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:
  - $X_{in1} = Z_{in1}$ = travel cost
  - $X_{in2} = log(Z_{in2})$ = log (travel time)
  - $X_{in3} = Z_{in1}/S_{n1}$ = travel cost / 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*

n	**Income**	**Auto Time**	**Transit Time**	**Choice**
1	35	15.4	58.2	Auto
2	45	14.2	31.0	Transit
3	37	19.6	43.6	Auto
4	42	50.8	59.9	Auto
5	32	55.5	33.8	Transit
6	15	N/A	48.4	Transit

Model Specification Example

$$V_{auto} = \beta_0 + \beta_1 \, TT_{auto} + \beta_2 \, ln(Income)$$
$$V_{transit} = \beta_1 \, TT_{transit}$$

	β_0	β_1	β_2
Auto	*1*	TT_{auto}	*ln(Income)*
Transit	*0*	$TT_{transit}$	*0*

Probabilities of Observed Choices

- Individual 1:

$V_{auto} = \beta_0 + \beta_1\ 15.4 + \beta_2\ ln(35)$

$V_{transit} = \beta_1\ 58.2$

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

- Individual 2:

$V_{auto} = \beta_0 + \beta_1\ 14.2 + \beta_2\ ln(45)$

$V_{transit} = \beta_1\ 31.0$

$$P(Transit) = \frac{e^{31.0\beta_1}}{e^{\beta_0 + 14.2\beta_1 + \ln(45)\beta_2} + e^{31.0\beta_1}}$$

Maximum Likelihood Estimation

- Find the values of β 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, $P(i \mid X_n)$
  - 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:

$$W(i) = \int_X P(i|X)p(X)dX$$, *p*(*X*) is the density function of *X*

$$= \frac{1}{N_T}\sum_{n=1}^{N_T} P(i|X_n)$$, N_T 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:

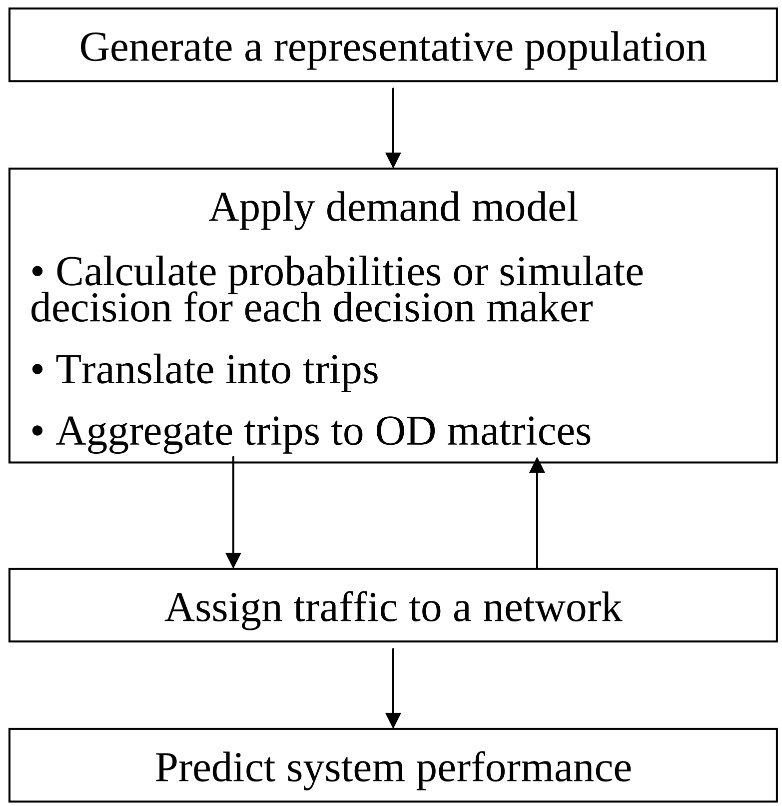
$$\hat{W}(i) = \frac{1}{N_s} \sum_{n=1}^{N_s} \hat{P}(i \mid x_n)$$ where N_s is the # of obs. in sample

- For a weighted sample:

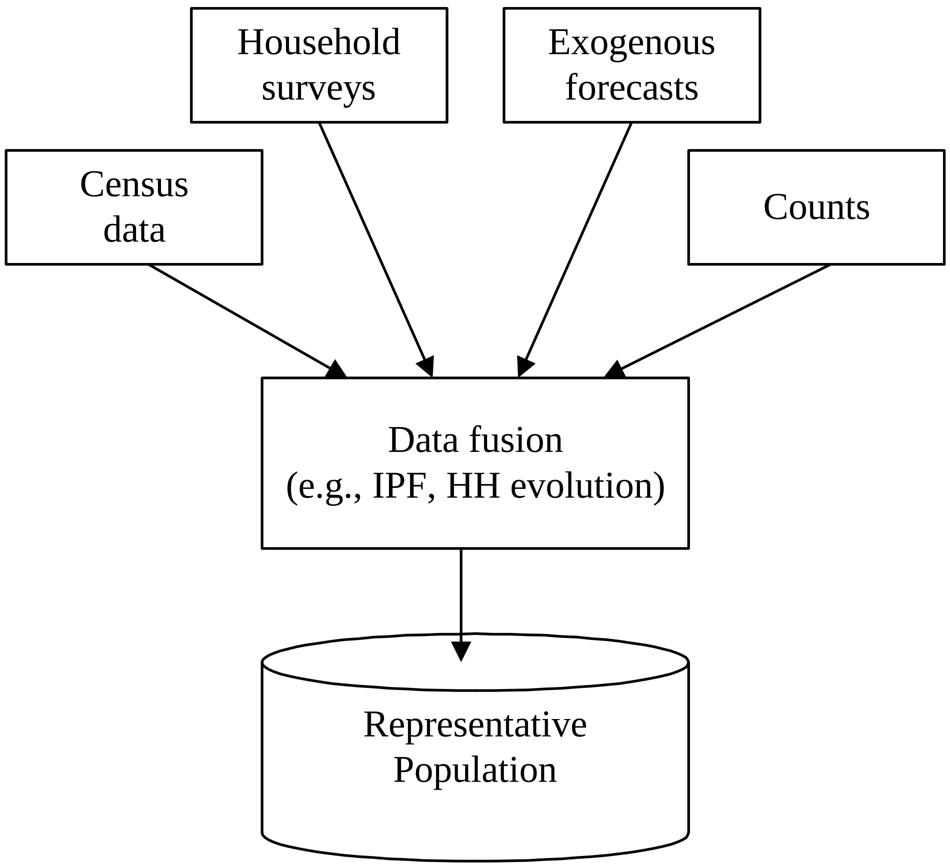
$$\hat{W}(i) = \sum_{n=1}^{N_s} \frac{w_n}{\sum_n w_n} \hat{P}(i \mid x_n) \text{ , where } \frac{1}{w_n} \text{ is } x_n\text{'s selection prob.}$$

- No aggregations bias, but there is sampling error

Disaggregate Prediction



Generating Disaggregate Populations



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:

$$P_n(i|C_n) = \Phi(V_n) = \int_{-\infty}^{V_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\varepsilon^2} d\varepsilon$$

- 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)

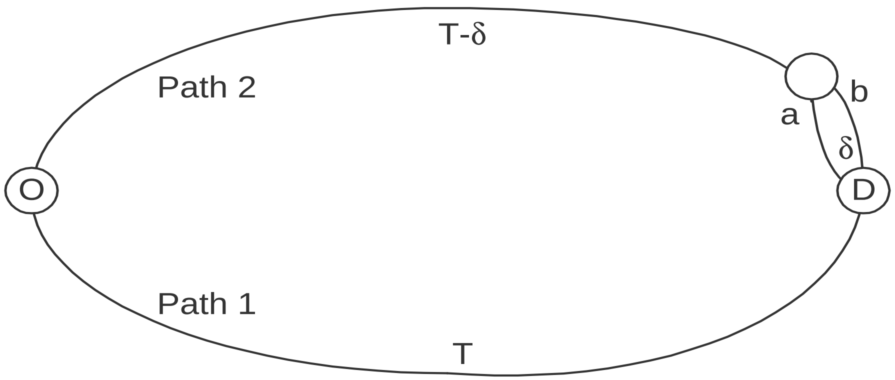
- Property of the Multinomial Logit Model
  - ε_{jn} independent identically distributed (i.i.d.)
  - $\varepsilon_{jn} \sim ExtremeValue(0,\mu) \quad \forall j$
  - $$P_n(i|C_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}}$$

so $$\frac{P(i|C_1)}{P(j|C_1)} = \frac{P(i|C_2)}{P(j|C_2)} \qquad \forall\, i, j, C_1, C_2$$

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 = {auto, bus} *and* $V_{auto} = V_{bus} = V$
  - P(auto) = P(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 = {auto, red bus, blue bus}; $V_{\text{red bus}} = V_{\text{blue bus}} = V$
  - Logit now predicts
    P(auto) = P(red bus) = P(blue bus) =1/3
  - We'd expect
    P(auto) =1/2, P(red bus) = P(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:

Population	Auto Share	Red Bus Share	
Auto people	90%	10%	P(auto)/P(red bus) = 9
Transit people	10%	90%	P(auto)/P(red bus) = 1/9
Total	**50%**	**50%**	

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

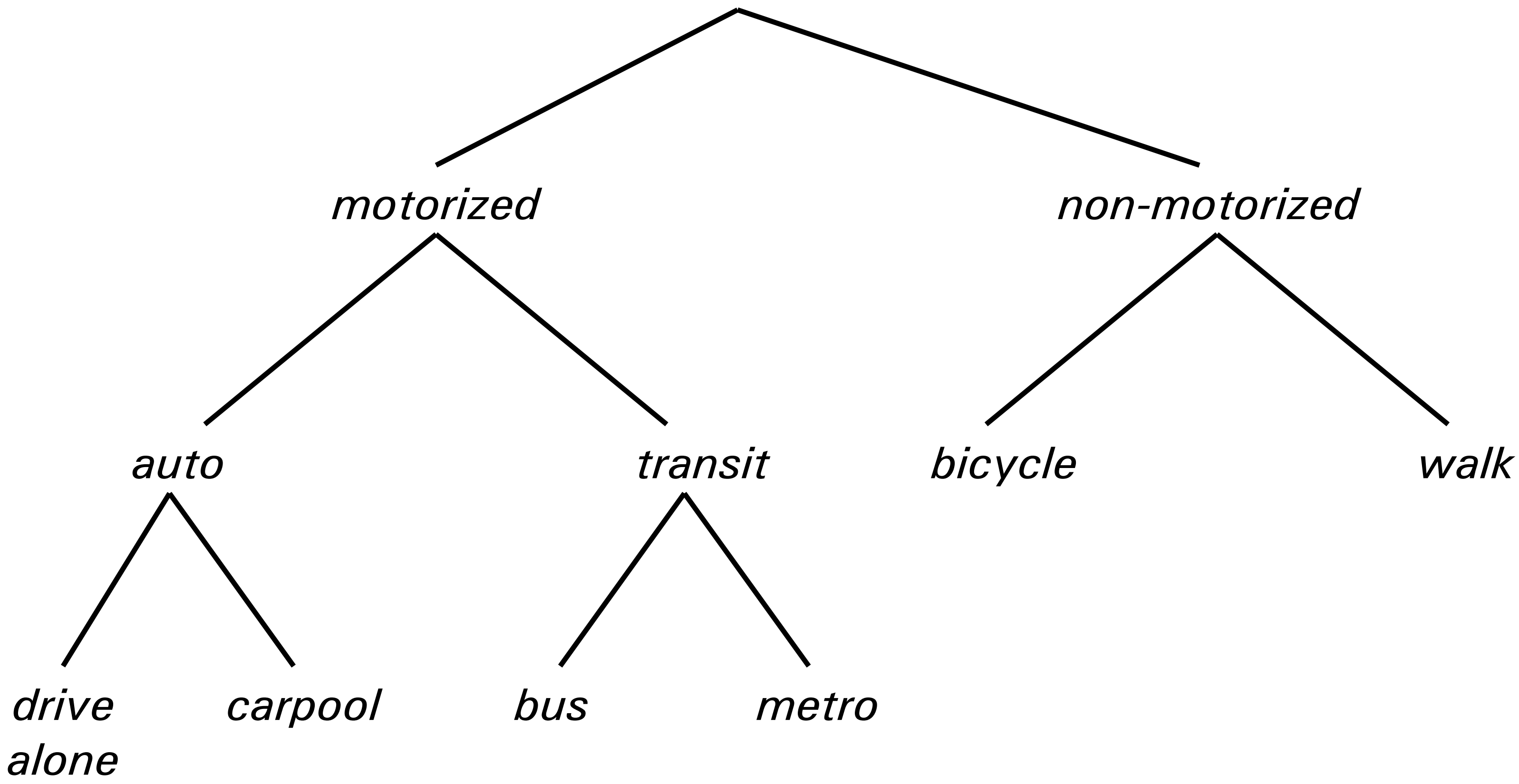
Population	Auto Share	Red Bus Share	Blue Bus Share
Auto people	81.8%	9.1%	9.1%
Transit people	5.2%	47.4%	47.4%
Total	**43.5%**	**28.25%**	**28.25%**

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)

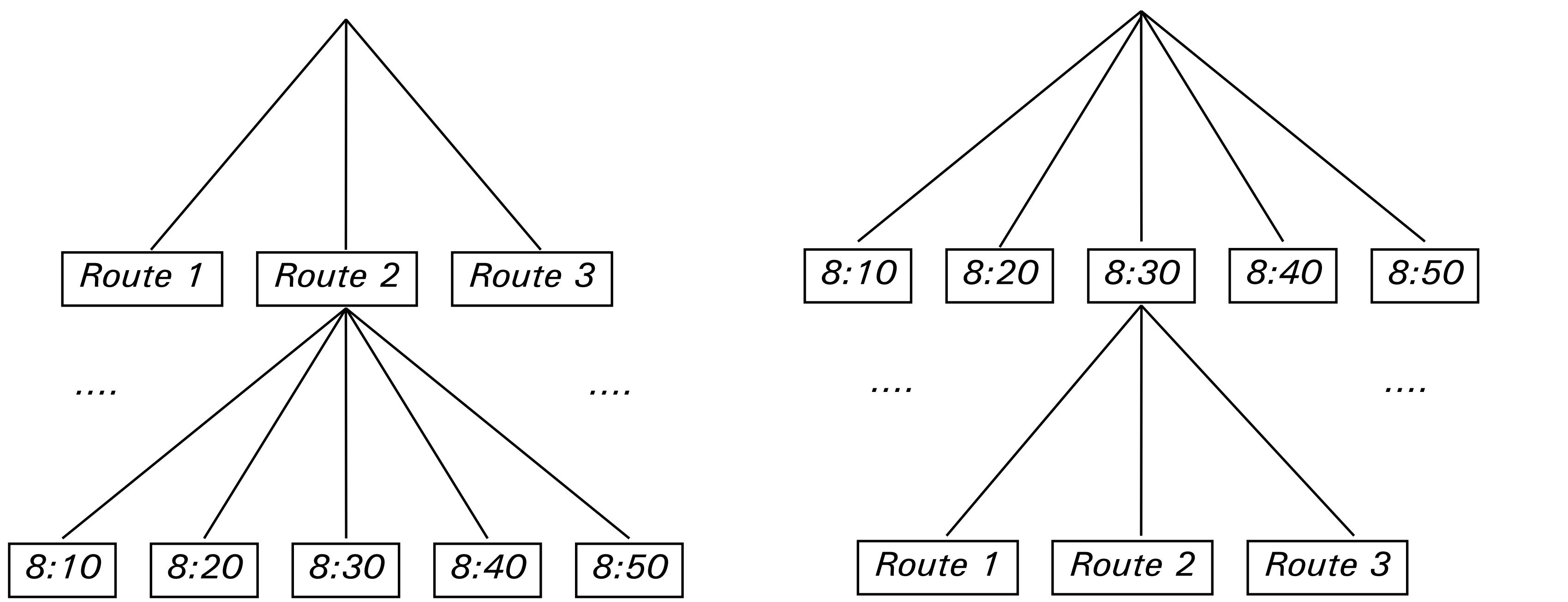
Tree Representation of Nested Logit

- Example: Mode Choice (Correlated Alternatives)



Tree Representation of Nested Logit

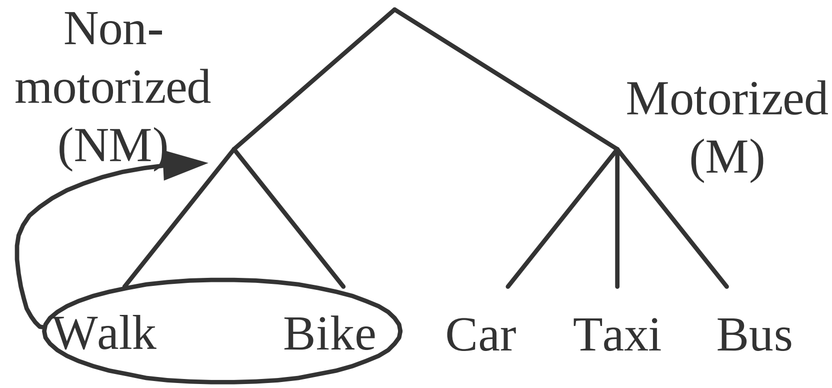
- Example: Route and Departure Time Choice (Multidimensional Choice)



Nested Model Estimation

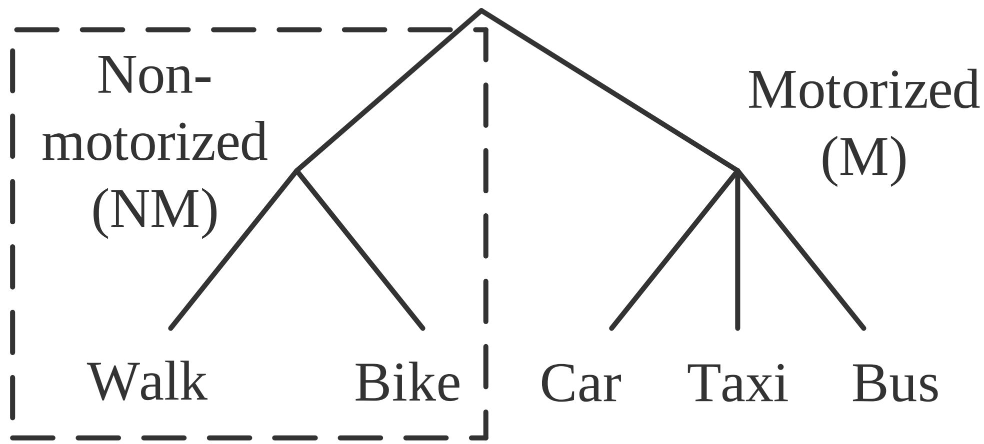
- Logit at each node
- Utilities at lower level enter at the node as the *inclusive* value

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



- The inclusive value is often referred to as *logsum*

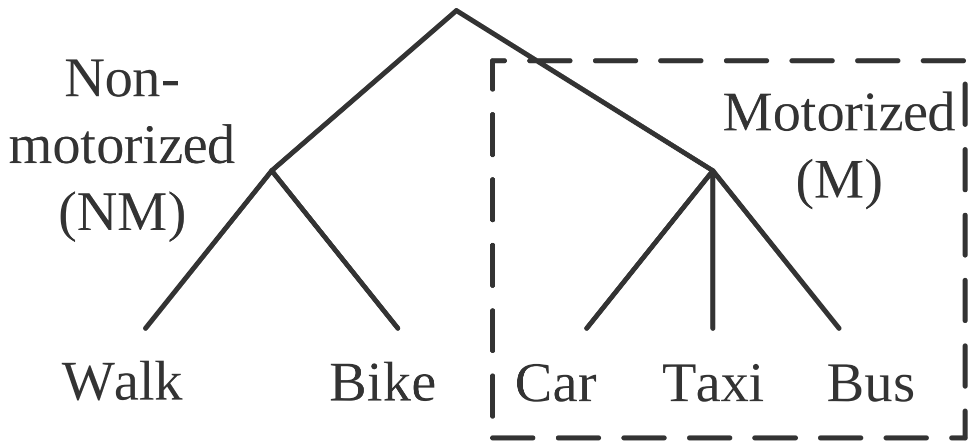
Nested Model – Example



$$P(i \mid NM) = \frac{e^{\mu_{NM} V_i}}{e^{\mu_{NM} V_{Walk}} + e^{\mu_{NM} V_{Bike}}} \qquad i = Walk, Bike$$

$$I_{NM} = \frac{1}{\mu_{NM}} \ln\left(e^{\mu_{NM} V_{Walk}} + e^{\mu_{NM} V_{Bike}}\right)$$

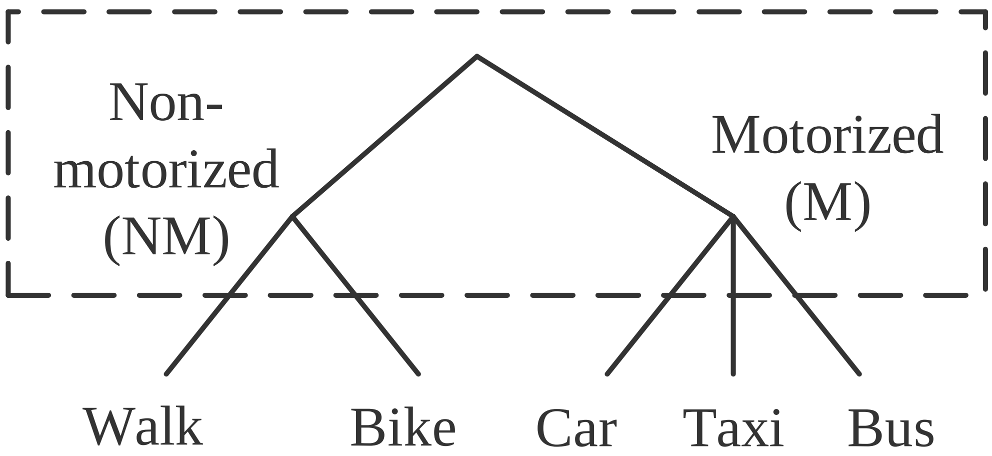
Nested Model – Example



$$\mathrm{P}(i \mid M) = \frac{e^{\mu_M V_i}}{e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}}} \qquad i = Car, Taxi, Bus$$

$$I_M = \frac{1}{\mu_M} \ln(e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}})$$

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

$$P(Bus) = P(Bus|M) \cdot P(M)$$

$$= \left[\frac{e^{\mu_M V_{Bus}}}{e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}}} \right] \cdot \left[\frac{e^{\mu I_M}}{e^{\mu I_{NM}} + e^{\mu I_M}} \right]$$

$$= \left[\frac{e^{\mu_M V_{Bus}}}{e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}}} \right] \cdot \left[\frac{e^{\frac{\mu}{\mu_M} \ln(e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}})}}{e^{\frac{\mu}{\mu_{NM}} \ln(e^{\mu_{NM} V_{Walk}} + e^{\mu_{NM} V_{Bike}})} + e^{\frac{\mu}{\mu_M} \ln(e^{\mu_M V_{Car}} + e^{\mu_M V_{Taxi}} + e^{\mu_M V_{Bus}})}} \right]$$

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

- Ben-Akiva, M. and Bierlaire, M. (2003), ‘Discrete Choice Models With Applications to Departure Time and Route Choice,’ *The Handbook of Transportation Science,* 2nd ed., (eds.) R.W. Hall, Kluwer, pp. 7 – 38.
- Ben-Akiva, M. and Lerman, S. (1985), *Discrete Choice Analysis*, MIT Press, Cambridge, Massachusetts.
- Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, United Kingdom.

- 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:

$$C_{mn} \cap C_{m'n} = \varnothing \quad \forall m \neq m'$$

- 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: $P(i \mid C_n) = P(C_{mn} \mid C_n) P(i \mid C_{mn}), i \in C_{mn} \subseteq C_n$

  where

$$P(C_{mn} \mid C_n) = \frac{e^{\mu V_{C_{mn}}}}{\sum_l e^{\mu V_{C_{ln}}}} \quad \text{and} \quad P(i \mid C_{mn}) = \frac{e^{\mu_m \tilde{V}_{in}}}{\sum_{j \in C_{mn}} e^{\mu_m \tilde{V}_{jn}}}$$

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:

$$
\begin{aligned}
U_{auto} &= \beta X_{auto} + \xi_{auto} + \nu_{auto} \\
U_{bus} &= \beta X_{bus} + \xi_{bus} + \nu_{bus} \\
U_{subway} &= \beta X_{subway} + \xi_{subway} + \nu_{subway}
\end{aligned}
$$

ε — ν *i.i.d. Extreme Value*

e.g. $\xi \sim N(0,\Sigma)$

- Probability:

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

ξ *unknown* →

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

Continuous Logit Mixture

Random Taste Variation

- Logit: β is a constant vector
  - Can segment, e.g. $\beta_{low\ inc}$, $\beta_{med\ inc}$, $\beta_{high\ inc}$
- Logit Mixture: β 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 *latent classes*

Latent Class Choice Model

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

Class-specific Choice Model

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

Class Membership Model

(probability of belonging to class s)

Summary of Discrete Choice Models

	Logit	NL/CNL	Probit	Logit Mixture
Handles unobserved taste heterogeneity	No	No	Yes	Yes
Flexible substitution pattern	No	Yes	Yes	Yes
Handles panel data	No	No	Yes	Yes
Requires error terms normally distributed	No	No	Yes	No
Closed-form choice probabilities available	Yes	Yes	No	No (cont.) Yes (discrete)
Numerical approximation and/or simulation needed	No	No	Yes	Yes (cont.) No (discrete)

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