forecasts
- Census data
- Counts

Data fusion (e.g., IPF, HH evolution)

Representative Population
Review

● Empirical issues
  – Model specification and estimation
  – Aggregate forecasting

● Next…More theoretical issues
  – Independence from Irrelevant Alternatives (IIA) property – Motivation for Nested Logit
  – Nested Logit - specification and an example
Summary of Basic Discrete Choice Models

- **Binary Probit:**
  \[ P_n(i|C_n) = \Phi(V_n) = \int_{-\infty}^{V_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \epsilon^2} d\epsilon \]

- **Binary Logit:**
  \[ P_n(i|C_n) = \frac{1}{1 + e^{-V_n}} = \frac{e^{V_{in}}}{e^{V_{in}} + e^{V_{jn}}} \]

- **Multinomial Logit:**
  \[ P_n(i|C_n) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \]
Independence from Irrelevant Alternatives (IIA)

Property of the Multinomial Logit Model

- $\epsilon_{jn}$ independent identically distributed (i.i.d.)
- $\epsilon_{jn} \sim \text{ExtremeValue}(0, \mu)$ \( \forall j \)

\[
 P_n(i|C_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}}
\]

so

\[
 \frac{P(i|C_1)}{P(j|C_1)} = \frac{P(i|C_2)}{P(j|C_2)} \quad \forall i, j, C_1, C_2
\]

such that $i, j \in C_1$, $i, j \in C_2$, $C_1 \subseteq C_n$ and $C_2 \subseteq C_n$
Examples of IIA

- Route choice with an overlapping segment

\[ P(1|\{1,2a,2b\}) = P(2a|\{1,2a,2b\}) = P(2b|\{1,2a,2b\}) = \frac{e^{\mu T}}{\sum_{j \in \{1,2a,2b\}} e^{\mu T}} = \frac{1}{3} \]
Red Bus / Blue Bus Paradox

- Consider that initially auto and bus have the same utility
  - \( C_n = \{\text{auto, bus}\} \) and \( V_{\text{auto}} = V_{\text{bus}} = V \)
  - \( P(\text{auto}) = P(\text{bus}) = 1/2 \)

- Suppose that a new bus service is introduced that is identical to the existing bus service, except the buses are painted differently (red vs. blue)
  - \( C_n = \{\text{auto, red bus, blue bus}\}; V_{\text{red bus}} = V_{\text{blue bus}} = V \)
  - Logit now predicts
    - \( P(\text{auto}) = P(\text{red bus}) = P(\text{blue bus}) = 1/3 \)
    - We’d expect
      - \( P(\text{auto}) = 1/2, P(\text{red bus}) = P(\text{blue bus}) = 1/4 \)
IIA and Aggregation

- Divide the population into two equally-sized groups: those who prefer autos, and those who prefer transit
- Mode shares before introducing blue bus:

<table>
<thead>
<tr>
<th>Population</th>
<th>Auto Share</th>
<th>Red Bus Share</th>
<th>P(auto)/P(red bus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto people</td>
<td>90%</td>
<td>10%</td>
<td>9</td>
</tr>
<tr>
<td>Transit people</td>
<td>10%</td>
<td>90%</td>
<td>1/9</td>
</tr>
<tr>
<td>Total</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

- Auto and red bus share ratios remain constant for each group after introducing blue bus:

<table>
<thead>
<tr>
<th>Population</th>
<th>Auto Share</th>
<th>Red Bus Share</th>
<th>Blue Bus Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto people</td>
<td>81.8%</td>
<td>9.1%</td>
<td>9.1%</td>
</tr>
<tr>
<td>Transit people</td>
<td>5.2%</td>
<td>47.4%</td>
<td>47.4%</td>
</tr>
<tr>
<td>Total</td>
<td>43.5%</td>
<td>28.25%</td>
<td>28.25%</td>
</tr>
</tbody>
</table>
Motivation for Nested Logit

- Overcome the IIA Problem of Multinomial Logit when
  - Alternatives are correlated (e.g., red bus and blue bus)
  - Multidimensional choices are considered (e.g., departure time and route)
Tree Representation of Nested Logit

- Example: Mode Choice (Correlated Alternatives)

![Diagram of a tree representing nested logit model for mode choice, with categories motorized, non-motorized, auto, carpool, bus, metro, bicycle, and walk.](image)
Tree Representation of Nested Logit

- Example: Route and Departure Time Choice (Multidimensional Choice)
Nested Model Estimation

- Logit at each node
- Utilities at lower level enter at the node as the *inclusive* value

$$I_{NM} = \ln \left( \sum_{i \in C_{NM}} e^{V_i} \right)$$

- The inclusive value is often referred to as *logsum*
Nested Model – Example

\[ P(i \mid NM) = \frac{e^{\mu_{NM}V_i}}{e^{\mu_{NM}V_{Walk}} + e^{\mu_{NM}V_{Bike}}} \quad i = \text{Walk, Bike} \]

\[ I_{NM} = \frac{1}{\mu_{NM}} \ln\left(e^{\mu_{NM}V_{Walk}} + e^{\mu_{NM}V_{Bike}}\right) \]
Nested Model – Example

\[ P(i \mid M) = \frac{e^{\mu_i M V_i}}{e^{\mu_{\text{Car}} M V_{\text{Car}}} + e^{\mu_{\text{Taxi}} M V_{\text{Taxi}}} + e^{\mu_{\text{Bus}} M V_{\text{Bus}}}} \quad i = \text{Car, Taxi, Bus} \]

\[ I_M = \frac{1}{\mu_M} \ln \left( e^{\mu_{\text{Car}} M V_{\text{Car}}} + e^{\mu_{\text{Taxi}} M V_{\text{Taxi}}} + e^{\mu_{\text{Bus}} M V_{\text{Bus}}} \right) \]
Nested Model – Example

\[ P(NM) = \frac{e^{\mu_I NM}}{e^{\mu_I NM} + e^{\mu_I M}} \]

\[ P(M) = \frac{e^{\mu_I M}}{e^{\mu_I NM} + e^{\mu_I M}} \]
Nested Model – Example

- Calculation of choice probabilities

\[ P(\text{Bus}) = P(\text{Bus} | M) \cdot P(M) \]

\[
\begin{align*}
\frac{e^{\mu_M V_{Bus}}}{e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}}} \cdot \frac{e^{\mu_{NM} V_{Car}} + e^{\mu_{NM} V_{Taxi}} + e^{\mu_{NM} V_{Bus}}}{e^{\mu_{NM}} + e^{\mu_M}} \\
\frac{e^{\mu_M V_{Bus}}}{e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}}} \cdot \frac{-\mu \ln(e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}})}{e^{\mu_M}} \\
\frac{-\mu \ln(e^{\mu_{NM} V_{Walk}} + e^{\mu_{NM} V_{Bike}})}{e^{\mu_{NM}}} + e^{\mu_M}
\end{align*}
\]
Extensions to Discrete Choice Modeling

- Multinomial Probit (MNP)
- Sampling and Estimation Methods
- Combined Data Sets
- Taste Heterogeneity
- Cross Nested Logit and GEV Models
- Mixed Logit and Probit (Hybrid Models)
- Latent Variables (e.g., Attitudes and Perceptions)
- Choice Set Generation
Summary

- Introduction to Discrete Choice Analysis
- A simple example
- The Random Utility Model
- Specification and Estimation of Discrete Choice Models
- Forecasting with Discrete Choice Models
- IIA Property - Motivation for Nested Logit Models
- Nested Logit
Additional Readings


- And/Or take 1.202 next semester!
Appendix

Nested Logit model specification
Cross-Nested Logit
Logit Mixtures (Continuous/Discrete)
Revealed + Stated Preferences
Nested Logit Model Specification

- Partition $C_n$ into $M$ non-overlapping nests:
  \[ C_{mn} \cap C_{m'n} = \emptyset \quad \forall \ m \neq m' \]

- Deterministic utility term for nest $C_{mn}$:
  \[
  V_{C_{mn}} = \tilde{V}_{C_{mn}} + \frac{1}{\mu_m} \ln \sum_{j \in C_{mn}} e^{\mu_m \tilde{V}_{jn}}
  \]

- Model:
  \[
  P(i \mid C_n) = P(C_{mn} \mid C_n) P(i \mid C_{mn}), \quad i \in C_{mn} \subseteq C_n
  \]
  where
  \[
  P(C_{mn} \mid C_n) = \frac{e^{\mu \tilde{V}_{C_{mn}}}}{\sum_l e^{\mu \tilde{V}_{C_{ln}}}} \quad \text{and} \quad P(i \mid C_{mn}) = \frac{e^{\mu_m \tilde{V}_{in}}}{\sum_{j \in C_{mn}} e^{\mu_m \tilde{V}_{jn}}}
  \]
Continuous Logit Mixture

Example:

● Combining Probit and Logit

● Error decomposed into two parts
  – Probit-type portion for *flexibility*
  – i.i.d. Extreme Value for *tractability*

● An intuitive, practical, and powerful method
  – Correlations across alternatives
  – Taste heterogeneity
  – Correlations across space and time

● Requires simulation-based estimation
Cont. Logit Mixture: Error Component

Illustration

- Utility:
  
  \[ U_{\text{auto}} = \beta X_{\text{auto}} + \xi_{\text{auto}} + \nu_{\text{auto}} \]
  
  \[ U_{\text{bus}} = \beta X_{\text{bus}} + \xi_{\text{bus}} + \nu_{\text{bus}} \]
  
  \[ U_{\text{subway}} = \beta X_{\text{subway}} + \xi_{\text{subway}} + \nu_{\text{subway}} \]

- Probability:

  \[ \Lambda(\text{auto}|X, \xi) = \frac{e^{\beta X_{\text{auto}} + \xi_{\text{auto}}}}{e^{\beta X_{\text{auto}} + \xi_{\text{auto}}} + e^{\beta X_{\text{bus}} + \xi_{\text{bus}}} + e^{\beta X_{\text{subway}} + \xi_{\text{subway}}}} \]

  \[ \xi \text{ unknown} \rightarrow \]

  \[ P(\text{auto}|X) = \int_{\xi} \Lambda(\text{auto} | X, \xi) f(\xi) d\xi \]

\[ \nu \text{ i.i.d. Extreme Value} \]
\[ e.g. \; \xi \sim N(0, \Sigma) \]
Continuous Logit Mixture

Random Taste Variation

- Logit: \( \beta \) is a constant vector
  - Can segment, e.g. \( \beta_{low \text{ inc}} \), \( \beta_{med \text{ inc}} \), \( \beta_{high \text{ inc}} \)

- Logit Mixture: \( \beta \) can be randomly distributed
  - Can be a function of personal characteristics
  - Distribution can be Normal, Lognormal, Triangular, etc
Discrete Logit Mixture

Latent Classes

Main Postulate:

- Unobserved heterogeneity is “generated” by discrete or categorical constructs such as
  - Different decision protocols adopted
  - Choice sets considered may vary
  - Segments of the population with varying tastes
- Above constructs characterized as \textit{latent classes}
Latent Class Choice Model

\[ P(i) = \sum_{s=1}^{S} \Lambda(i \mid s)Q(s) \]

Class-specific Choice Model

Class Membership Model

(probability of choosing \( i \) conditional on belonging to class \( s \))

(probability of belonging to class \( s \))
## Summary of Discrete Choice Models

<table>
<thead>
<tr>
<th>Feature</th>
<th>Logit</th>
<th>NL/CNL</th>
<th>Probit</th>
<th>Logit Mixture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handles unobserved taste heterogeneity</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Flexible substitution pattern</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Handles panel data</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Requires error terms normally distributed</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Closed-form choice probabilities available</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No (cont.) Yes (discrete)</td>
</tr>
<tr>
<td>Numerical approximation and/or simulation</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes (cont.) No (discrete)</td>
</tr>
</tbody>
</table>
6. Revealed and Stated Preferences

• Revealed Preferences Data
  – Travel Diaries
  – Field Tests

• Stated Preferences Data
  – Surveys
  – Simulators
Stated Preferences / Conjoint Experiments

• Used for product design and pricing
  – For products with significantly different attributes
  – When attributes are strongly correlated in real markets
  – Where market tests are expensive or infeasible

• Uses data from survey “trade-off” experiments in which attributes of the product are systematically varied

• Applied in transportation studies since the early 1980s

• Can be combined with Revealed Preferences Data
  – Benefit from strengths
  – Correct for weaknesses
  – Improve efficiency
Framework for Combining Data

Attributes of Alternatives & Characteristics of Decision-Maker

Utility

Revealed Preferences

Stated Preferences