



# Estimating discrete choice models with Biogeme

Onur Akman

Complex Social Systems (2025)

19/03/2025

[onur.akman@uj.edu.pl](mailto:onur.akman@uj.edu.pl)

# Introduction

- We will explore **discrete choice modeling**: Modeling how people make decisions among discrete alternatives.
  - How do people choose where to eat?
  - How do people choose what to watch?
  - How do people choose where to sit in this lab?
- We will use **Biogeme**, a Python package for estimating choice models.

# Why Model Discrete Choices?

- Many real-world problems involve choices:
  - Product selection (brand A vs. B)
  - Online decisions (click vs. no click)
- We are interested in
  - **understanding how people make choices,**
  - **what are the factors in their thought processes** and
  - **replicating them in the future with maximum accuracy.**
- Understanding choices helps with policy-making, marketing, and system design.
- Unlike simple statistics, **DCM helps analyze the causal effects of factors on choices.**

Natural question:  
**Why Not Just  
Use Machine  
Learning?**

ML excels at prediction but **often  
lacks interpretability.**



With DCM, we can:

Interpret coefficients  
**meaningfully.**

Make **counterfactual  
analyses** (e.g., “What  
if we reduce bus fares  
by 10%?”).

Ensure choices follow  
rational behavior (e.g.,  
consistency with  
economic theory).

# DCM: Toy example

- Step 1: Define the problem
  - How do people choose which one to buy among the **three shampoo products**?
- Step 2: Obtain data
  - We need a dataset containing:
    - decisions of consumers
    - useful information about the consumers and the products (eg., age of the consumer, income of the consumer, price of the shampoo)

# DCM: Toy example (2)

- Step 3: Define utilities
  - We say that each consumer **maximizes** a “**utility function**” while making a decision.
  - We design **utility functions associated with each option**. For example, say that the following is the utility function of Shampoo A:
    - $U_A = \beta_1 \cdot Price + \beta_2 \cdot Availability + \beta_3 \cdot Organic$
    - Where  $\beta_1, \beta_2, \beta_3$  are parameters we estimate from data.
- Step 4: Estimation
  - Based on our observations in our dataset, estimate parameters in our model (betas) which maximizes the likelihood of our observations.
- Step 5: Interpret
  - If  $\beta_1 < 0$ , people prefer cheaper shampoos.
  - If  $\beta_3 > 0$ , people prefer organic shampoos.

# Biogeme

- Biogeme provides a unified framework for:
  - Defining parameters and utility functions
  - Abstraction in estimation
  - A catalog of discrete choice models
- It is developed and maintained by Prof. Michel Bierlaire, Ecole Polytechnique Fédérale de Lausanne, Switzerland.
- Website: <https://biogeme.epfl.ch/>



**BIOGEME**

# Tutorial

- Colab link:

[https://colab.research.google.com/drive/1a5wheL60vVIXnq\\_PxmmTL5Ky8ShyH3Ks?usp=sharing](https://colab.research.google.com/drive/1a5wheL60vVIXnq_PxmmTL5Ky8ShyH3Ks?usp=sharing)

- Colab link (shortened):

<https://tinyurl.com/486akz5a>

- Download the notebook from:

ComplexSocialSystemsCourse > labs > lab2