Mathematical models of choice behavior

Michel Bierlaire
michel.bierlaire@epfl.ch

Ecole Polytechnique Fédérale de Lausanne
Introduction

- What kind of behavior can be mathematically modeled?
Psychohistory

Branch of mathematics which deals with the reactions of human conglomerates to fixed social and economic stimuli. The necessary size of such a conglomerate may be determined by Seldon’s First Theorem.

Encyclopedia Galactica, 116th Edition (1020 F.E.)
Encyclopedia Galactica Publishing Co., Terminus

Motivation: shorten the period of barbarism after the Fall of the Galactic Empire
Asimov, I. (1951) Foundation, Gnome Press
Here...

- Individual behavior (vs. aggregate behavior)
- Theory of behavior which is
  - descriptive: how people behave and not how they should
  - abstract: not too specific
  - operational: can be used in practice for forecasting
- Type of behavior: choice
Motivations

WELL, EACH DECISION WE MAKE DETERMINES THE RANGE OF CHOICES WE’LL FACE NEXT.
Motivations

“It is our choices that show what we truly are, far more than our abilities” Albus Dumbledore
“Liberty, taking the word in its concrete sense, consists in the ability to choose.” Simone Weil (French philosopher, 1909-1943)

Field:
- Marketing
- Transportation
- Politics
- Management
- New technologies

Type of behavior:
- Choice of a brand
- Choice of a transportation mode
- Choice of a president
- Choice of a management policy
- Choice of investments
Applications

Case studies

- Choice-lab marketing
  - Context: B2B, data provider (financial, demographic, etc.)
  - Objective: understand why clients quit

- Quebec energy
  - Context: space and water heating in households
  - Objective: importance of the type of household and price

- Transportation mode choice in the Netherlands
  - Context: car vs rail in Nijmegen
  - Objective: sensitivity to travel time and cost, inertia.
Applications

• Swissmetro
  • Context: new transportation technology
  • Objective: demand pattern, pricing

• Residential telephone services
  • Context: flat rate vs. measured
  • Objective: offer the most appropriate service
Importance

Daniel L. McFadden

1937–

• 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”
Framework

Choice: outcome of a sequential decision-making process

- Definition of the choice problem: How do I get to EPFL?
- Generation of alternatives: car as driver, car as passenger, train
- Evaluation of the attributes of the alternatives: price, time, flexibility, comfort
- Choice: decision rule
- Implementation: travel
Framework

A choice theory defines
1. decision maker
2. alternatives
3. attributes of alternatives
4. decision rule
Decision-maker

- Individual or a group of persons
- If group of persons, we ignore internal interactions
- Important to capture difference in tastes and decision-making process
- Socio-economic characteristics: age, gender, income, education, etc.
Alternatives

- Environment: *universal choice set* (\( \mathcal{U} \))
- Individual \( n \): *choice set* (\( \mathcal{C}_n \))

**Choice set generation:**

- Availability
- Awareness

Ph.D. dissertation, Department of Civil Engineering, MIT, Cambridge, Ma.
Alternatives

Continuous vs. discrete

Continuous choice set:

\[ p_{\text{Beer}} q_{\text{Beer}} + p_{\text{Milk}} q_{\text{Milk}} = I \]

Discrete choice set:

\[ C_n = \{ \text{Car, Bus, Bike} \} \]
Attributes

- cost
- travel time
- walking time
- comfort
- bus frequency
- etc.

- ✔ Generic vs. specific
- ✔ Quantitative vs. qualitative
- ✔ Perception
Decision rules

Neoclassical economic theory
Preference-indifference operator $\succeq$

(i) reflexivity
\[ a \succeq a \quad \forall a \in C_n \]

(ii) transitivity
\[ a \succeq b \text{ and } b \succeq c \Rightarrow a \succeq c \quad \forall a, b, c \in C_n \]

(iii) comparability
\[ a \succeq b \text{ or } b \succeq a \quad \forall a, b \in C_n \]
Decision rules

Neoclassical economic theory (ctd)

☞ Numerical function

\[ \exists \ U_n : C_n \rightarrow \mathbb{R} : a \sim U_n(a) \text{ such that} \]

\[ a \succeq b \iff U_n(a) \geq U_n(b) \quad \forall a, b \in C_n \]
Decision rules

- Utility is a latent concept
- It cannot be directly observed
Example: continuous choice

Continuous choice set

- \( Q = \{ q_1, \ldots, q_L \} \) consumption bundle
- \( q_i \) is the quantity of product \( i \) consumed
- Utility of the bundle:
  \[ U(q_1, \ldots, q_L) \]
- \( Q_a \succsim Q_b \) iff \( U(q_1^a, \ldots, q_L^a) \geq U(q_1^b, \ldots, q_L^b) \)
- Budget constraint:
  \[ \sum_{i=1}^{L} p_i q_i \leq I. \]
Example: continuous choice

Decision-maker solves the optimization problem

$$\max_{q \in \mathbb{R}^L} U(q_1, \ldots, q_L)$$

subject to

$$\sum_{i=1}^{L} p_i q_i = I.$$ 

Example with two products...
Example: continuous choice

$$\max_{q_1, q_2} U = \beta_0 q_1^{\beta_1} q_2^{\beta_2}$$

subject to

$$p_1 q_1 + p_2 q_2 = I.$$ 

Lagrangian of the problem:

$$L(q_1, q_2, \lambda) = \beta_0 q_1^{\beta_1} q_2^{\beta_2} + \lambda(I - p_1 q_1 - p_2 q_2).$$

Necessary optimality condition

$$\nabla L(q_1, q_2, \lambda) = 0$$
Example: continuous choice

Necessary optimality conditions

\[ \beta_0 \beta_1 q_1^{\beta_1 - 1} q_2^{\beta_2} - \lambda p_1 = 0 \]
\[ \beta_0 \beta_2 q_1^{\beta_1} q_2^{\beta_2 - 1} - \lambda p_2 = 0 \]
\[ p_1 q_1 + p_2 q_2 - I = 0. \]

We have

\[ \beta_0 \beta_1 q_1^{\beta_1} q_2^{\beta_2} - \lambda p_1 q_1 = 0 \]
\[ \beta_0 \beta_2 q_1^{\beta_1} q_2^{\beta_2} - \lambda p_2 q_2 = 0 \]

so that

\[ \lambda I = \beta_0 q_1^{\beta_1} q_2^{\beta_2} (\beta_1 + \beta_2) \]
Example: continuous choice

Therefore

$$\beta_0 q_1^{\beta_1} q_2^{\beta_2} = \frac{\lambda I}{(\beta_1 + \beta_2)}$$

As $\beta_0 \beta_2 q_1^{\beta_1} q_2^{\beta_2} = \lambda p_2 q_2$, we obtain (assuming $\lambda \neq 0$)

$$q_2 = \frac{I \beta_2}{p_2(\beta_1 + \beta_2)}$$

Similarly, we obtain

$$q_1 = \frac{I \beta_1}{p_1(\beta_1 + \beta_2)}$$
Example: continuous choice

\[ q_1 = \frac{I\beta_1}{p_1(\beta_1 + \beta_2)} \]

\[ q_2 = \frac{I\beta_2}{p_2(\beta_1 + \beta_2)} \]

Demand functions
Discrete choice

- Similarities with Knapsack problem
- Calculus cannot be used anymore

\[ U = U(q_1, \ldots, q_L) \]

with

\[ q_i = \begin{cases} 
1 & \text{if product } i \text{ is chosen} \\
0 & \text{otherwise} 
\end{cases} \]

and

\[ \sum_i q_i = 1. \]
Discrete choice

- Do not work with demand functions anymore
- Work with utility functions
- $U$ is the “global” utility
- Define $U_i$ the utility associated with product $i$.
- It is a function of the attributes of the product (price, quality, etc.)
- We say that product $i$ is chosen if

\[ U_i \geq U_j \quad \forall j. \]
Example

Example: two transportation modes

\[ U_1 = -\beta t_1 - \gamma c_1 \]
\[ U_2 = -\beta t_2 - \gamma c_2 \]

with \( \beta, \gamma > 0 \)

\[ U_1 \geq U_2 \text{ iff } -\beta t_1 - \gamma c_1 \geq -\beta t_2 - \gamma c_2 \]

that is

\[ \frac{\beta}{\gamma} t_1 - c_1 \geq \frac{\beta}{\gamma} t_2 - c_2 \]

or

\[ c_1 - c_2 \leq -\frac{\beta}{\gamma} (t_1 - t_2) \]
Example

Obvious cases:

- \( c_1 \geq c_2 \) and \( t_1 \geq t_2 \): 2 dominates 1.
- \( c_2 \geq c_1 \) and \( t_2 \geq t_1 \): 1 dominates 2.
- Trade-offs in over quadrants
Example

![Graph showing a line with points labeled 1 is chosen and 2 is chosen. The line passes through the points (1, 1) and (0, 2).]
Example

1 is chosen

2 is chosen

Introduction – p. 30/74
Assumptions

Decision rules
Neoclassical economic theory (ctd)
Decision-maker

✔️ perfect discriminating capability
✔️ full rationality
✔️ permanent consistency

Analyst

✔️ knowledge of all attributes
✔️ perfect knowledge of \( \succsim \) (or \( U_n(\cdot) \))
✔️ no measurement error
Assumptions

Uncertainty

Source of uncertainty?

☞ Decision-maker: stochastic decision rules
☞ Analyst: lack of information

☞ Bohr: “Nature is stochastic”
☞ Einstein: “God does not play dice”
Assumptions

Lack of information: random utility models


Sources of uncertainty:

☞ Unobserved attributes
☞ Unobserved taste variations
☞ Measurement errors
☞ Instrumental variables

For each individual \( n \),

\[
U_{in} = V_{in} + \epsilon_{in}
\]

Dependent variable is latent. Therefore, we prefer the model

\[
P(i|C_n) = P[U_{in} = \max_{j \in C_n} U_{jn}] = P(U_{in} \geq U_{jn} \ \forall j \in C_n)
\]
## Example

### Data:

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

The distribution

Assumption: \( \varepsilon_T \) and \( \varepsilon_C \) 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.

\[
\varepsilon_C \sim \text{EV}(0, \mu)
\]
Error term

$\text{EV}(\eta, \mu)$, with $\mu > 0$:

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

If $\varepsilon \sim \text{EV}(\eta, \mu)$, then

The distribution

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

$$= e^{-e^{-\mu(c-\eta)}}$$
Error term

Gumbel PDF $\mu = 0, \sigma = 1$
Error term
Error term

If

\[ \varepsilon \sim EV(\eta, \mu) \]

then

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

where \( \gamma \) is Euler’s constant

\[ \gamma = \lim_{k \to \infty} \sum_{i=1}^{k} \frac{1}{i} - \ln k \]

\[ = - \int_{\infty}^{\infty} e^{-x} \ln x \, dx \]

\[ \approx 0.5772 \]
Error term

The distribution

\[ P(C|\{C, T\}) = P(\varepsilon \leq V_C - V_T) = P(\varepsilon \leq \beta_1(T_C - T_T) - \beta_0) \]

where \( \varepsilon = \varepsilon_T - \varepsilon_C \).

\[ \varepsilon_C \sim \text{EV}(0, \mu) \]
\[ \varepsilon_T \sim \text{EV}(0, \mu) \]
\[ \varepsilon \sim \text{Logistic}(0, \mu) \]

Logit Model
Error term

The distribution
For the Logistic(0, µ), we have

\[ P(c \geq \varepsilon) = F(c) = \frac{1}{1 + e^{-\mu c}} \]

\[ P(C|\{C, T\}) = P(\varepsilon \leq V_C - V_T) \]
\[ = F(V_C - V_T) \]
\[ = \frac{1}{1 + e^{-\mu(V_C - V_T)}} \]
\[ = \frac{e^{\mu V_C}}{e^{\mu V_C} + e^{\mu V_T}} \]
Binary Logistic Unit Model or Binary Logit Model

Normalize $\mu = 1$
Back to the example

Let’s assume that $\beta_0 = 0.5$ and $\beta_1 = -0.1$

Let’s consider the first observation:

- $T_C = 52.9$
- $T_T = 4.4$
- **Choice = transit**

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

$$V_C = \beta_1 T_C = -5.29$$
$$V_T = \beta_1 T_T + \beta_0 = 0.06$$
Back to the example

\[ P(\text{transit}) = \frac{e^{V_T}}{e^{V_T} + e^{V_C}} \]

\[ P(\text{transit}) = \frac{e^{0.06}}{e^{0.06} + e^{-5.29}} \approx 1 \]

The model almost perfectly predicts this observation
Back to the example

Let’s assume again that $\beta_0 = 0.5$ and $\beta_1 = -0.1$

Let’s consider the second observation:

- $T_C = 4.1$
- $T_T = 28.5$
- Choice = *transit*

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

\[
V_C = \beta_1 T_C = -0.41 \\
V_T = \beta_1 T_T + \beta_0 = -2.35
\]
Back to the example

\[ P(\text{transit}) = \frac{e^{VT}}{e^{VT} + e^{VC}} \]

\[ P(\text{transit}) = \frac{e^{-2.35}}{e^{-2.35} + e^{-0.41}} \approx 0.13 \]

The model does not correctly predict this observation
The probability that the model reproduces both observations is

\[ P_1(\text{transit})P_2(\text{transit}) = 0.13 \]

The probability that the model reproduces all observations is

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

In general

\[ \mathcal{L}^* = \prod \limits_n (P_n(\text{auto})^{y_{\text{auto},n}} P_n(\text{transit})^{y_{\text{transit},n}}) \]

where \( y_{j,n} \) is 1 if individual \( n \) has chosen alternative \( j \), 0 otherwise
$\mathcal{L}^*$ is called the likelihood of the sample for a given model. It is a probability. We report this value for some values of $\beta_0$ and $\beta_1$

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\mathcal{L}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$4.57 \times 10^{-7}$</td>
</tr>
<tr>
<td>0</td>
<td>-1</td>
<td>$1.97 \times 10^{-30}$</td>
</tr>
<tr>
<td>0</td>
<td>-0.1</td>
<td>$4.1 \times 10^{-4}$</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.1</td>
<td>$4.62 \times 10^{-4}$</td>
</tr>
</tbody>
</table>
Back to the example
Maximum likelihood estimation

\[
\max_\beta \prod_n (P_n(\text{auto})^{y_{\text{auto},n}} P_n(\text{transit})^{y_{\text{transit},n}})
\]

Alternatively, we prefer to maximize the log-likelihood

\[
\max_\beta \ln \prod_n (P_n(\text{auto})^{y_{\text{auto},n}} P_n(\text{transit})^{y_{\text{transit},n}})
\]

\[
\max_\beta \sum_n \ln (y_{\text{auto},n} P_n(\text{auto}) + y_{\text{transit},n} P_n(\text{transit}))
\]
Maximum likelihood estimation
Maximum likelihood estimation

In general, the likelihood of a sample composed of $N$ observations is

$$
\mathcal{L}^*(\beta_1, \ldots, \beta_K) = \prod_{n=1}^{N} P_n(1)^{y_{1n}} P_n(2)^{y_{2n}}
$$

where $y_{1n}$ is 1 if individual $n$ has chosen alternative 1, and 0 otherwise. We also have

$$
P_n(2) = 1 - P_n(1) \text{ and } y_{2n} = 1 - y_{1n}
$$
Maximum likelihood estimation

The log-likelihood is more convenient:

\[ \mathcal{L}(\beta_1, \ldots, \beta_K) = \sum_{n=1}^{N} (y_{1n} \log P_n(1) + y_{2n} \log P_n(2)) \]

Problem to solve

\[ \max_{\beta \in \mathbb{R}^K} \mathcal{L}(\beta) \]
Logit model

Binary logit:

\[ P(i|\{i,j\}) = \frac{e^{V_i}}{e^{V_i} + e^{V_j}} \]

Logit: assumes that the error terms are i.i.d. EV

\[ P(i|C_n = \{1, 2, \ldots, i, \ldots, J\}) = \frac{e^{V_i}}{\sum_{j \in C_n} e^{V_j}} \]

Property:

- If the \( V_i \) are linear-in-parameter
- the likelihood function for the logit model is globally concave.
- Other choice models do no have this nice property
Example

Transportation mode choice:

\[
V_{\text{car}} = \beta_C \text{cost}_{\text{car}} + \beta_T \text{time}_{\text{car}} \\
V_{\text{bus}} = \beta_C \text{cost}_{\text{bus}} + \beta_T \text{time}_{\text{bus}} \\
V_{\text{bike}} = \beta_C \text{cost}_{\text{bike}} + \beta_T \text{time}_{\text{bike}}
\]

- $\beta_{\text{time}}$ can be different depending on the choice context
- Example: trip purpose

\[
V_{\text{car}} = \beta_C \text{cost}_{\text{car}} + \beta_T^{b} \text{time}_{\text{car}} \delta(\text{business}) + \beta_T^{s} \text{time}_{\text{car}} \delta(\text{shopping}) \\
V_{\text{bus}} = \beta_C \text{cost}_{\text{bus}} + \beta_T^{b} \text{time}_{\text{bus}} \delta(\text{business}) + \beta_T^{s} \text{time}_{\text{car}} \delta(\text{shopping}) \\
V_{\text{bike}} = \beta_C \text{cost}_{\text{bike}} + \beta_T^{b} \text{time}_{\text{bike}} \delta(\text{business}) + \beta_T^{s} \text{time}_{\text{car}} \delta(\text{shopping})
\]
Make it more complex: mixtures

- $\beta_T$ is distributed in the population
- We cannot characterize the segmentation
- Assume a distribution, e.g. normal

$$\beta_T \sim N(\bar{\beta}_T, \sigma^2_T)$$

- If $\beta_T$ were known, we would have

$$P(\text{car}|\beta_T, C_n) = \frac{e^{V_{\text{car}}}}{\sum_{j \in C_n} V_j}$$

- But $\beta_T$ is distributed. Therefore,

$$P(\text{car}|C_n) = \int_{\beta_T} P(\text{car}|\beta_T, C_n) f(\beta_T) d\beta_T$$
Mixtures

In statistics, a mixture probability distribution function is a convex combination of other probability distribution functions. If $f(\varepsilon, \theta)$ is a distribution function, and if $w(\theta)$ is a non negative function such that

$$\int_\theta w(\theta)d\theta = 1$$

then

$$g(\varepsilon) = \int_\theta w(\theta)f(\varepsilon, \theta)d\theta$$

is also a distribution function. We say that $g$ is a $w$-mixture of $f$. If $f$ is a logit model, $g$ is a continuous $w$-mixture of logit. If $f$ is a MEV model, $g$ is a continuous $w$-mixture of MEV.
Mixtures

Discrete mixtures are also possible. If $w_i, i = 1, \ldots, n$ are non-negative weights such that

$$\sum_{i=1}^{n} w_i = 1$$

then

$$g(\varepsilon) = \sum_{i=1}^{n} w_i f(\varepsilon, \theta_i)$$

is also a distribution function where $\theta_i, i = 1, \ldots, n$ are parameters. We say that $g$ is a discrete $w$-mixture of $f$. 

Example: discrete mixture of normal distributions

$0.6 \ N(5,0.16) + 0.4 \ N(8,1)$
Example: discrete mixture of binary logit models

\[ P(1|s=1,x) + 0.6 \ P(1|s=2,x) \]
Simulation

\[ P(i) = \int_{\xi} \Lambda(i|\xi) f(\xi) d\xi \]

No closed form formula

- Randomly draw numbers such that their frequency matches the density \( f(\xi) \)
- Let \( \xi^1, \ldots, \xi^R \) be these numbers
- The choice model can be approximated by

\[ P(i) \approx \tilde{P}(i) = \frac{1}{R} \sum_{r=1}^{R} \Lambda(i|\xi^r), \text{ as} \]

\[ \lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} \Lambda(i|\xi^r) = \int_{\xi} \Lambda(i|\xi) f(\xi) d\xi \]
Simulation

\[ P(i) \approx \tilde{P}(i) = \frac{1}{R} \sum_{r=1}^{R} \Lambda(i|\xi^r). \]

The kernel is a logit model, easy to compute.

\[ \Lambda(i|\xi^r) = \frac{e^{V_{1n}+\xi^r}}{e^{V_{1n}+\xi^r} + e^{V_{2n}+\xi^r} + e^{V_{3n}}} \]

Therefore, it amounts to generating the appropriate draws.
Simulation: uniform distribution

- Almost all programming languages provide generators for a uniform $U(0, 1)$
- If $r$ is a draw from a $U(0, 1)$, then

\[ s = (b - a)r + a \]

is a draw from a $U(a, b)$
Simulation: standard normal

- If $r_1$ and $r_2$ are independent draws from $U(0, 1)$, then

$$s_1 = \sqrt{-2 \ln r_1 \sin(2\pi r_2)}$$
$$s_2 = \sqrt{-2 \ln r_1 \cos(2\pi r_2)}$$

are independent draws from $N(0, 1)$
Simulation: standard normal

Histogram of 100 random samples from a univariate Gaussian PDF with unit variance and zero mean

scaled bin frequency
Gaussian p.d.f.
Simulation: standard normal

Histogram of 500 random samples from a univariate Gaussian PDF with unit variance and zero mean
Simulation: standard normal

Histogram of 1000 random samples from a univariate Gaussian PDF with unit variance and zero mean
Simulation: standard normal

Histogram of 5000 random samples from a univariate Gaussian PDF with unit variance and zero mean

Gaussian p.d.f.
Simulation: standard normal

Histogram of 10000 random samples from a univariate Gaussian PDF with unit variance and zero mean
Maximum simulated likelihood

\[
\max_{\theta} \mathcal{L}(\theta) = \sum_{n=1}^{N} \left( \sum_{j=1}^{J} y_{jn} \ln \tilde{P}(j; \theta) \right) = \sum_{n=1}^{N} \left( \sum_{j=1}^{J} y_{jn} \ln \frac{1}{R} \sum_{r=1}^{R} \Lambda(i|\xi^r) \right)
\]

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

Vector of parameters \(\theta\) contains:

- usual (fixed) parameters of the choice model
- parameters of the density of the random parameters
- For instance, if \(\beta_j \sim N(\mu_j, \sigma_j^2)\), \(\mu_j\) and \(\sigma_j\) are parameters to be estimated
Maximum simulated likelihood

Warning:

- $\tilde{P}(j; \theta)$ is an unbiased estimator of $P(j; \theta)$

$$E[\tilde{P}_n(j; \theta)] = P(j; \theta)$$

- $\ln \tilde{P}(j; \theta)$ is not an unbiased estimator of $\ln P(j; \theta)$

$$\ln E[\tilde{P}(j; \theta) \neq E[\ln \tilde{P}(j; \theta)]$$

- Under some conditions, it is a consistent (asymptotically unbiased) estimator, so that many draws are necessary.
Maximum simulated likelihood

In practice:

- Generate the draws once for all.
- The function to maximize has a closed form

\[
\max_{\theta} \mathcal{L}(\theta) = \sum_{n=1}^{N} \left( \sum_{j=1}^{J} y_{jn} \ln \left( \frac{1}{R} \sum_{r=1}^{R} \Lambda(j|\xi^r; \theta) \right) \right)
\]

- The value of \( R \) can be adjusted during the algorithm: Bastin et al.
Summary

- Behavioral theory based on utility maximization
- Operational models derived from random utility
- Simplest model: the logit model
- Estimation of the parameters: maximum likelihood
- Maximize a globally concave nonlinear differentiable function
- With other models, same thing without concavity
- Advanced models: mixtures
- Rely on simulation
- Time consuming