Introduction to choice modeling

Michel Bierlaire

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

January 6, 2017

Outline

1. Motivation
2. Theoretical foundations
   - Decision maker
   - Characteristics
   - Choice set
   - Alternative attributes
   - Decision rule
3. Probabilistic choice theory
4. Binary choice
5. Model specification
   - Error term
6. Applying the model
7. Parameter estimation

Motivation

Human dimension in

- engineering
- business
- marketing
- planning
- policy making

Need for

- behavioral *theories*
- quantitative *methods*
- operational mathematical *models*

Motivation

Concept of demand

- marketing
- transportation
- energy
- finance

Concept of choice

- brand, product
- mode, destination
- type, usage
- buy/sell, product

Importance

Daniel L. McFadden

- UC Berkeley 1963, MIT 1977, UC Berkeley 1991
- Laureate of *The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 2000*
- Owns a farm and vineyard in Napa Valley
- “Farm work clears the mind, and the vineyard is a great place to prove theorems”

Choice theory

Choice: outcome of a sequential decision-making process

- defining the choice problem
- generating alternatives
- evaluating alternatives
- making a choice,
- executing the choice.

Theory of behavior that is

- descriptive: how people behave and not how they should
- abstract: not too specific
- operational: can be used in practice for forecasting

Building the theory

Define

1. who (or what) is the decision maker,
2. what are the characteristics of the decision maker,
3. what are the alternatives available for the choice,
4. what are the attributes of the alternatives, and
5. what is the decision rule that the decision maker uses to make a choice.

Decision maker

Individual

- a person
- a group of persons (internal interactions are ignored)
  - household, family
  - firm
  - government agency
- notation: n

Characteristics of the decision maker

Disaggregate models

Individuals

- face different choice situations
- have different tastes

Characteristics

- income
- sex
- age
- level of education
- household/firm size
- etc.

Alternatives

Choice set

- Non empty finite and countable set of alternatives
- Universal: $\mathcal{C}$
- Individual specific: $\mathcal{C}_n \subseteq \mathcal{C}$
- Availability, awareness

Example

Choice of a transportation model

- $\mathcal{C}$ ={car, bus, metro, walking }
- If the decision maker has no driver license, and the trip is 12km long

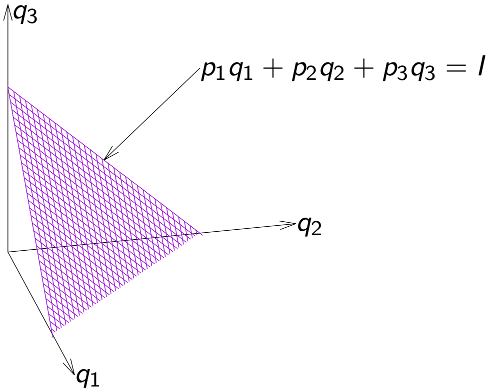
$$\mathcal{C}_n = \{\text{bus}, \text{metro}\}$$

Continuous choice set

Microeconomic demand analysis

Commodity bundle

- q_1: quantity of milk
- q_2: quantity of bread
- q_3: quantity of butter
- Unit price: p_i
- Budget: I

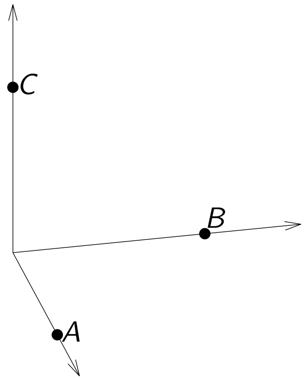


Discrete choice set

Discrete choice analysis

List of alternatives

- Brand A
- Brand B
- Brand C



Alternative attributes

Characterize each alternative i for each individual n

- price
- travel time
- frequency
- comfort
- color
- size
- etc.

Nature of the variables

- Discrete and continuous
- Generic and specific
- Measured or perceived

Decision rule

Homo economicus

Rational and narrowly self-interested economic actor who is optimizing her outcome

Utility

$$U_n : \mathcal{C}_n \longrightarrow \mathbb{R} : a \rightsquigarrow U_n(a)$$

- captures the attractiveness of an alternative
- measure that the decision maker wants to optimize

Behavioral assumption

- the decision maker associates a utility with each alternative
- the decision maker is a perfect optimizer
- the alternative with the highest utility is chosen

Simple example: mode choice

Attributes

	Attributes	
Alternatives	Travel time (t)	Travel cost (c)
Car (1)	t_1	c_1
Bus (2)	t_2	c_2

Utility

$$\widetilde{U} = \widetilde{U}(y_1, y_2),$$

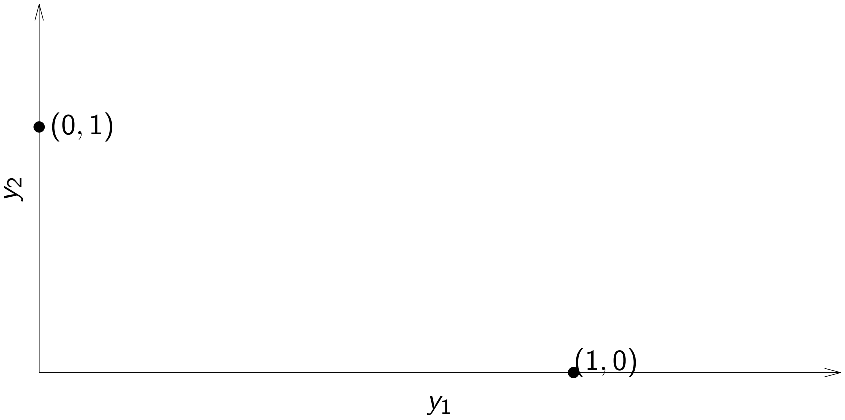
where we impose the restrictions that, for $i = 1, 2$,

$$y_i = \begin{cases} 1 & \text{if travel alternative i is chosen,} \\ 0 & \text{otherwise;} \end{cases}$$

and that only one alternative is chosen: $y_1 + y_2 = 1$.

Simple example: mode choice

Choice set



Simple example: mode choice

Utility functions

$$\begin{aligned} U_1 &= -\beta_t t_1 - \beta_c c_1, \\ U_2 &= -\beta_t t_2 - \beta_c c_2, \end{aligned}$$

where $\beta_t > 0$ and $\beta_c > 0$ are parameters.

Equivalent specification

$$\begin{aligned} U_1 &= -(\beta_t/\beta_c) t_1 - c_1 &= -\beta t_1 - c_1 \\ U_2 &= -(\beta_t/\beta_c) t_2 - c_2 &= -\beta t_2 - c_2 \end{aligned}$$

where $\beta > 0$ is a parameter.

Choice

- Alternative 1 is chosen if $U_1 \geq U_2$.
- Ties are ignored.

Simple example: mode choice

Choice

Alternative 1 is chosen if

$$-\beta t_1 - c_1 \geq -\beta t_2 - c_2$$

or

$$-\beta(t_1 - t_2) \geq c_1 - c_2$$

Alternative 2 is chosen if

$$-\beta t_1 - c_1 \leq -\beta t_2 - c_2$$

or

$$-\beta(t_1 - t_2) \leq c_1 - c_2$$

Dominated alternative

- If $c_2 > c_1$ and $t_2 > t_1$, $U_1 > U_2$ for any $\beta > 0$
- If $c_1 > c_2$ and $t_1 > t_2$, $U_2 > U_1$ for any $\beta > 0$

Simple example: mode choice

Trade-off

- Assume $c_2 > c_1$ and $t_1 > t_2$.
- Is the traveler willing to pay the extra cost $c_2 - c_1$ to save the extra time $t_1 - t_2$?
- Alternative 2 is chosen if

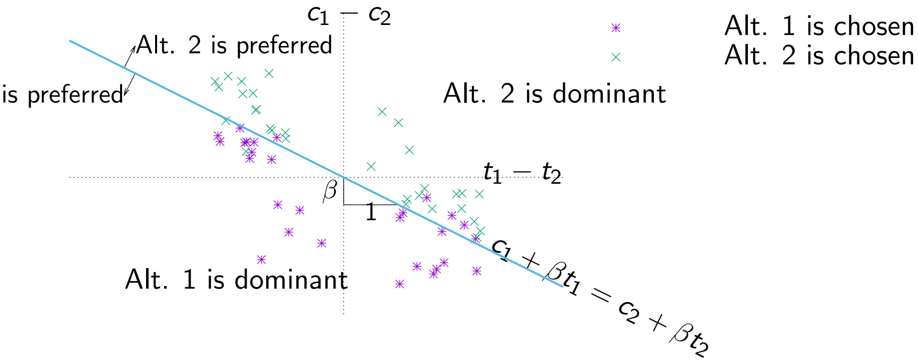
$$-\beta(t_1 - t_2) \leq c_1 - c_2$$

or

$$\beta \geq \frac{c_2 - c_1}{t_1 - t_2}$$

- β is called the *willingness to pay* or *value of time*

Simple example: mode choice



Outline

1. Motivation
2. Theoretical foundations
   - Decision maker
   - Characteristics
   - Choice set
   - Alternative attributes
   - Decision rule
3. **Probabilistic choice theory**
4. Binary choice
5. Model specification
   - Error term
6. Applying the model
7. Parameter estimation

Behavioral validity of the utility maximization?

Assumptions

Decision-makers

- are able to process information
- have perfect discrimination power
- have transitive preferences
- are perfect maximizer
- are always consistent

Relax the assumptions

Use a probabilistic approach: what is the probability that alternative i is chosen?

Introducing probability

Constant utility

- Human behavior is inherently random
- Utility is deterministic
- Consumer does not maximize utility
- Probability to use inferior alternative is non zero

Random utility

- Decision-maker are rational maximizers
- Analysts have no access to the utility used by the decision-maker
- Utility becomes a random variable

Niels Bohr

Nature is stochastic

Albert Einstein

God does not throw dice

Simple example

Choice

between *Auto* and *Transit*

Data

#	Time auto	Time transit	Choice	#	Time auto	Time transit	Choice
1	52.9	4.4	T	11	99.1	8.4	T
2	4.1	28.5	T	12	18.5	84.0	C
3	4.1	86.9	C	13	82.0	38.0	C
4	56.2	31.6	T	14	8.6	1.6	T
5	51.8	20.2	T	15	22.5	74.1	C
6	0.2	91.2	C	16	51.4	83.8	C
7	27.6	79.7	C	17	81.0	19.2	T
8	89.9	2.2	T	18	51.0	85.0	C
9	41.5	24.5	T	19	62.2	90.1	C
10	95.0	43.5	T	20	95.1	22.2	T
				21	41.6	91.5	C

Outline

1. Motivation
2. Theoretical foundations
   - Decision maker
   - Characteristics
   - Choice set
   - Alternative attributes
   - Decision rule
3. Probabilistic choice theory
4. Binary choice
5. Model specification
   - Error term
6. Applying the model
7. Parameter estimation

Binary choice model

Specification of the utilities

$$
\begin{aligned}
U_C &= \beta_1 T_C + \varepsilon_C \\
U_T &= \beta_1 T_T + \varepsilon_T
\end{aligned}
$$

where T_C is the travel time with car (min) and T_T the travel time with transit (min).

Choice model

$$
\begin{aligned}
P(C|\{C,T\}) &= \Pr(U_C \geq U_T) \\
&= \Pr(\beta_1 T_C + \varepsilon_C \geq \beta_1 T_T + \varepsilon_T) \\
&= \Pr(\beta_1 (T_C - T_T) \geq \varepsilon_T - \varepsilon_C) \\
&= \Pr\left(\varepsilon \leq \beta_1 (T_C - T_T)\right)
\end{aligned}
$$

where $\varepsilon = \varepsilon_T - \varepsilon_C$.

Error term

Three assumptions about the random variables ε_T and ε_C

1. What's their mean?
2. What's their variance?
3. What's their distribution?

Note

- For binary choice, it would be sufficient to make assumptions about $\varepsilon = \varepsilon_T - \varepsilon_C$.
- But we want to generalize later on.

The mean

Change of variables

- Define $\mathrm{E}[\varepsilon_C] = \beta_C$ and $\mathrm{E}[\varepsilon_T] = \beta_T$.
- Define $\varepsilon'_C = \varepsilon_C - \beta_C$ and $\varepsilon'_T = \varepsilon_T - \beta_T$,
- so that $\mathrm{E}[\varepsilon'_C] = \mathrm{E}[\varepsilon'_T] = 0$.

Choice model

$P(C|\{C,T\}) =$

$$\begin{array}{lcl} \Pr(\beta_1(T_C - T_T) & \geq & \varepsilon_T - \varepsilon_C) = \\ \Pr(\beta_1(T_C - T_T) & \geq & \varepsilon'_T + \beta_T - \varepsilon'_C - \beta_C) = \\ \Pr(\beta_1(T_C - T_T) + (\beta_C - \beta_T) & \geq & \varepsilon'_T - \varepsilon'_C) = \\ \Pr(\beta_1(T_C - T_T) + \beta_0 & \geq & \varepsilon') \end{array}$$

where $\beta_0 = \beta_C - \beta_T$ and $\varepsilon' = \varepsilon'_T - \varepsilon'_C$.

The mean

Specification

- The means of the error terms can be included as parameters of the deterministic part.
- Only the mean of the difference of the error terms is identified.

Alternative Specific Constant

Equivalent specifications:

$$\begin{array}{rcll} U_C & = & \beta_1 T_C & +\varepsilon_C \\ U_T & = & \beta_1 T_T + \beta_T & +\varepsilon_T \end{array} \text{ or } \begin{array}{rcll} U_C & = & \beta_1 T_C + \beta_C & +\varepsilon_C \\ U_T & = & \beta_1 T_T & +\varepsilon_T \end{array}$$

In practice: associate an alternative specific constant with all alternatives but one.

The mean

Note

Adding the same constant to all utility functions does not affect the choice model

$$\Pr(U_C \geq U_T) = \Pr(U_C + K \geq U_T + K) \quad \forall K \in \mathbb{R}^n.$$

The bottom line...

If the deterministic part of the utility functions contains an Alternative Specific Constant (ASC) for all alternatives but one, the mean of the error terms can be assumed to be zero without loss of generality.

The variance

Utility is ordinal

Utilities can be scaled up or down without changing the choice probability

$$\Pr(U_C \geq U_T) = \Pr(\alpha U_C \geq \alpha U_T) \quad \forall \alpha > 0$$

Link with the variance

$$\begin{aligned} \text{Var}(\alpha U_C) &= \alpha^2 \text{Var}(U_C) \\ \text{Var}(\alpha U_T) &= \alpha^2 \text{Var}(U_T) \end{aligned}$$

Variance is not identified

- As any α can be selected arbitrarily, any variance can be assumed.
- No way to identify the variance of the error terms from data.
- The scale has to be arbitrarily decided.

The distribution

Assumption 1

ε_T and ε_C are the sum of many r.v. capturing unobservable attributes (e.g. mood, experience), measurement and specification errors.

Central-limit theorem

The sum of many i.i.d. random variables approximately follows a normal distribution: $N(\mu, \sigma^2)$.

Assumed distribution

$$\varepsilon_C \sim N(0, 1), \quad \varepsilon_T \sim N(0, 1)$$

The Normal distribution $N(\mu, \sigma^2)$

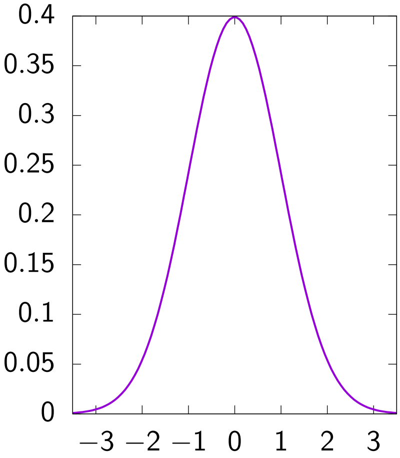
Probability density function (pdf)

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

Cumulative distribution function (CDF)

$$P(c \geq \varepsilon) = F(c) = \int_{-\infty}^{c} f(t)dt$$

No closed form.



The distribution

$\varepsilon = \varepsilon_T - \varepsilon_C$

- From the properties of the normal distribution, we have

$$\begin{aligned} \varepsilon_C &\sim N(0,1) \\ \varepsilon_T &\sim N(0,1) \\ \varepsilon = \varepsilon_T - \varepsilon_C &\sim N(0,2) \end{aligned}$$

- As the variance is arbitrary, we may also assume

$$\begin{aligned} \varepsilon_C &\sim N(0,0.5) \\ \varepsilon_T &\sim N(0,0.5) \\ \varepsilon = \varepsilon_T - \varepsilon_C &\sim N(0,1) \end{aligned}$$

The binary probit model

Choice model

$$P(C|\{C,T\}) = \Pr(\beta_1(T_C - T_T) + \beta_0 \geq \varepsilon) = F_\varepsilon(\beta_1(T_C - T_T) + \beta_0)$$

The binary probit model

$$P(C|\{C,T\}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1(T_C - T_T) - \beta_0} e^{-\frac{1}{2}t^2} dt$$

Not a closed form expression

The distribution

Assumption 2

ε_T and ε_C are the maximum of many r.v. capturing unobservable attributes (e.g. mood, experience), measurement and specification errors.

Gumbel theorem

The maximum of many i.i.d. random variables approximately follows an Extreme Value distribution: $\text{EV}(\eta, \mu)$.

Assumed distribution

$$\varepsilon_C \sim \text{EV}(0,1), \quad \varepsilon_T \sim \text{EV}(0,1).$$

The Extreme Value distribution EV(η, μ)

Probability density function (pdf)

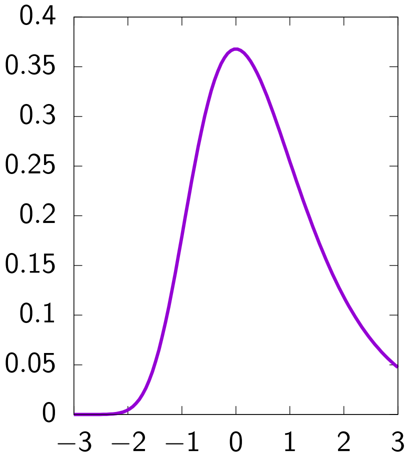
$$f(t) = \mu e^{-\mu(t-\eta)} e^{-e^{-\mu(t-\eta)}}$$

Cumulative distribution function (CDF)

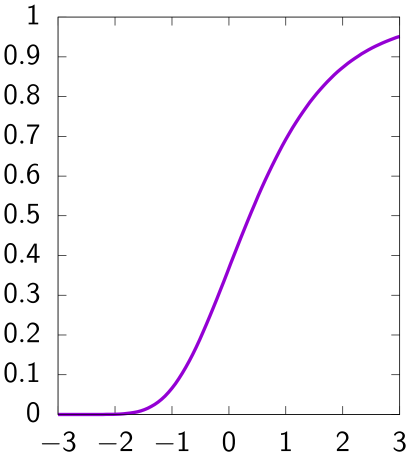
$$\begin{aligned} P(c \geq \varepsilon) = F(c) &= \int_{-\infty}^{c} f(t) dt \\ &= e^{-e^{-\mu(c-\eta)}} \end{aligned}$$

The Extreme Value distribution

pdf EV(0,1)



CDF EV(0,1)



The Extreme Value distribution

Properties

If

$$\varepsilon \sim \mathrm{EV}(\eta, \mu)$$

then

$$\mathrm{E}[\varepsilon] = \eta + \frac{\gamma}{\mu} \quad \text{and} \quad \mathrm{Var}[\varepsilon] = \frac{\pi^2}{6\mu^2}$$

where γ is Euler's constant.

Euler's constant

$$\gamma = \lim_{k \to \infty} \sum_{i=1}^{k} \frac{1}{i} - \ln k = -\int_0^\infty e^{-x} \ln x dx \approx 0.5772$$

The distribution

$\varepsilon = \varepsilon_T - \varepsilon_C$

From the properties of the extreme value distribution, we have

$$
\begin{aligned}
\varepsilon_C &\sim \text{EV}(0,1) \\
\varepsilon_T &\sim \text{EV}(0,1) \\
\varepsilon &\sim \text{Logistic}(0,1)
\end{aligned}
$$

The Logistic distribution: Logistic(η,μ)

Probability density function (pdf)

$$f(t) = \frac{\mu e^{-\mu(t-\eta)}}{(1 + e^{-\mu(t-\eta)})^2}$$

Cumulative distribution function (CDF)

$$P(c \geq \varepsilon) = F(c) = \int_{-\infty}^{c} f(t)dt = \frac{1}{1 + e^{-\mu(c-\eta)}}$$

with $\mu > 0$.

The binary logit model

Choice model

$$P(C|\{C,T\}) = \Pr(\beta_1(T_C - T_T) + \beta_0 \geq \varepsilon) = F_\varepsilon(\beta_1(T_C - T_T) + \beta_0)$$

The binary logit model

$$P(C|\{C,T\}) = \frac{1}{1 + e^{-(\beta_1(T_C - T_T) + \beta_0)}} = \frac{e^{\beta_1 T_C + \beta_0}}{e^{\beta_1 T_C + \beta_0} + e^{\beta_1 T_T}}$$

The binary logit model

$$P(C|\{C,T\}) = \frac{e^{V_C}}{e^{V_C} + e^{V_T}}$$

Outline

1. Motivation
2. Theoretical foundations
   - Decision maker
   - Characteristics
   - Choice set
   - Alternative attributes
   - Decision rule
3. Probabilistic choice theory
4. Binary choice
5. Model specification
   - Error term
6. Applying the model
7. Parameter estimation

Back to the example

Choice

between *Auto* and *Transit*

Data

#	Time auto	Time transit	Choice	#	Time auto	Time transit	Choice
1	52.9	4.4	T	11	99.1	8.4	T
2	4.1	28.5	T	12	18.5	84.0	C
3	4.1	86.9	C	13	82.0	38.0	C
4	56.2	31.6	T	14	8.6	1.6	T
5	51.8	20.2	T	15	22.5	74.1	C
6	0.2	91.2	C	16	51.4	83.8	C
7	27.6	79.7	C	17	81.0	19.2	T
8	89.9	2.2	T	18	51.0	85.0	C
9	41.5	24.5	T	19	62.2	90.1	C
10	95.0	43.5	T	20	95.1	22.2	T
				21	41.6	91.5	C

The model

Utility functions

$$
\begin{aligned}
V_{C1} &= \beta_1 T_{C1} \\
V_{T1} &= \beta_1 T_{T1} + \beta_T
\end{aligned}
$$

Parameters

Let's assume that $\beta_T = 0.5$ and $\beta_1 = -0.1$

First individual

Variables

Let's consider the first observation:

- $T_{C1} = 52.9$
- $T_{T1} = 4.4$
- Choice = *transit*: $y_{\text{auto},1} = 0$, $y_{\text{transit},1} = 1$

Choice

What's the probability given by the model that this individual indeed chooses *transit*?

First individual

Utility functions

$$\begin{aligned} V_{C1} &= \beta_1 T_{C1} &= -5.29 \\ V_{T1} &= \beta_1 T_{T1} + \beta_T &= 0.06 \end{aligned}$$

Choice model

$$P_1(\text{transit}) = \frac{e^{V_{T1}}}{e^{V_{T1}} + e^{V_{C1}}} = \frac{e^{0.06}}{e^{0.06} + e^{-5.29}} \cong 1$$

Comments

- The model fits the observation very well.
- Consistent with the assumption that travel time is the only explanatory variable.

Second individual

Variables

- $T_{C2} = 4.1$
- $T_{T2} = 28.5$
- Choice $=$ *transit*: $y_{\text{auto},2} = 0$, $y_{\text{transit},2} = 1$

Choice

What's the probability given by the model that this individual indeed chooses *transit*?

Second individual

Utility functions

$$\begin{array}{rcccl} V_{C2} & = & \beta_1 T_{C2} & = & -0.41 \\ V_{T2} & = & \beta_1 T_{T2} + \beta_T & = & -2.35 \end{array}$$

Choice model

$$P_2(\text{transit}) = \frac{e^{V_{T2}}}{e^{V_{T2}} + e^{V_{C2}}} = \frac{e^{-2.35}}{e^{-2.35} + e^{-0.41}} \cong 0.13$$

Comment

- The model poorly fits the observation.
- But the assumption is that travel time is the only explanatory variable.
- Still, the probability is not small.

Likelihood

Two observations

The probability that the model reproduces both observations is

$$P_1(\text{transit}) P_2(\text{transit}) = 0.13$$

All observations

The probability that the model reproduces all observations is

$$P_1(\text{transit}) P_2(\text{transit}) \ldots P_{21}(\text{auto}) = 4.62\ 10^{-4}$$

Likelihood of the sample

$$\mathcal{L}' = \prod_n \left(P_n(\text{auto})^{y_{\text{auto},n}} P_n(\text{transit})^{y_{\text{transit},n}} \right)$$

where $y_{j,n}$ is 1 if individual n has chosen alternative j, 0 otherwise

Likelihood

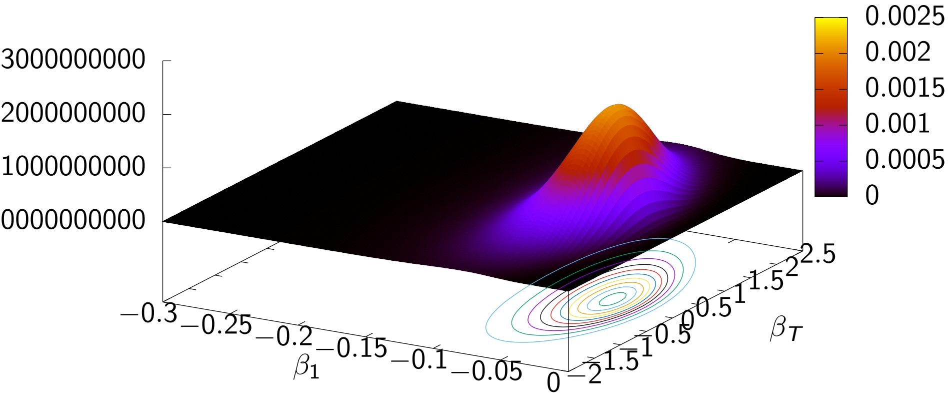
Likelihood

- Probability that the model fits all observations.
- It is a function of the parameters.

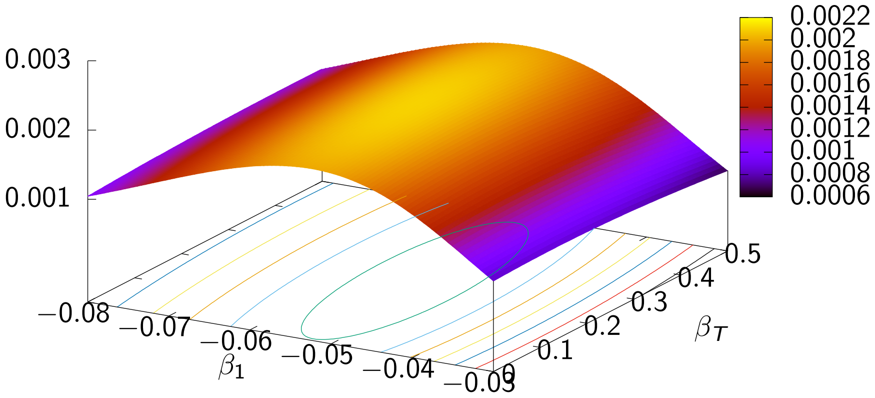
Examples

β_T	β_1	$\mathcal{L}'$
0	0	$4.57\ 10^{-07}$
0	-1	$1.97\ 10^{-30}$
0	-0.1	$4.1\ 10^{-04}$
0.5	-0.1	$4.62\ 10^{-04}$

Likelihood function



Likelihood function (zoom)



Maximum likelihood estimation

Estimators for the parameters

Parameters that achieve the maximum likelihood

$$\max_{\beta} \prod_{n} \left(P_n(\text{auto}; \beta)^{y_{\text{auto},n}} P_n(\text{transit}; \beta)^{y_{\text{transit},n}} \right)$$

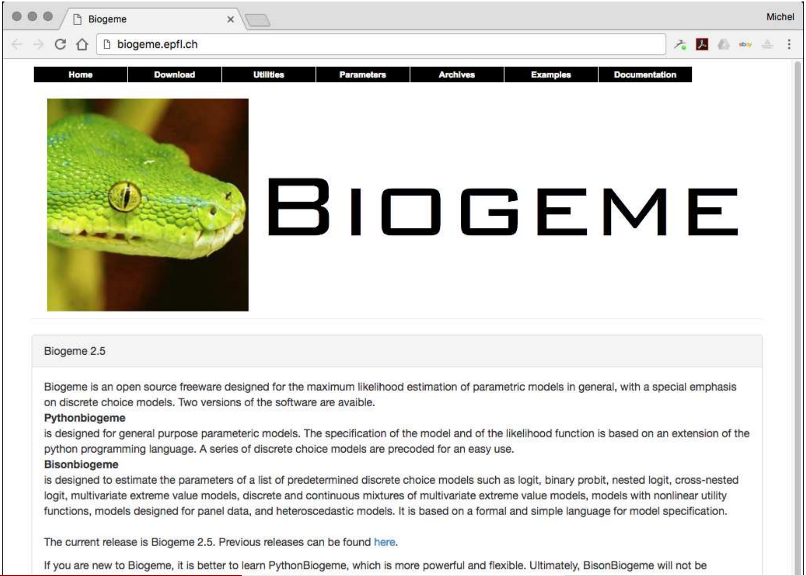
Log likelihood

Alternatively, we prefer to maximize the log likelihood

$$\max_{\beta} \ln \prod_{n} \left(P_n(\text{auto})^{y_{\text{auto},n}} P_n(\text{transit})^{y_{\text{transit},n}} \right) =$$

$$\max_{\beta} \sum_{n} y_{\text{auto},n} \ln P_n(\text{auto}) + y_{\text{transit},n} \ln P_n(\text{transit})$$

Biogeme



Home | Download | Utilities | Parameters | Archives | Examples | Documentation

BIOGEME

Biogeme 2.5

Biogeme is an open source freeware designed for the maximum likelihood estimation of parametric models in general, with a special emphasis on discrete choice models. Two versions of the software are avaible.

Pythonbiogeme

is designed for general purpose parameteric models. The specification of the model and of the likelihood function is based on an extension of the python programming language. A series of discrete choice models are precoded for an easy use.

Bisonbiogeme

is designed to estimate the parameters of a list of predetermined discrete choice models such as logit, binary probit, nested logit, cross-nested logit, multivariate extreme value models, discrete and continuous mixtures of multivariate extreme value models, models with nonlinear utility functions, models designed for panel data, and heteroscedastic models. It is based on a formal and simple language for model specification.

The current release is Biogeme 2.5. Previous releases can be found here.

If you are new to Biogeme, it is better to learn PythonBiogeme, which is more powerful and flexible. Ultimately, BisonBiogeme will not be

Short course

Discrete Choice Analysis

- Every year at EPFL, Switzerland
- Next event: February 12 — February 16, 2017
- Lecturers: Moshe Ben-Akiva (MIT), Michel Bierlaire (EPFL)
- URL: `transp-or.epfl.ch/dca`