Logit with multiple alternatives

Matthieu de Lapparent
matthieu.delapparent@epfl.ch

Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne

Outline

Outline

Random Utility

For each $i \in \mathcal{C}_n$

$$U_{in} = V_{in} + \varepsilon_{in}$$

- What is $\mathcal{C}_n$?
- What is V_{in}?
- What is ε_{in}?

Choice set

Universal choice set $\mathcal{C}$

- All potential alternatives for the population
- Restricted to relevant alternatives

Mode choice:

driving alone	sharing a ride	taxi
motorcycle	bicycle	walking
bus	rail rapid transit	horse

Choice set

Individual's choice set

- No driver's license
- No auto
- Awareness of bus services
- Rail transit services unreachable
- Walking not an option for long distance

Individual's mode choice

- ~~driving alone~~
- sharing a ride
- taxi
- motorcycle
- bicycle
- ~~walking~~
- bus
- ~~rail rapid transit~~
- ~~horse~~

Choice set

Choice set generation is tricky

- How to model "awareness"?
- What does "unreachable" mean exactly?
- What does "long distance" mean exactly?

We will continue assuming a deterministic rule

Error terms

Main assumption

ε_{in} are

- extreme value EV(0,μ),
- independent and
- identically distributed.

Comments

- Independence: across i and n.
- Identical distribution: same scale parameter μ across i and n.
- Scale must be normalized, e.g. $\mu = 1$

Illustration of μ: A rising tide lifts all boats

Derivation of the logit model

Reminder: binary case

- $\mathcal{C}_n = \{i, j\}$
- $U_{in} = V_{in} + \varepsilon_{in}$
- $\varepsilon_{in} \sim \text{EV}(0, \mu)$
- ε_{in} i.i.d.

Choice model

$$P(i|\mathcal{C}_n = \{i, j\}) = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}}$$

Derivation of the logit model

Multiple alternatives

- $\mathcal{C}_n = \{1, \ldots, J_n\}$
- $U_{in} = V_{in} + \varepsilon_{in}$
- $\varepsilon_{in} \sim \text{EV}(0, \mu)$
- ε_{in} i.i.d.

Choice model

$$P(i|\mathcal{C}_n) = P(V_{in} + \varepsilon_{in} \geq \max_{j=1,\ldots,J_n} V_{jn} + \varepsilon_{jn})$$

Assume without loss of generality (wlog) that $i = 1$

$$P(1|\mathcal{C}_n) = P(V_{1n} + \varepsilon_{1n} \geq \max_{j=2,\ldots,J_n} V_{jn} + \varepsilon_{jn})$$

Derivation of the logit model

Composite alternative

- Define a composite alternative as "anything but alternative one"
- Associated utility:

$$U^* = \max_{j=2,\ldots,J_n} (V_{jn} + \varepsilon_{jn})$$

- From a property of the EV distribution

$$U^* \sim \text{EV}\left(\frac{1}{\mu} \ln \sum_{j=2}^{J_n} e^{\mu V_{jn}}, \mu\right)$$

Derivation of the logit model

- From another property of the EV distribution

$$U^* = V^* + \varepsilon^*$$

where

$$V^* = \frac{1}{\mu} \ln \sum_{j=2}^{J_n} e^{\mu V_{jn}}$$

and

$$\varepsilon^* \sim \mathrm{EV}(0, \mu)$$

Derivation of the logit model

- Therefore

$$\begin{aligned} P(1|\mathcal{C}_n) &= P(V_{1n} + \varepsilon_{1n} \geq \max_{j=2,\ldots,J_n} V_{jn} + \varepsilon_{jn}) \\ &= P(V_{1n} + \varepsilon_{1n} \geq V^* + \varepsilon^*) \end{aligned}$$

- This is a binary choice model with a systematic composite alternative

$$P(1|\mathcal{C}_n) = \frac{e^{\mu V_{1n}}}{e^{\mu V_{1n}} + e^{\mu V^*}}$$

where

$$V^* = \frac{1}{\mu} \ln \sum_{j=2}^{J_n} e^{\mu V_{jn}}$$

Derivation of the logit model

- and can be rewritten as

$$
\begin{aligned}
P(1|\mathcal{C}_n) &= \frac{e^{\mu V_{1n}}}{e^{\mu V_{1n}} + e^{\mu V^*}} \\
&= \frac{e^{\mu V_{1n}}}{e^{\mu V_{1n}} + \sum_{j=2}^{J_n} e^{\mu V_{jn}}} \\
&= \frac{e^{\mu V_{1n}}}{\sum_{j=1}^{J_n} e^{\mu V_{jn}}}
\end{aligned}
$$

Derivation of the logit model

- The scale parameter μ is not identifiable: $\mu = 1$.
- Warning: not identifiable $\neq$ not existing
- Limiting cases
  - $\mu \to 0$, that is variance goes to infinity

$$\lim_{\mu \to 0} P(i|\mathcal{C}_n) = \frac{1}{J_n} \quad \forall i \in \mathcal{C}_n$$

  - $\mu \to +\infty$, that is variance goes to zero

$$\begin{aligned} \lim_{\mu\to\infty} P(i|C_n) &= \lim_{\mu\to\infty} \frac{1}{1+\sum_{j\neq i} e^{\mu(V_{jn}-V_{in})}} \\ &= \begin{cases} 1 & \text{if } V_{in} > \max_{j\neq i} V_{jn} \\ 0 & \text{if } V_{in} < \max_{j\neq i} V_{jn} \end{cases} \end{aligned}$$

Another derivation of the Multinomial logit model

$$P(i|\mathcal{C}_n) = P(j \in \mathcal{C}_n, j \neq i, V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn})$$

$$P(i|\mathcal{C}_n) = P(j \in \mathcal{C}_n, j \neq i, V_{in} - V_{jn} + \varepsilon_{in} \geq \varepsilon_{jn})$$

$$P(i|\mathcal{C}_n) = \int_{-\infty}^{+\infty} \left[\prod_{j \in \mathcal{C}_n, j \neq i} \int_{-\infty}^{V_{in} - V_{jn} + \varepsilon_{in}} f(\varepsilon_{jn}) d\varepsilon_{jn} \right] f(\varepsilon_{in}) d\varepsilon_{in}$$

$$P(i|\mathcal{C}_n) = \int_{-\infty}^{+\infty} \prod_{j \in \mathcal{C}_n, j \neq i} e^{-e^{V_{jn} - V_{in} - \varepsilon_{in}}} f(\varepsilon_{in}) d\varepsilon_{in}$$

Another derivation of the Multinomial logit model, cont.

$$P(i|\mathcal{C}_n) = \int_{-\infty}^{+\infty} e^{-\sum_{j\in\mathcal{C}_n, j\neq i} e^{V_{jn}-V_{in}-\varepsilon_{in}}} f(\varepsilon_{in}) d\varepsilon_{in}$$

$$P(i|\mathcal{C}_n) = \int_{-\infty}^{+\infty} e^{-e^{-\varepsilon_{in}} \sum_{j\in\mathcal{C}_n, j\neq i} e^{V_{jn}-V_{in}}} e^{-\varepsilon_{in}} e^{-e^{-\varepsilon_{in}}} d\varepsilon_{in}$$

$$y_{in} = e^{-\varepsilon_{in}}, dy_{in} = -e^{-\varepsilon_{in}} d\varepsilon_{in}, y_{in} \to]0; +\infty[$$

$$P(i|\mathcal{C}_n) = \int_{0}^{+\infty} e^{-y_{in}\left(1+\sum_{j\in\mathcal{C}_n, j\neq i} e^{V_{jn}-V_{in}}\right)} dy_{in}$$

$$P(i|\mathcal{C}_n) = \frac{1}{1+\sum_{j\in\mathcal{C}_n, j\neq i} e^{V_{jn}-V_{in}}} = \frac{e^{V_{in}}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}}}$$

Outline

Systematic part of the utility function

Shift focus from ε to V

$$V_{in} = V(z_{in}, S_n)$$

where

- z_{in} is a vector of attributes of alternative i for individual n
- S_n is a vector of socio-economic characteristics of n

Outline:

- Functional form: linear utility
- Explanatory variables: What exactly is contained in z_{in} and S_n?
- Functional form: capturing nonlinearities
- Interactions

Functional form: linear utility

Notation for explanatory variables

$$x_{in} = (z_{in}, S_n)$$

In general, linear-in-parameters utility functions are used

$$V_{in} = V(z_{in}, S_n) = V(x_{in}) = \sum_k \beta_k (x_{in})_k$$

Not as restrictive as it may seem

Explanatory variables: attributes of alternatives

Numerical and continuous

- $(z_{in})_k \in \mathbb{R}, \forall i, n, k$
- Associated with a specific unit

Examples

- Auto in-vehicle time (in min.)
- Transit in-vehicle time (in min.)
- Auto out-of-pocket cost (in cents)
- Transit fare (in cents)
- Walking time to the bus stop (in min.)

Straightforward modeling

Explanatory variables: attributes of alternatives

- V_{in} is unitless
- Therefore, β depends on the unit of the associated attribute
- Example: consider two specifications

$$\begin{aligned} V_{in} &= \beta_1 \mathrm{TT}_{in} + \cdots \\ V_{in} &= \beta_1' \mathrm{TT}_{in}' + \cdots \end{aligned}$$

- If TT_{in} is measured in minutes, the unit of β_1 is 1/min
- If TT_{in}' is measured in hours, the unit of β_1' is 1/hour
- Both models are equivalent, but the estimated β will be scaled differently

$$\beta_1 \mathrm{TT}_{in} = \beta_1' \mathrm{TT}_{in}' \Longrightarrow \frac{\mathrm{TT}_{in}}{\mathrm{TT}_{in}'} = \frac{\beta_1'}{\beta_1} = 60$$

Explanatory variables: attributes of alternatives

Impact of attributes on different alternatives

- Generic, or

$$\begin{aligned} V_{\text{auto}} &= \beta_1 \text{TT}_{\text{auto}} \\ V_{\text{bus}} &= \beta_1 \text{TT}_{\text{bus}} \end{aligned}$$

- Alternative specific parameters

$$\begin{aligned} V_{\text{auto}} &= \beta_1 \text{TT}_{\text{auto}} \\ V_{\text{bus}} &= \beta_2 \text{TT}_{\text{bus}} \end{aligned}$$

Modeling assumption: a minute has/doesn't have the same marginal utility whether it is incurred on the auto or bus mode

Explanatory variables: socio-eco. characteristics

Numerical and continuous

- Numerical and continuous
- $(S_n)_k \in \mathbb{R}, \forall n, k$
- Associated with a specific unit

Examples

- Annual income (in KCHF)
- Age (in years)

Warning: S_n do not depend on i

Explanatory variables: socio-eco. characteristics

They cannot appear in all utility functions

$$\left.\begin{array}{rcl} V_1 & = & \beta_1 x_{11} + \beta_2 \text{income} \\ V_2 & = & \beta_1 x_{21} + \beta_2 \text{income} \\ V_3 & = & \beta_1 x_{31} + \beta_2 \text{income} \end{array}\right\} \Longleftrightarrow \left\{\begin{array}{rcl} V_1' & = & \beta_1 x_{11} \\ V_2' & = & \beta_1 x_{21} \\ V_3' & = & \beta_1 x_{31} \end{array}\right.$$

Need to specify as alternative specific, e.g.

$$\begin{array}{rcll} V_1 & = & \beta_1 x_{11} + \beta_2 \text{income} & +\beta_4 \text{age} \\ V_2 & = & \beta_1 x_{21} + \beta_3 \text{income} & +\beta_5 \text{age} \\ V_3 & = & \beta_1 x_{31} & \end{array}$$

Functional form: dealing with nonlinearities

- Discrete and qualitative variables
- Continuous variables
  - Categories
  - Splines
  - Box-Cox
  - Power series

Discrete variables

Mainly used to capture impact of qualitative attributes

- Level of comfort for the train
- Reliability of the bus
- Color of car
- etc...

or discrete characteristics

- Sex
- Education
- Professional status
- etc.

Discrete variables

Procedure for model specification

- Identify all possible levels of the attribute:
  - Very high comfort (V),
  - High comfort (H),
  - Moderate comfort (M),
  - Low comfort (L)
- Select a base case: Very high comfort
- Define numerical attributes
- Adopt a coding convention

Discrete variables

Introduce a 0/1 attribute code for all levels except the base case

- z_H for High comfort
- z_M for Moderate comfort
- z_L for Low comfort

	z_H	z_M	z_L
Very high comfort	0	0	0
High comfort	1	0	0
Moderate comfort	0	1	0
Low comfort	0	0	1

If a qualitative attribute has n levels, we introduce $n-1$ (0/1) variables in the model

Comparing two coding conventions

Very high comfort fixed as base

$$V = \cdots + \beta_\text{V} z_\text{V} + \beta_\text{H} z_\text{H} + \beta_\text{M} z_\text{M} + \beta_\text{L} z_\text{L} \quad \text{where } \beta_\text{V} = 0$$

- β_H: difference of utility between high comfort and very high comfort (supposedly negative)
- β_M: difference of utility between moderate comfort and very high comfort (supposedly more negative)
- β_L: difference of utility between low comfort and very high comfort (supposedly even more negative)

Comparing two ways of coding

High comfort fixed as base

$$V' = \cdots + \beta_V z_V + \beta_H z_H + \beta_M z_M + \beta_L z_L \quad \text{where } \beta_H = 0$$

- β'_V: difference of utility between very high comfort and high comfort (supposedly positive)
- β'_M: difference of utility between moderate comfort and high comfort (supposedly negative)
- β'_L: difference of utility between low comfort and high comfort (supposedly more negative)

Discrete variables

Example of estimation with Biogeme:

	Model 1	Model 2
ASC	0.574	0.574
BETA_V	0.000	0.918
BETA_H	-0.919	0.000
BETA_M	-1.015	-0.096
BETA_L	-2.128	-1.210

Nonlinear transformations of the variables

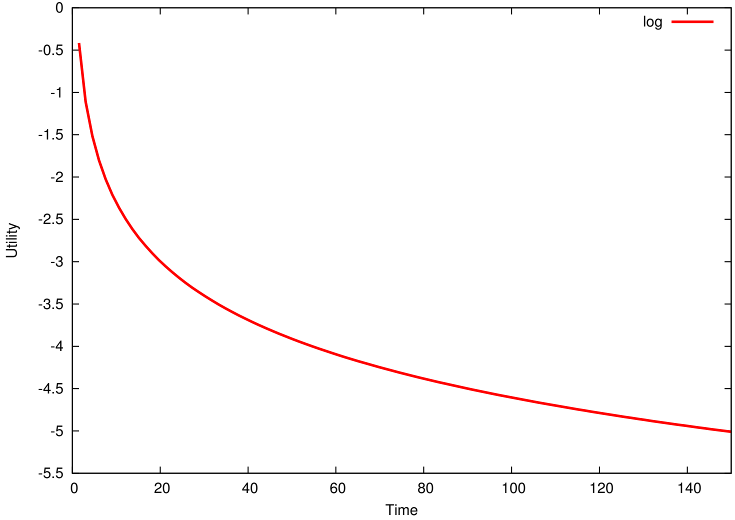
Example with travel time

- Compare a trip of 5 min with a trip of 10 min (+5 minutes)
- Compare a trip of 120 min with a trip of 125 min (+5 minutes)

Behavioral assumption

One additional minute of travel time is not perceived in the same way for short trips as for long trips

Nonlinear transformations of the variables



Nonlinear transformations of the variables

Assumption 1: the marginal impact of travel time is constant

$$V_i = \beta_T \text{time}_i + \cdots$$

Assumption 2: the marginal impact of travel time decreases with longer travel time

$$V_i = \beta_T \ln(\text{time}_i) + \cdots$$

Remarks

- Still a linear-in-parameters form
- The unit, the value, and the interpretation of β_T is different

Continuous variables: split into categories

Like earlier assumption: sensitivity to travel time varies with travel time level

- Logarithmic transformation not the only specification
- Another possibility is to split travel time into categories (here TT in minutes)
  - Short: 0-90 min
  - Medium: 91 - 180 min
  - Long: 181 - 270 min
  - Very long: over 271 min

Possible specifications

- Categories with constants (inferior solution)
- Piecewise linear specification (spline)

Continuous variables: categories with constants

Same specification as for discrete variables

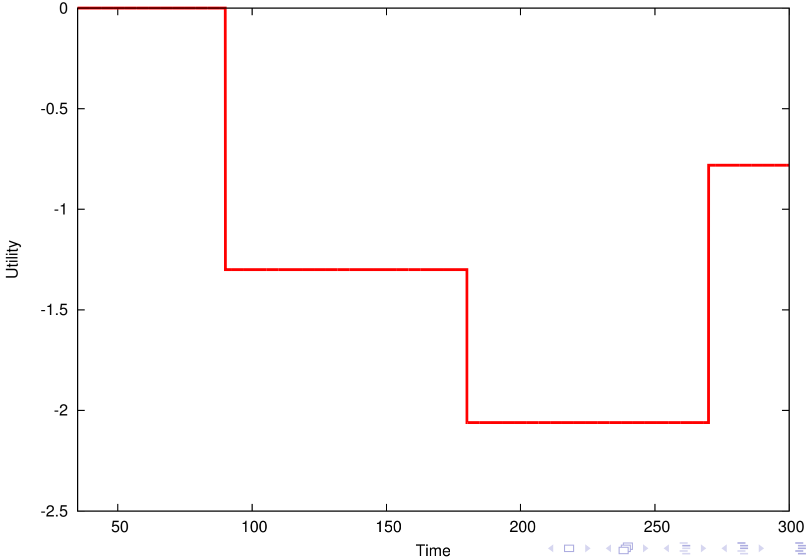
$$V_i = \beta_{T1}x_{T1} + \beta_{T2}x_{T2} + \beta_{T3}x_{T3} + \beta_{T4}x_{T4} + \ldots$$

with

- $x_{T1} = 1$ if $TT_i \in$ [0–90[, 0 otherwise
- $x_{T2} = 1$ if $TT_i \in$ [91–180[, 0 otherwise
- $x_{T3} = 1$ if $TT_i \in$ [181–270[, 0 otherwise
- $x_{T4} = 1$ if $TT_i \in$ [271–[, 0 otherwise

One β must be normalized to 0.

Continuous variables: categories with constants



Continuous variables: categories with constants

Drawbacks

- No sensitivity to travel time within the intervals
- Discontinuous utility function (jumps)
- Need for many small intervals
- Results may vary significantly with the definition of the intervals

Appropriate when

- Categories have been used in the survey (income, age)
- Definition of categories is natural (weekday)

Continuous variables: Piecewise linear specification

Piecewise linear specification (spline)

- Captures the sensitivity within the intervals
- Enforces continuity of the utility function

Piecewise linear specification

Features

- Capture the sensitivity within the intervals
- Enforce continuity of the utility function

$$V_i = \beta_{T1} x_{T1} + \beta_{T2} x_{T2} + \beta_{T3} x_{T3} + \beta_{T4} x_{T4} + \ldots$$

where

$$x_{T1} = \begin{cases} t & \text{if } t < 90 \\ 90 & \text{otherwise} \end{cases} \qquad x_{T2} = \begin{cases} 0 & \text{if } t < 90 \\ t - 90 & \text{if } 90 \leq t < 180 \\ 90 & \text{otherwise} \end{cases}$$

$$x_{T3} = \begin{cases} 0 & \text{if } t < 180 \\ t - 180 & \text{if } 180 \leq t < 270 \\ 90 & \text{otherwise} \end{cases} \qquad x_{T4} = \begin{cases} 0 & \text{if } t < 270 \\ t - 270 & \text{otherwise} \end{cases}$$

Piecewise linear specification

Note: coding in Biogeme for interval [a:a+b[

$$x_{Ti} = \begin{cases} 0 & \text{if } t < a \\ t - a & \text{if } a \leq t < a + b \\ b & \text{otherwise} \end{cases} \qquad x_{Ti} = \max(0, \min(t - a, b))$$

$$\begin{aligned} x_{T1} &= \min(t, 90) \\ x_{T2} &= \max(0, \min(t - 90, 90)) \\ x_{T3} &= \max(0, \min(t - 180, 90)) \\ x_{T4} &= \max(0, t - 270) \end{aligned}$$

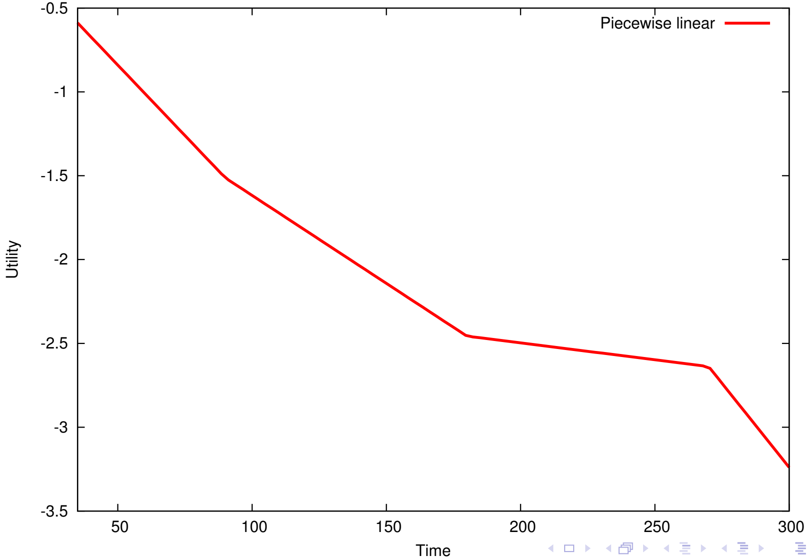
```
TRAIN_TT1  =  min( TRAIN_TT , 90)
TRAIN_TT2  =  max(0,min( TRAIN_TT - 90, 90))
TRAIN_TT3  =  max(0,min( TRAIN_TT - 180 , 90))
TRAIN_TT4  =  max(0,TRAIN_TT - 270)
```

Piecewise linear specification

Examples:

t	TT1	TT2	TT3	TT4
40	40	0	0	0
100	90	10	0	0
200	90	90	20	0
300	90	90	90	30

Piecewise linear specification



Continuous variables: Box-Cox transforms

Box-Cox transform

Box and Cox, J. of the Royal Statistical Society (1964)

$$V_i = \beta x_i(\lambda) + \cdots$$

where

$$x_i(\lambda) = \begin{cases} \dfrac{x_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln x_i & \text{if } \lambda = 0. \end{cases}$$

where $x_i > 0$.

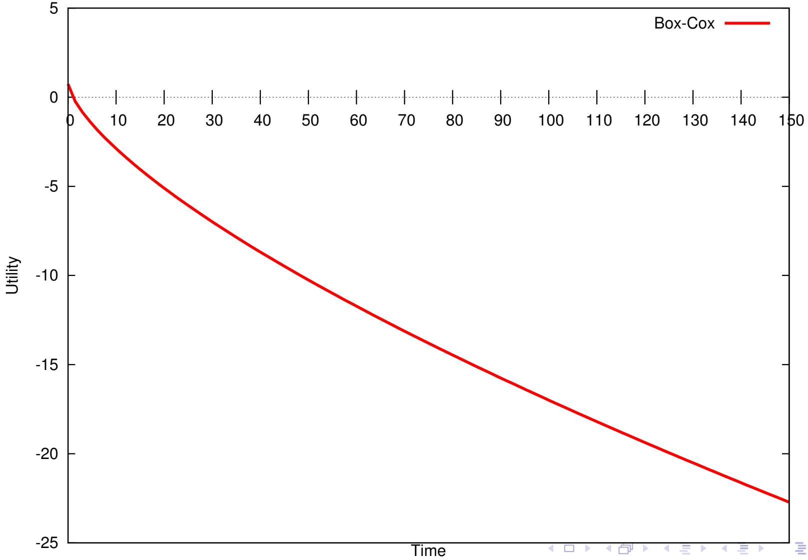
Box-Cox transforms

Box-Tukey transform

If $x_i \leq 0$, include a constant α such that $x_i + \alpha > 0$ and

$$x_i(\lambda, \alpha) = \begin{cases} \dfrac{(x_i + \alpha)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(x_i + \alpha) & \text{if } \lambda = 0. \end{cases}$$

Box-Cox transforms ($\lambda = 0.7$)



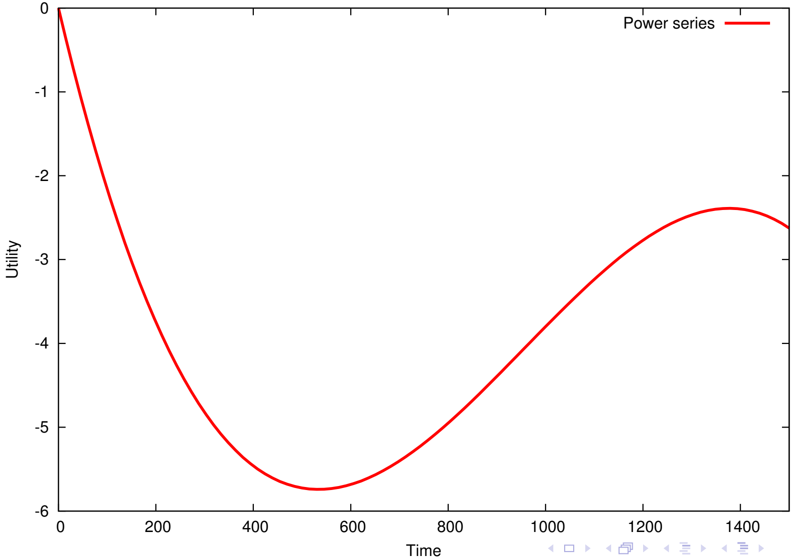
Power series

Taylor expansion

$$V_i = \beta_1 T + \beta_2 T^2 + \beta_3 T^3 + \dots$$

- In practice, these terms can be very correlated
- Difficult to interpret
- Risk of over fitting

Power series



Interactions

- All individuals in a population are not alike
- Socio-economic characteristics define segments in the population
- How to capture heterogeneity?
  - Interactions of attributes and characteristics
  - Discrete segmentation
  - Continuous segmentation

Interactions of attributes and characteristics

Combination of attributes

- cost / income
- fare / disposable income
- distance / out-of-vehicle time (=speed)

warning: correlation of attributes may produce degeneracy in the model

Interactions: discrete segmentation

Example with discrete segments

- Hypothesis: different sensitivities for combinations of:
  - Gender (M,F)
  - House location (metro, suburb, periphery areas)
- Each individual belongs to exactly one of the 6 segments
- Specification of 6 segments

$$\beta_{M,m} TT_{M,m} + \beta_{M,s} TT_{M,s} + \beta_{M,p} TT_{M,p} +$$
$$\beta_{F,m} TT_{F,m} + \beta_{F,s} TT_{F,s} + \beta_{F,p} TT_{F,p} +$$

- $TT_i = TT$ if indiv. belongs to segment i, and 0 otherwise

Interactions: continuous segmentation

Example with continuous characteristics

- Hypothesis: the cost parameter varies with income

$$\beta_{\text{cost}} = \hat{\beta}_{\text{cost}} \left(\frac{\text{inc}}{\text{inc}_{\text{ref}}} \right)^{\lambda} \text{ with } \lambda = \frac{\partial \beta_{\text{cost}}}{\partial \text{inc}} \frac{\text{inc}}{\beta_{\text{cost}}}$$

- Reference value is arbitrary
- Several characteristics can be combined:

$$\beta_{\text{cost}} = \hat{\beta}_{\text{cost}} \left(\frac{\text{inc}}{\text{inc}_{\text{ref}}} \right)^{\lambda_1} \left(\frac{\text{age}}{\text{age}_{\text{ref}}} \right)^{\lambda_2}$$

warning: λ must be estimated and utility is no longer linear-in-parameters

Heteroscedasticity

Assumption: variance of error terms is different across individuals

Assume there are two different groups such that

$$\begin{aligned} U_{in_1} &= V_{in_1} + \varepsilon_{in_1} \\ U_{in_2} &= V_{in_2} + \varepsilon_{in_2} \end{aligned}$$

and $\text{var}(\varepsilon_{in_2}) = \alpha^2 \text{var}(\varepsilon_{in_1})$

Logit is homoscedastic

- ε_{in} i.i.d. across both i and n.
- How can we specify the model in order to use logit?

Motivation

- People have different level of knowledge (e.g. taxi drivers)
- Different sources of data

Heteroscedasticity

Solution: include scale parameters

$$\begin{array}{rcccccccc} \alpha U_{in_1} & = & \alpha V_{in_1} & + & \alpha\varepsilon_{in_1} & = & \alpha V_{in_1} & + & \varepsilon'_{in_1} \\ U_{in_2} & = & V_{in_2} & + & \varepsilon_{in_2} & = & V_{in_2} & + & \varepsilon'_{in_2} \end{array}$$

where ε'_{in_1} and ε'_{in_2} are i.i.d.

Remarks

- Even if $V_{in_1} = \sum_j \beta_j x_{jin_1}$ is linear-in-parameters, $\alpha V_{in_1} = \sum_j \alpha\beta_j x_{jin_1}$ is not.
- Normalization: a different scale parameter can be estimated for each segment of the population, except one that must be normalized.

Outline

A case study

Choice of residential telephone services

- Household survey conducted in Pennsylvania, USA, 1984
- Revealed preferences
- 434 observations

A case study

Telephone services and availability

	metro, suburban & some perimeter areas	other perimeter areas	non-metro areas
Budget Measured	yes	yes	yes
Standard Measured	yes	yes	yes
Local Flat	yes	yes	yes
Extended Area Flat	no	yes	no
Metro Area Flat	yes	yes	no

A case study

Universal choice set

$$\mathcal{C} = \{\text{BM}, \text{SM}, \text{LF}, \text{EF}, \text{MF}\}$$

Specific choice sets

- Metro, suburban & some perimeter areas: {BM,SM,LF,MF}
- Other perimeter areas: $\mathcal{C}$
- Non-metro areas: {BM,SM,LF}

A case study

Specification table

	β_1	β_2	β_3	β_4	β_5
BM	0	0	0	0	ln(cost(BM))
SM	1	0	0	0	ln(cost(SM))
LF	0	1	0	0	ln(cost(LF))
EF	0	0	1	0	ln(cost(EF))
MF	0	0	0	1	ln(cost(MF))

A case study

Utility functions

$$
\begin{array}{rcccl}
V_{\mathrm{BM}} & = & & & \beta_5 \ln(\mathrm{cost}_{\mathrm{BM}}) \\
V_{\mathrm{SM}} & = & \beta_1 & + & \beta_5 \ln(\mathrm{cost}_{\mathrm{SM}}) \\
V_{\mathrm{LF}} & = & \beta_2 & + & \beta_5 \ln(\mathrm{cost}_{\mathrm{LF}}) \\
V_{\mathrm{EF}} & = & \beta_3 & + & \beta_5 \ln(\mathrm{cost}_{\mathrm{EF}}) \\
V_{\mathrm{MF}} & = & \beta_4 & + & \beta_5 \ln(\mathrm{cost}_{\mathrm{MF}})
\end{array}
$$

A case study

Specification table II

	β_1	β_2	β_3	β_4	β_5	β_6	β_7
BM	0	0	0	0	ln(cost(BM))	users	0
SM	1	0	0	0	ln(cost(SM))	users	0
LF	0	1	0	0	ln(cost(LF))	0	1 if metro/suburb
EF	0	0	1	0	ln(cost(EF))	0	0
MF	0	0	0	1	ln(cost(MF))	0	0

A case study

Utility functions

$$
\begin{array}{rcccccccc}
V_{\text{BM}} & = & & & \beta_5 \ln(\text{cost}_{\text{BM}}) & + & \beta_6 \text{users} & & \\
V_{\text{SM}} & = & \beta_1 & + & \beta_5 \ln(\text{cost}_{\text{SM}}) & + & \beta_6 \text{users} & & \\
V_{\text{LF}} & = & \beta_2 & + & \beta_5 \ln(\text{cost}_{\text{LF}}) & & & + & \beta_7 \text{MS} \\
V_{\text{EF}} & = & \beta_3 & + & \beta_5 \ln(\text{cost}_{\text{EF}}) & & & & \\
V_{\text{MF}} & = & \beta_4 & + & \beta_5 \ln(\text{cost}_{\text{MF}}) & & & &
\end{array}
$$

Outline

Maximum likelihood estimation

Logit Model

$$P_n(i|\mathcal{C}_n) = \frac{e^{V_{in}}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}}}$$

Log-likelihood of a sample

$$\mathcal{L}(\beta_1,\ldots,\beta_K) = \sum_{n=1}^{N}\left(\sum_{j=1}^{J} y_{jn} \ln P_n(j|\mathcal{C}_n)\right)$$

where $y_{jn} = 1$ if ind. n has chosen alt. j, 0 otherwise

Maximum likelihood estimation

Logit model

$$\begin{aligned}\ln P_n(i|\mathcal{C}_n) &= \ln \frac{e^{V_{in}}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}}} \\ &= V_{in} - \ln(\textstyle\sum_{j\in\mathcal{C}_n} e^{V_{jn}})\end{aligned}$$

Log-likelihood of a sample for logit

$$\mathcal{L}(\beta_1,\ldots,\beta_K) = \sum_{n=1}^{N}\sum_{i=1}^{J} y_{in}\left(V_{in} - \ln \sum_{j\in\mathcal{C}_n} e^{V_{jn}}\right)$$

Maximum likelihood estimation

The maximum likelihood estimation problem

$$\max_{\beta \in \mathbb{R}^K} \mathcal{L}(\beta)$$

- Nonlinear optimization
- If the V's are linear-in-parameters, the function is concave

Maximum likelihood estimation

Numerical issue

$$P_n(i|\mathcal{C}_n) = \frac{e^{V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}}}$$

Largest value that can be stored in a computer $\approx 10^{308}$, that is

$$e^{709.783}$$

It is equivalent to compute

$$P_n(i|\mathcal{C}_n) = \frac{e^{V_{in}-V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}-V_{in}}} = \frac{1}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}-V_{in}}}$$

Outline

Simple models

Null model

$$U_i = \varepsilon_i \quad \forall i$$

$$P_n(i|\mathcal{C}_n) = \frac{e^{V_{in}}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}}} = \frac{e^0}{\sum_{j\in\mathcal{C}_n} e^0} = \frac{1}{\#\mathcal{C}_n}$$

$$\mathcal{L} = \sum_n \ln \frac{1}{\#\mathcal{C}_n} = -\sum_n \ln(\#\mathcal{C}_n)$$

Simple models

Constants only [Assume $\mathcal{C}_n = \mathcal{C}, \forall n$]

$$U_i = c_i + \varepsilon_i \quad \forall i$$

In the sample of size n, there are n_i persons choosing alt. i.

$$\ln P(i) = c_i - \ln(\sum_j e^{c_j})$$

If $\mathcal{C}_n$ is the same for all people choosing i, the log-likelihood for this part of the sample is

$$\mathcal{L}_i = n_i c_i - n_i \ln(\sum_j e^{c_j})$$

Simple models

Constants only (ctd)

The total log-likelihood is

$$\mathcal{L} = \sum_j n_j c_j - n \ln(\sum_j e^{c_j})$$

At the maximum, the derivatives must be zero

$$\frac{\partial \mathcal{L}}{\partial c_1} = n_1 - n\frac{e^{c_1}}{\sum_j e^{c_j}} = n_1 - nP(1) = 0.$$

Simple models

Constants only (ctd.)

Therefore,

$$P(1) = \frac{n_1}{n}$$

Conclusion

If all alternatives are always available, a model with only Alternative Specific Constants reproduces exactly the market shares in the sample

Back to the case study

Alt.	n_i	n_i/n	c_i	e^{c_i}	P(i)
BM	73	0.168	0.247	1.281	0.168
SM	123	0.283	0.769	2.158	0.283
LF	178	0.410	1.139	3.123	0.410
EF	3	0.007	-2.944	0.053	0.007
MF	57	0.131	0.000	1.000	0.131
	434	1.000			

Null-model: $\mathcal{L}$ = -434 ln(5) = -698.496

Warning: results have been obtained assuming that all alternatives are always available