
Computer Lab II

Further introduction to Biogeme Binary Logit Model Estimation

Antonin Danalet, Amanda Stathopoulos, Bilge Atasoy

Today

- **Few words about semester projects for mathematicians**
- Further introduction to BIOGEME
- Estimation of Binary Logit models

Pedestrian Activities from WiFi traces

- Model for pedestrian activities from WiFi traces
- Code in Python, using triangulation data (Cisco)
- To Do:
 - Needs to be adapted for Access Point data (Radius). Comparison with other dataset (Cisco)
 - Calibration of the model: number of people in the restaurants
 - Integrate a shortest path algorithm in Python
 - Visualization of results, e.g., how many people in Sat'
- Skills: some coding in Python, possibly visualization
- If interested talk to Antonin

Modelling the choice of vehicle

- Wider model of vehicle choice (work with car-manufacturer)
- Factor in richer behavioral variables; attitudes, dynamic factors
- To Do:
 - data elaboration
 - exploratory analysis
 - confirmatory models
 - formulate behavioral model
- Skills: large data, statistics and bonus = behavioral theories of decision-making
- If interested talk to Amanda

Today

- Few words about semester projects for mathematicians
- **Further introduction to BIOGEME**
- Estimation of Binary Logit models

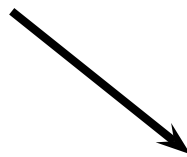
How does BIOGEME work?

How does BIOGEME work?

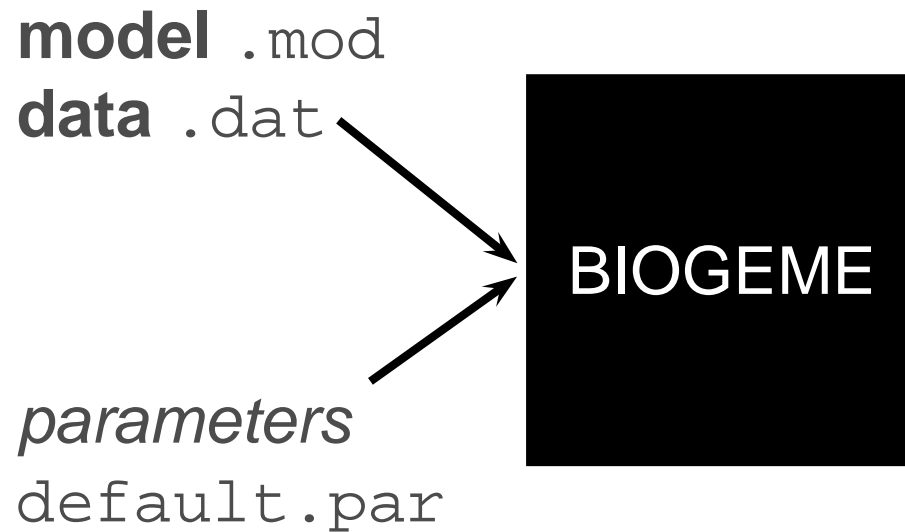
BIOGEME

How does BIOGEME work?

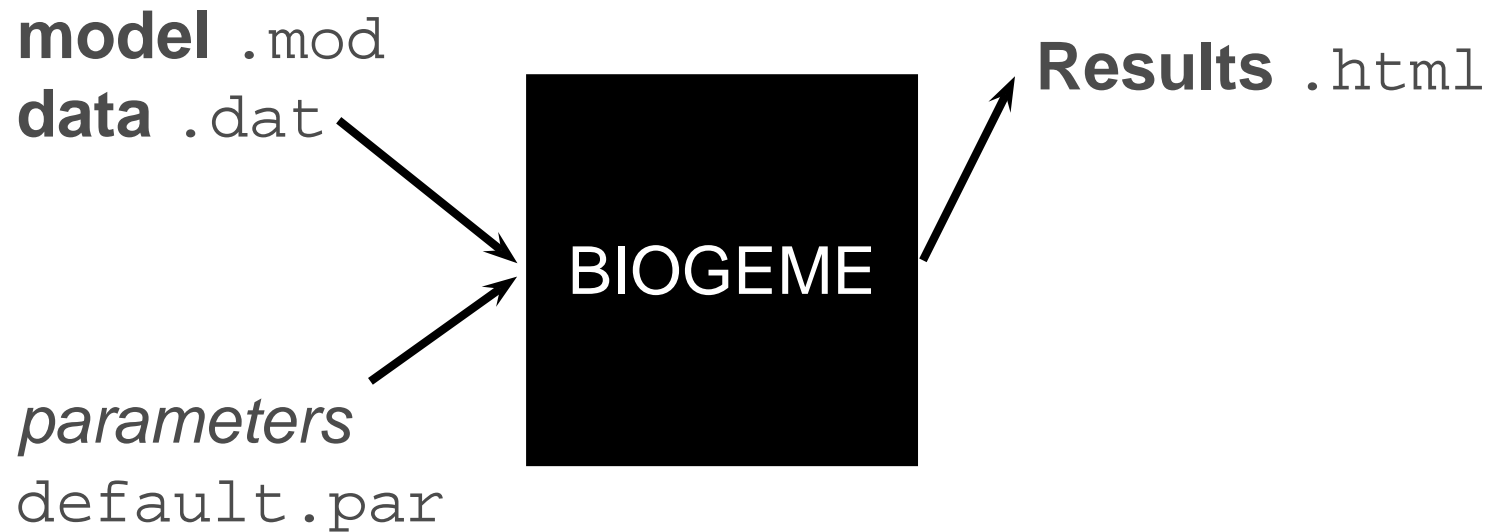
model .mod
data .dat



How does BIOGEME work?



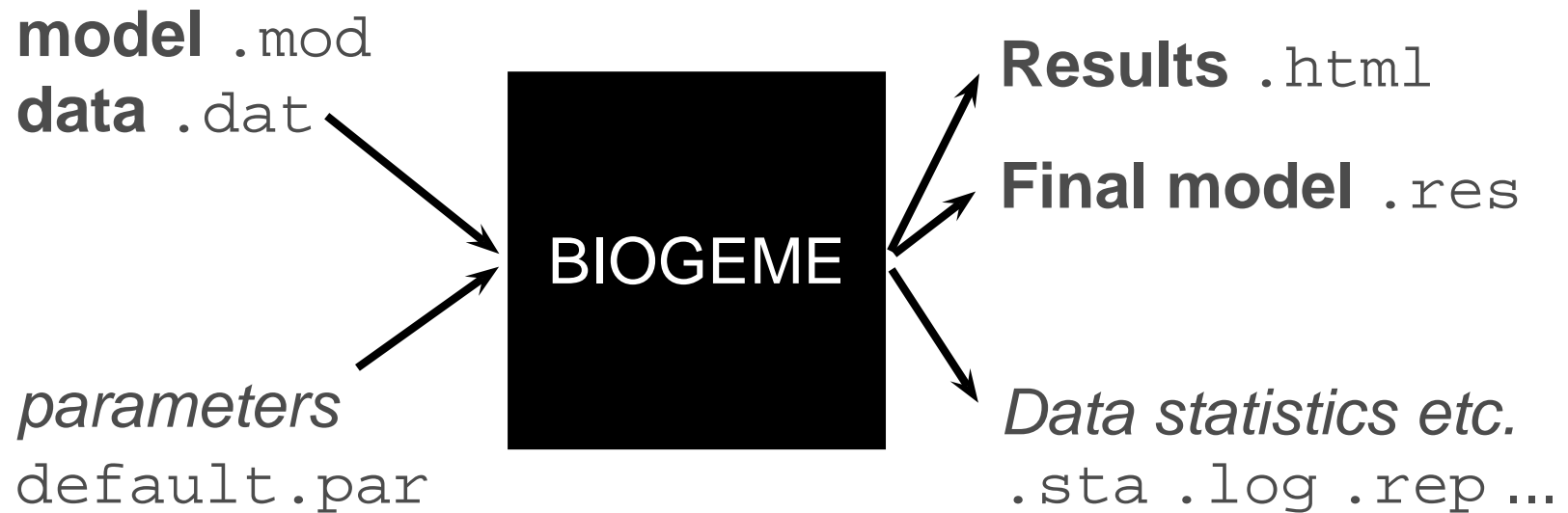
How does BIOGEME work?



How does BIOGEME work?



How does BIOGEME work?



BIOGEME - Data file

- File extension `.dat`
- First row contains column / variable names
- One observation per row
- Each row must contain a choice indicator
- Example with the Netherlands transportation mode choice data:
choice between car and train

BIOGEME - Data file

netherlands.dat

id	choice	rail_cost	rail_time	car_cost	car_time
1	0	40	2.5	5	1.167
2	0	35	2.016	9	1.517
3	0	24	2.017	11.5	1.966
4	0	7.8	1.75	8.333	2
5	0	28	2.034	5	1.267
...					
219	1	35	2.416	6.4	1.283
220	1	30	2.334	2.083	1.667
221	1	35.7	1.834	16.667	2.017
222	1	47	1.833	72	1.533
223	1	30	1.967	30	1.267

BIOGEME - Data file

netherlands.dat

id	choice	rail_cost	rail_time	car_cost	car_time
1	0	40	2.5	5	1.167
2	0	35	2.016	9	1.517
3	0	24	2.017	11.5	1.966
4	0	7.8	1.75	8.333	2
5	0	28	2.034	5	1.267
...					
219	1	35	2.416	6.4	1.283
220	1	30	2.334	2.083	1.667
221	1	35.7	1.834	16.667	2.017
222	1	47	1.833	72	1.533
223	1	30	1.967	30	1.267

Unique identifier of observations

BIOGEME - Data file

netherlands.dat

id	choice	rail_cost	rail_time	car_cost	car_time
1	0	40	2.5	5	1.167
2	0	35	2.016	9	1.517
3	0	24	2.017	11.5	1.966
4	0	7.8	1.75	8.333	2
5	0	28	2.034	5	1.267
...					
219	1	35	2.416	6.4	1.283
220	1	30	2.334	2.083	1.667
221	1	35.7	1.834	16.667	2.017
222	1	47	1.833	72	1.533
223	1	30	1.967	30	1.267

Choice indicator, 0: car and 1: train

BIOGEME - Model file

- File extension `.mod`
- Must be consistent with data file
- Contains deterministic utility specifications, model type etc.
- The model file contains different *sections* describing different elements of the model specification

BIOGEME - Model file

- How can we write the following deterministic utility functions for BIOGEME?

$$V_{\text{car}} = \text{ASC}_{\text{car}} + \beta_{\text{time}} \text{car}_{\text{time}} + \beta_{\text{cost}} \text{car}_{\text{cost}}$$

$$V_{\text{rail}} = \beta_{\text{time}} \text{rail}_{\text{time}} + \beta_{\text{cost}} \text{rail}_{\text{cost}}$$

BIOGEME - Model file

```
[Choice]
```

```
choice
```

```
[Beta]
```

```
// Name          DefaultValue LowerBound UpperBound  status
ASC_CAR          0.0          -100.0    100.0       0
ASC_RAIL         0.0          -100.0    100.0       1
BETA_COST        0.0          -100.0    100.0       0
BETA_TIME        0.0          -100.0    100.0       0
```

```
[Utilities]
```

```
//Id Name Avail linear-in-parameter expression
0   Car  one   ASC_CAR * one + BETA_COST * car_cost +
      BETA_TIME * car_time
1   Rail one   ASC_RAIL * one + BETA_COST * rail_cost +
      BETA_TIME * rail_time
```

BIOGEME - Model file

```
[Choice]
choice
```

```
[Beta]
// Name      DefaultValue LowerBound UpperBound  status
ASC_CAR      0.0          -100.0    100.0       0
ASC_RAIL     0.0          -100.0    100.0       1
BETA_COST    0.0          -100.0    100.0       0
BETA_TIME    0.0          -100.0    100.0       0
```

```
[Utilities]
//Id Name Avail linear-in-parameter expression
0 Car one ASC_CAR * one + BETA_COST * car_cost +
BETA_TIME * car_time
1 Rail one ASC_RAIL * one + BETA_COST * rail_cost +
BETA_TIME * rail_time
```

BIOGEME - Model file

```
[Choice]
```

```
choice
```

```
[Beta]
```

// Name	DefaultValue	LowerBound	UpperBound	status
ASC_CAR	0.0	-100.0	100.0	0
ASC_RAIL	0.0	-100.0	100.0	1
BETA_COST	0.0	-100.0	100.0	0
BETA_TIME	0.0	-100.0	100.0	0

```
[Utilities]
```

//Id	Name	Avail	linear-in-parameter expression
0	Car	one	ASC_CAR * one + BETA_COST * car_cost + BETA_TIME * car_time
1	Rail	one	ASC_RAIL * one + BETA_COST * rail_cost + BETA_TIME * rail_time

BIOGEME - Model file

[Choice]

What is one?

choice

Which is the type of model?

[Beta]

// Name	DefaultValue	LowerBound	UpperBound	status
ASC_CAR	0.0	-100.0	100.0	0
ASC_RAIL	0.0	-100.0	100.0	1
BETA_COST	0.0	-100.0	100.0	0
BETA_TIME	0.0	-100.0	100.0	0

[Utilities]

//Id	Name	Avail	linear-in-parameter	expression
0	Car	one	ASC_CAR * one + BETA_COST * car_cost + BETA_TIME * car_time	
1	Rail	one	ASC_RAIL * one + BETA_COST * rail_cost + BETA_TIME * rail_time	

BIOGEME - Model file

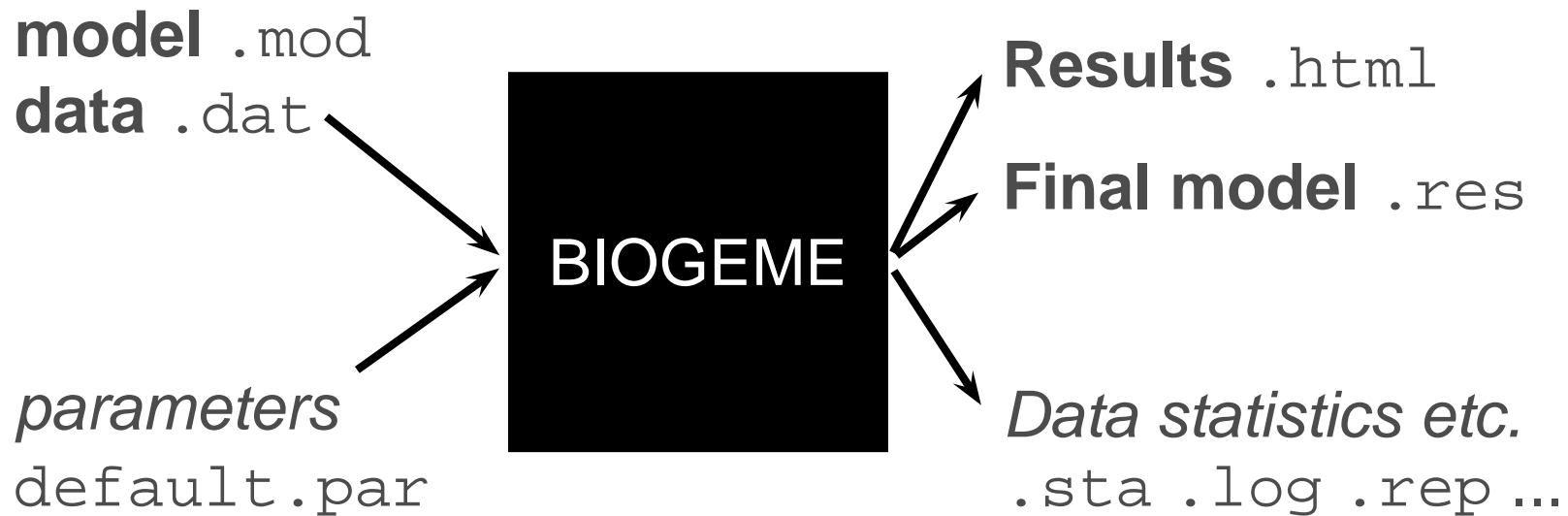
```
[Expressions]
```

```
// Define here arithmetic expressions for name that are not directly  
// available from the data  
one = 1
```

```
[Model]
```

```
// Currently, only $MNL (multinomial logit), $NL (nested logit), $CNL  
// (cross-nested logit) and $NGEV (Network GEV model) are valid keywords  
//  
$MNL
```

How does BIOGEME work?



Model and Data Files

- How to read and modify model files?
How to read data files?
 - GNU Emacs, vi, TextEdit (Mac) or Wordpad (Windows)
 - **Notepad (Windows) should not be used!**

BIOGEME - Results - Netherlands dataset

```
binary_generic_netherlands.html - Report from BIOGEME Version 1.8 [Sat Mar 7 14:36:56 CEST 2009] - Moz...
Echier  Édition  Affichage  Historique  Marque-pages  Outils  ?
file:///C:/BinaryNetherlands/binary_generic_netherlands.html
Les plus visités  Débuter avec Firefox  À la une
binary_generic_netherlands...
BIOGEME Version 1.8 [Sat Mar 7 14:36:56 CEST 2009]
Michel Bierlaire, EPFL
This file has automatically been generated.
09/23/10 12:08:00
Model: Multinomial Logit
Number of estimated parameters: 3
Number of observations: 228
Number of individuals: 228
Null log-likelihood: -158.038
Cte log-likelihood: -148.347
Init log-likelihood: -158.038
Final log-likelihood: -123.133
Likelihood ratio test: 69.809
Rho-square: 0.221
Adjusted rho-square: 0.202
Final gradient norm: +6.045e-004
Diagnostic: Convergence reached...
Iterations: 7
Run time: 00:00
Variance-covariance: from analytical hessian
Sample file: netherlands05.dat
Utility parameters


| Name      | Value   | Std err | t-test | p-value | Robust Std err | Robust t-test | p-value |
|-----------|---------|---------|--------|---------|----------------|---------------|---------|
| ASC_CAR   | -0.798  | 0.270   | -2.95  | 0.00    | 0.275          | -2.90         | 0.00    |
| ASC_RAIL  | 0.00    | fixed   |        |         |                |               |         |
| BETA_COST | -0.0499 | 0.0103  | -4.85  | 0.00    | 0.0107         | -4.67         | 0.00    |
| BETA_TIME | -1.33   | 0.344   | -3.86  | 0.00    | 0.354          | -3.75         | 0.00    |


Terminé
```

BIOGEME - Results

```
Model: Multinomial Logit
Number of estimated parameters: 3
Number of observations: 228
Number of individuals: 228
Null log-likelihood: -158.038
Cte log-likelihood: -148.347
Init log-likelihood: -158.038
Final log-likelihood: -123.133
Likelihood ratio test: 69.809
Rho-square: 0.221
Adjusted rho-square: 0.202
Final gradient norm: +6.045e-004
Diagnostic: Convergence reached...
Iterations: 7
Run time: 00:00
Variance-covariance: from analytical hessian
Sample file: netherlands05.dat
```

Name	Value	Std err	t-test	p-value	Robust Std err	Robust t-test	p-value
ASC_CAR	-0.798	0.270	-2.95	0.00	0.275	-2.90	0.00
ASC_RAIL	0.00	fixed					
BETA_COST	-0.0499	0.0103	-4.85	0.00	0.0107	-4.67	0.00
BETA_TIME	-1.33	0.344	-3.86	0.00	0.354	-3.75	0.00

Terminé

General model information

BIOGEME - Results

```
binary_generic_netherlands.html - Report from BIOGEME Version 1.8 [Sat Mar 7 14:36:56 CEST 2009] - Moz...
Bichier  Édition  Affichage  Historique  Marque-pages  Outils  ?
file:///C:/BinaryNetherlands/binary_generic_netherlands.html
Les plus visités  Débuter avec Firefox  À la une
binary_generic_netherlands...
BIOGEME Version 1.8 [Sat Mar 7 14:36:56 CEST 2009]
Michel Bierlaire, EPFL
This file has automatically been generated.
09/23/10 12:08:00
Model: Multinomial Logit
Number of estimated parameters: 3
Number of observations: 228
Number of individuals: 228
Null log-likelihood: -158.038
Cte log-likelihood: -148.347
Init log-likelihood: -158.038
Final log-likelihood: -123.133
Likelihood ratio test: 69.809
Rho-square: 0.221
Adjusted rho-square: 0.202
Final gradient norm: +6.045e-004
Diagnostic: Convergence reached...
Iterations: 7
Run time: 00:00
Variance-covariance: from analytical hessian
Sample file: netherlands05.dat
Utility parameters
```

Name	Value	Std err	t-test	p-value	Robust Std err	Robust t-test	p-value
ASC_CAR	-0.798	0.270	-2.95	0.00	0.275	-2.90	0.00
ASC_RAIL	0.00	fixed					
BETA_COST	-0.0499	0.0103	-4.85	0.00	0.0107	-4.67	0.00
BETA_TIME	-1.33	0.344	-3.86	0.00	0.354	-3.75	0.00

Terminé

Coefficient estimates

Today

- Further introduction to BIOGEME
- **Estimation of Binary Logit models**

Binary Logit Case Study

- Available datasets:
 - Airline itinerary choice (Boeing)
 - Choice-Lab marketing
 - Mode choice in Netherlands
 - Residential Telephone Services
 - Mode choice in Switzerland (Optima)
- Descriptions available on the course web site
- Optima dataset does not contain `.mod` files. A specification has to be proposed as an assignment.

How to go through the Case Studies

- Choose a dataset to work with (data descriptions are available on the course website)
- Copy the files related to the chosen dataset and case study from the course website.
- Study the `.mod` files with the help of the descriptions
- Run the `.mod` files with BIOGEME
- Interpret the results and compare your interpretation with the one we have proposed
- Develop other model specifications

Course website

- <http://transp-or.epfl.ch/> → Teaching → Mathematical modeling of behavior → Laboratories
- BIOGEME software (including documentation and utilities)
- For each Case Study
 - Data files for available datasets
 - Model specification files
 - Possible interpretation of results

Running Biogeme

In the MXF014

- accessing your folder My Documents:
type `L:` in the DOS command window

To run Biogeme on your own computer

- download GUI from the course web site:
 - BIOGEME v2.2: `Windows executables.zip`
 - put `biogeme.exe` in `C:\Program Files`
 - **OR:** put `winbiogeme.exe` in the folder where you would like to work
 - run it by double clicking on the icon

Today's plan

1. Independent work on 2-3 Case Studies

- choose case
- estimate
- interpret

2. Group work

- listen to description of data
- gather in groups
- generate .mod file (base)
- test an idea / hypothesis

OPTIMA

- Motivation: study mode choice in low density areas of switzerland
- Data-collection: Revealed Preferences (RP)
- Survey-type: Mail survey
- Collects 'loops' = cyclical sequence of trips
- database with 1906 observations
 - trip features
 - socioeconomic variables

OPTIMA - Data file

Table 1: extract 'optimaTOT3_valid.dat'

ID	TripPurpose	Choice	CoutTP1	CoutAutoCHF	age	CodeLangue
10350017	1	1	12.4	4.54	27	1
10350025	3	0	3	0.64	-1	1
10350075	1	1	24	3.38	63	1
10350085	1	1	10.8	1.66	57	1
10350086	1	1	9.6	2.71	58	1
10350100	3	1	3	1.63	80	1
10350120	3	1	24.8	6.98	64	1
10350125	3	1	4.8	0.99	68	1
10350125	3	1	2.2	0.19	68	1

Lab assignment 1

- gather group and work on your own specification of a Binary Logit on the Optima data
- examine data & variable description
- write a .mod file
- make up your own hypothesis
- test hypothesis
- results sent by mail on Friday 5th October
 - the .html file
 - .mod file
 - 2-page description