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?

Introduction

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

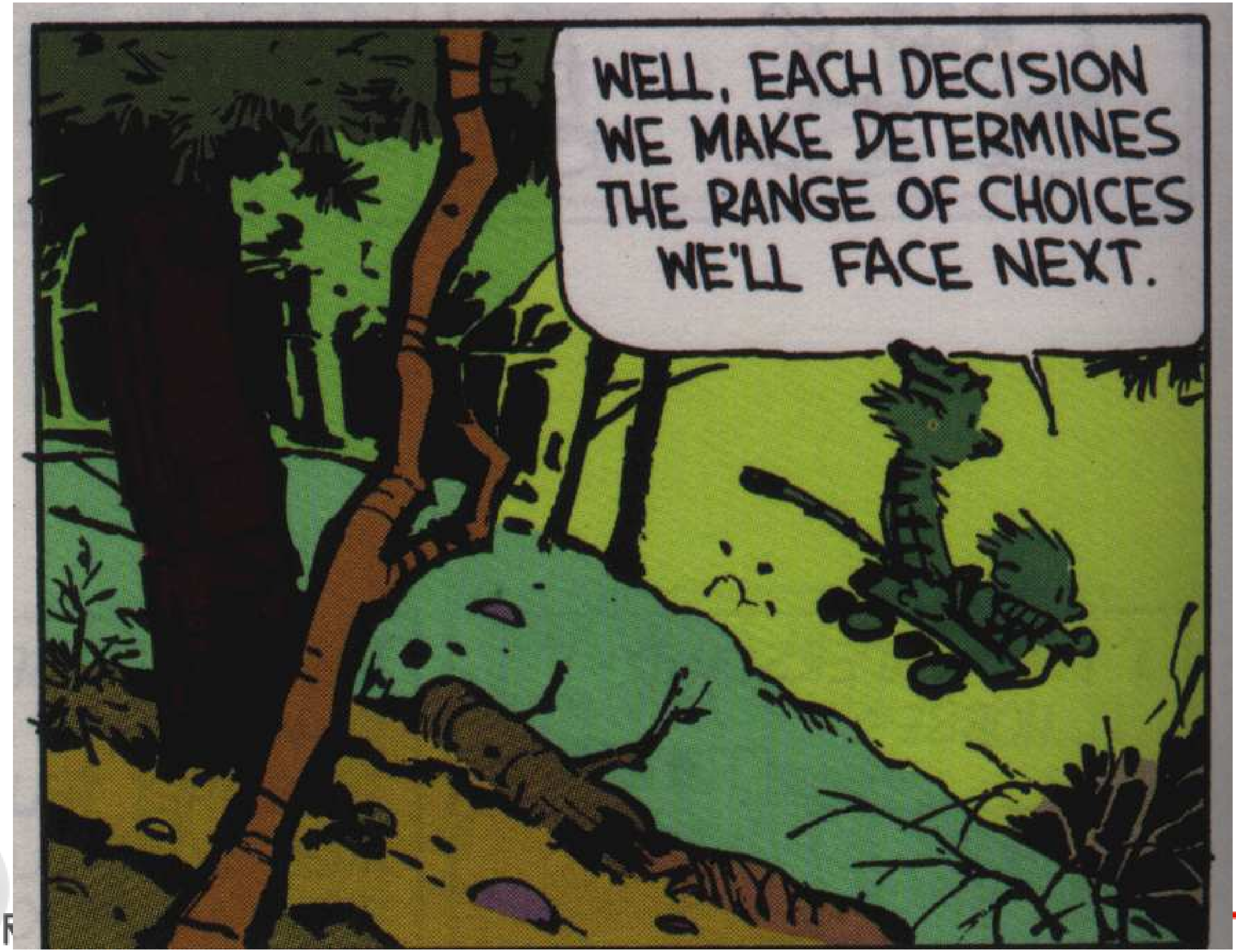
Motivation: shorten the period of barbarism after the Fall of the Galactic Empire

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



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 :	Type of behavior:
▶ Marketing	▶ Choice of a brand
▶ Transportation	▶ Choice of a transportation mode
▶ Politics	▶ Choice of a president
▶ Management	▶ Choice of a management policy
▶ New technologies	▶ Choice of investments

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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:

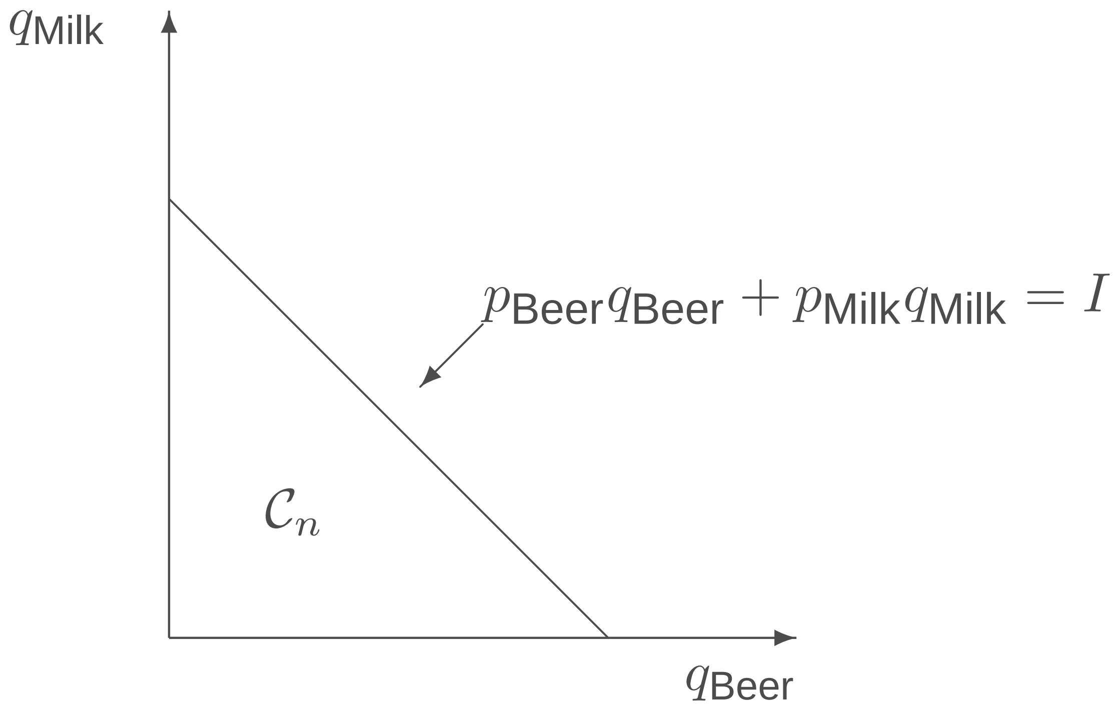
- Availability
- Awareness

Swait, J. (1984) *Probabilistic Choice Set Formation in Transportation Demand Models*
Ph.D. dissertation, Department of Civil Engineering, MIT, Cambridge, Ma.

Alternatives

Continuous vs. discrete

Continuous choice set:



Discrete choice set:

$$\mathcal{C}_n = \{\text{ Car, Bus, Bike}\}$$

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 $\gtrsim$

(i) reflexivity

$$a \gtrsim a \quad \forall a \in \mathcal{C}_n$$

(ii) transitivity

$$a \gtrsim b \text{ and } b \gtrsim c \Rightarrow a \gtrsim c \quad \forall a, b, c \in \mathcal{C}_n$$

(iii) comparability

$$a \gtrsim b \text{ or } b \gtrsim a \quad \forall a, b \in \mathcal{C}_n$$

Decision rules

Neoclassical economic theory (ctd)

☞ Numerical function

$$\exists\, U_n : \mathcal{C}_n \longrightarrow \mathbb{R} : a \rightsquigarrow U_n(a) \text{ such that}$$

$$a \gtrsim b \Leftrightarrow U_n(a) \geq U_n(b) \quad \forall a, b \in \mathcal{C}_n$$

Utility

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:

$$U(q_1, \ldots, q_L)$$

- $Q_a \gtrsim Q_b$ 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

$$\begin{array}{rcrcl}\beta_0\beta_1 q_1^{\beta_1-1}q_2^{\beta_2} & - & \lambda p_1 & = & 0\\ \beta_0\beta_2 q_1^{\beta_1}q_2^{\beta_2-1} & - & \lambda p_2 & = & 0\\ p_1q_1+p_2q_2 & - & I & = & 0.\end{array}$$

We have

$$\begin{array}{rcrcl}\beta_0\beta_1 q_1^{\beta_1}q_2^{\beta_2} & - & \lambda p_1 q_1 & = & 0\\ \beta_0\beta_2 q_1^{\beta_1}q_2^{\beta_2} & - & \lambda p_2 q_2 & = & 0\end{array}$$

so that

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

$$U_i \geq U_j \quad \forall j.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Example

Example: two transportation modes

$$\begin{aligned} U_1 &= -\beta t_1 - \gamma c_1 \\ U_2 &= -\beta t_2 - \gamma c_2 \end{aligned}$$

with β, $\gamma > 0$

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

that is

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

or

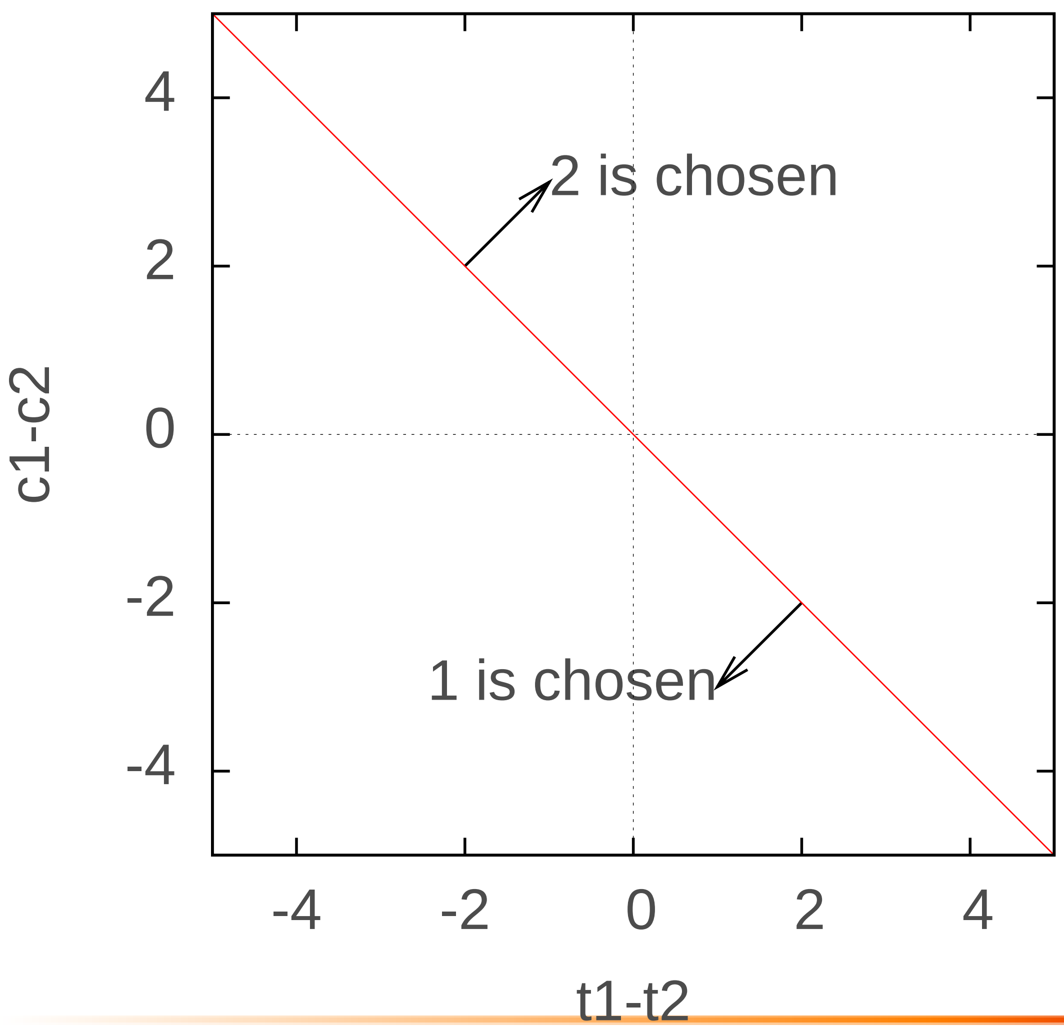
$$c_1 - c_2 \leq -\frac{\beta}{\gamma}(t_1 - t_2)$$

Example

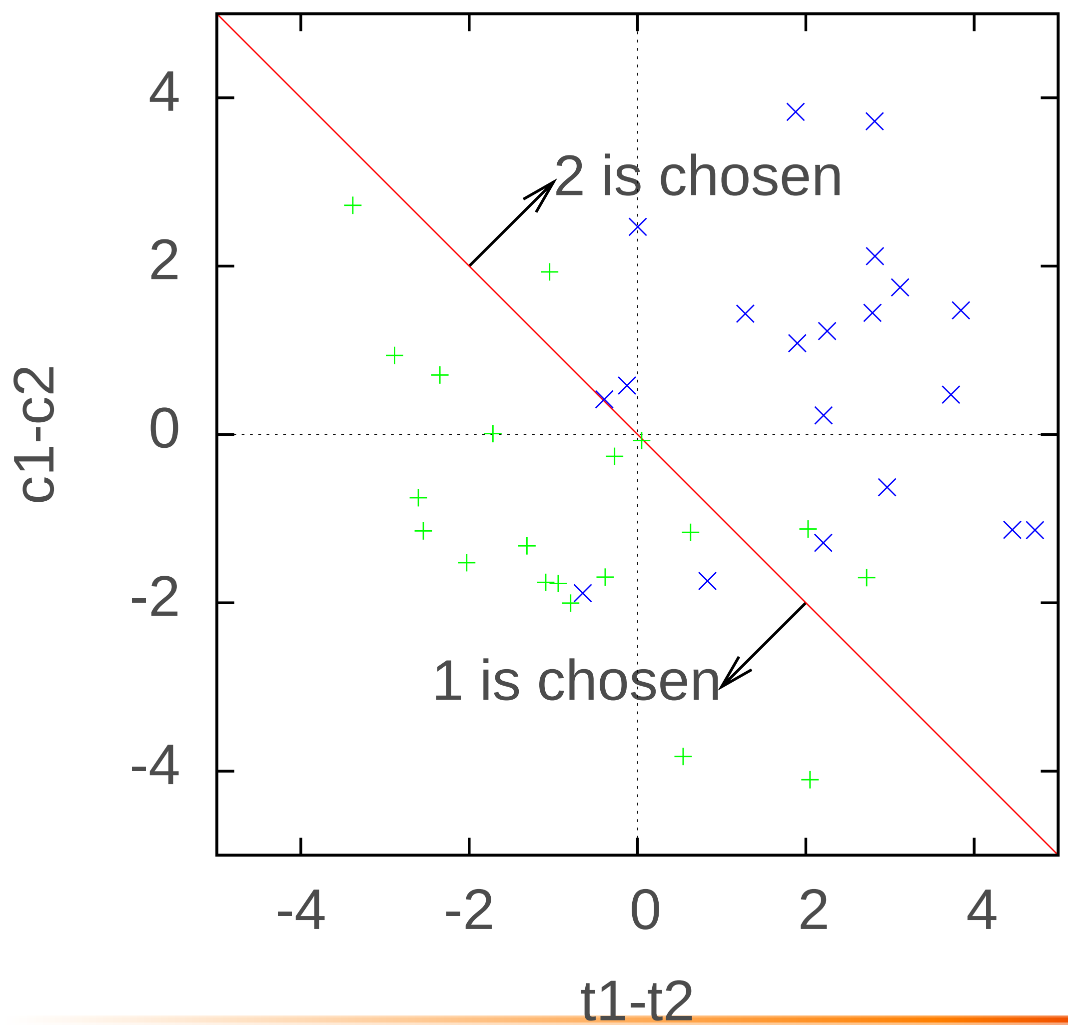
Obvious cases:

- $c_1 \geq c_2$ 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



Example



Assumptions

Decision rules

Neoclassical economic theory (ctd)

Decision-maker

- ✔ perfect discriminating capability
- ✔ full rationality
- ✔ permanent consistency

Analyst

- ✔ knowledge of all attributes
- ✔ perfect knowledge of $\gtrsim$ (or $U_n(\cdot)$)
- ✔ no measurement error

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

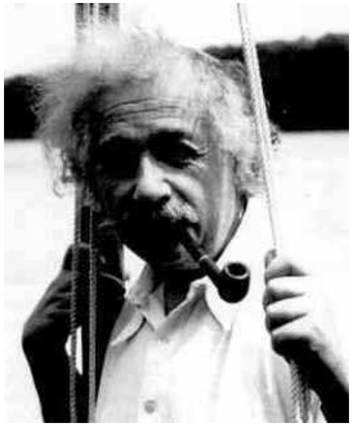
Assumptions

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

Manski 1973 The structure of Random Utility Models *Theory and Decision* 8:229–254

Sources of uncertainty:

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

For each individual n,

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

Dependent variable is latent. Therefore, we prefer the model

$$P(i|\mathcal{C}_n) = P[U_{in} = \max_{j \in \mathcal{C}_n} U_{jn}] = P(U_{in} \geq U_{jn} \; \forall j \in \mathcal{C}_n)$$

Example

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

Error term

The distribution

Assumption: 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.

$$\varepsilon_C \sim \text{EV}(0, \mu)$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Error term

EV(η, μ), with $\mu > 0$:

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

If $\varepsilon \sim$ EV(η, μ), then

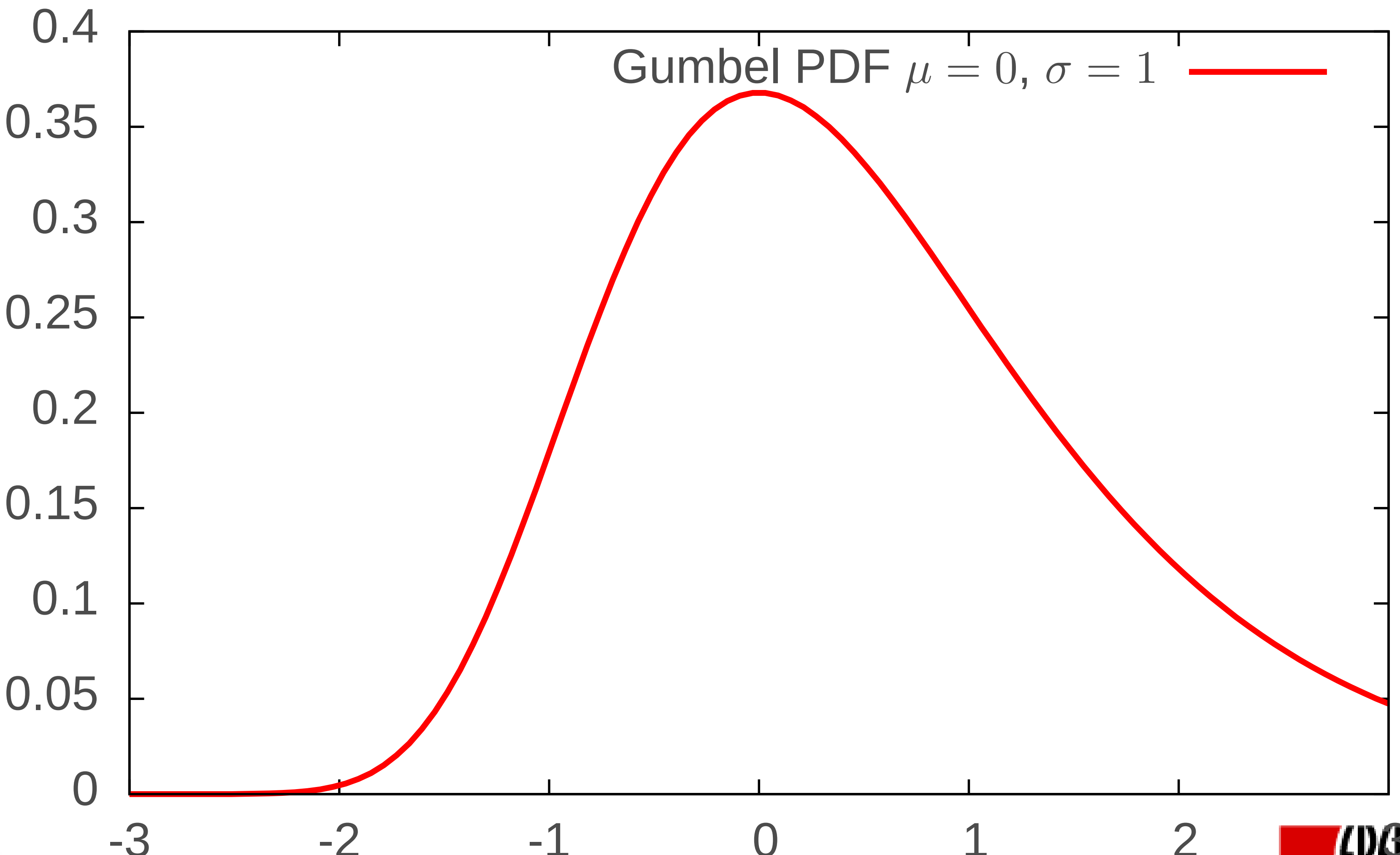
The distribution

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

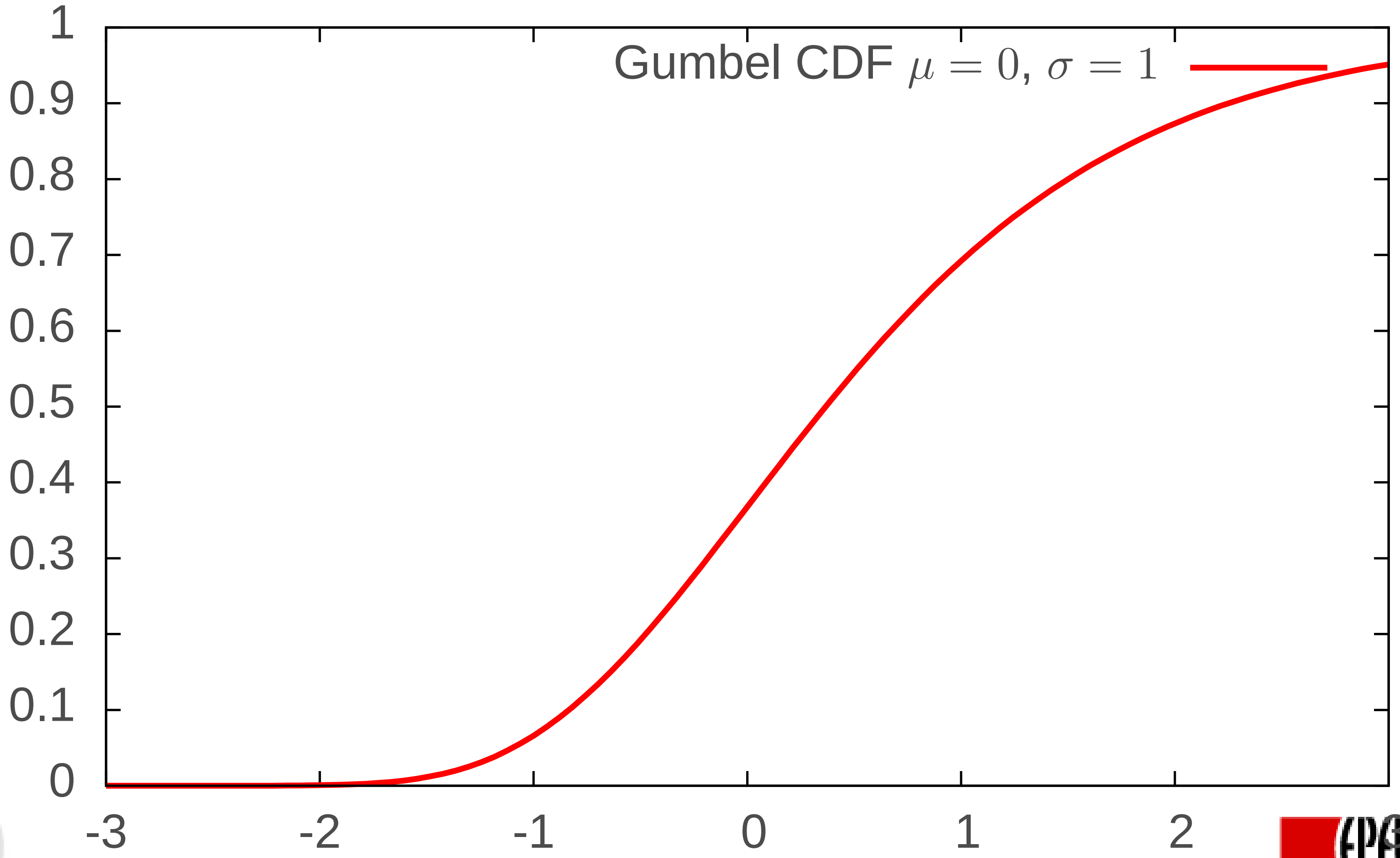
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Error term



Error term



Error term

If

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

then

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

where γ is Euler's constant

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

Error term

The distribution

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

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

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

Logit Model

Error term

The distribution

For the Logistic(0,μ), we have

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

$$\begin{aligned}
P(C|\{C,T\}) &= P\left(\varepsilon \leq V_C - V_T\right) \\
&= F(V_C - V_T) \\
&= \frac{1}{1+e^{-\mu(V_C - V_T)}} \\
&= \frac{e^{\mu V_C}}{e^{\mu V_C} + e^{\mu V_T}}
\end{aligned}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Error term

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

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*?

$$\begin{array}{rclcl} V_C & = & \beta_1 T_C & = & -5.29 \\ V_T & = & \beta_1 T_T + \beta_0 & = & 0.06 \end{array}$$

Back to the example

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

$$P(\text{transit}) = \frac{e^{0.06}}{e^{0.06} + e^{-5.29}} \cong 1$$

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*?

$$\begin{aligned} V_C &= \beta_1 T_C &= -0.41 \\ V_T &= \beta_1 T_T + \beta_0 &= -2.35 \end{aligned}$$

Back to the example

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

$$P(\text{transit}) = \frac{e^{-2.35}}{e^{-2.35} + e^{-0.41}} \cong 0.13$$

The model does not correctly predict this observation

Back to the example

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

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

In general

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

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

Back to the example

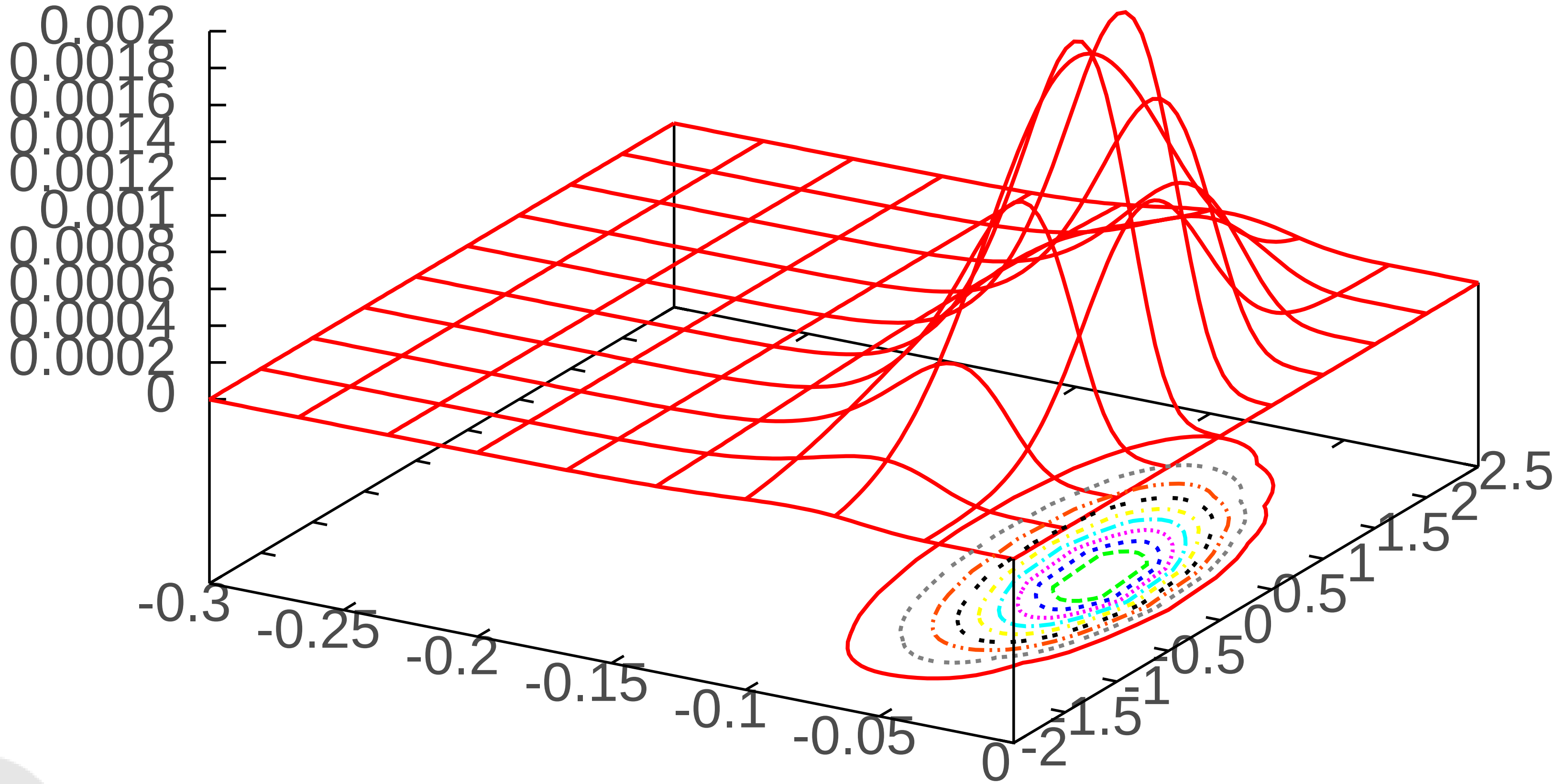
$\mathcal{L}^*$ is called the likelihood of the sample for a given model.
It is a probability.
We report this value for some values of β_0 and β_1

β_0	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}

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Back to the example



Maximum likelihood estimation

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

Alternatively, we prefer to maximize the log-likelihood

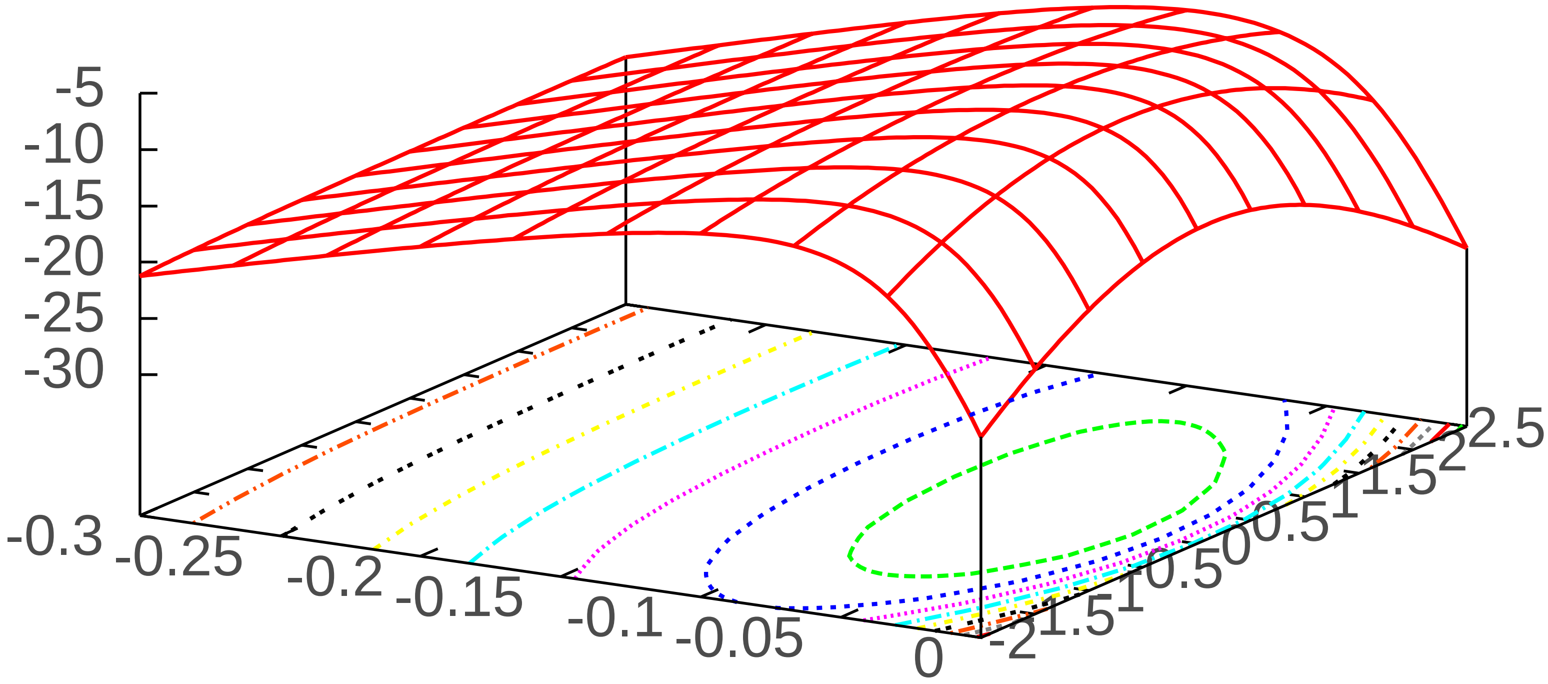
$$\max_{\beta} \ln \prod_{n} \left(P_n(\text{auto})^{y_{\text{auto},n}} P_n(\text{transit})^{y_{\text{transit},n}} \right)$$

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

$$P(i|\mathcal{C}_n = \{1,2,\ldots,i,\ldots,J\}) = \frac{e^{V_i}}{\sum_{j \in C_n} e^{V_j}}$$

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Example

Transportation mode choice:

$$
\begin{array}{lcl}
V_{\text{car}} & = & \beta_C \text{cost}_{\text{car}} + \beta_T \text{time}_{\text{car}} \\
V_{\text{bus}} & = & \beta_C \text{cost}_{\text{bus}} + \beta_T \text{time}_{\text{bus}} \\
V_{\text{bike}} & = & \beta_C \text{cost}_{\text{bike}} + \beta_T \text{time}_{\text{bike}}
\end{array}
$$

- β_{time} can be different depending on the choice context
- Example: trip purpose

$$
\begin{array}{lcl}
V_{\text{car}} & = & \beta_C \text{cost}_{\text{car}} + \beta_T^b \text{time}_{\text{car}} \delta(\text{business}) + \beta_T^s \text{time}_{\text{car}} \delta(\text{shopp} \\
V_{\text{bus}} & = & \beta_C \text{cost}_{\text{bus}} + \beta_T^b \text{time}_{\text{bus}} \delta(\text{business}) + \beta_T^s \text{time}_{\text{car}} \delta(\text{shopp} \\
V_{\text{bike}} & = & \beta_C \text{cost}_{\text{bike}} + \beta_T^b \text{time}_{\text{bike}} \delta(\text{business}) + \beta_T^s \text{time}_{\text{car}} \delta(\text{shopp}
\end{array}
$$

Make it more complex: mixtures

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

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

- If β_T were known, we would have

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

- But β_T is distributed. Therefore,

$$P(\text{car}|\mathcal{C}_n) = \int_{\beta_T} P(\text{car}|\beta_T, \mathcal{C}_n) f(\beta_T) d\beta_T$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

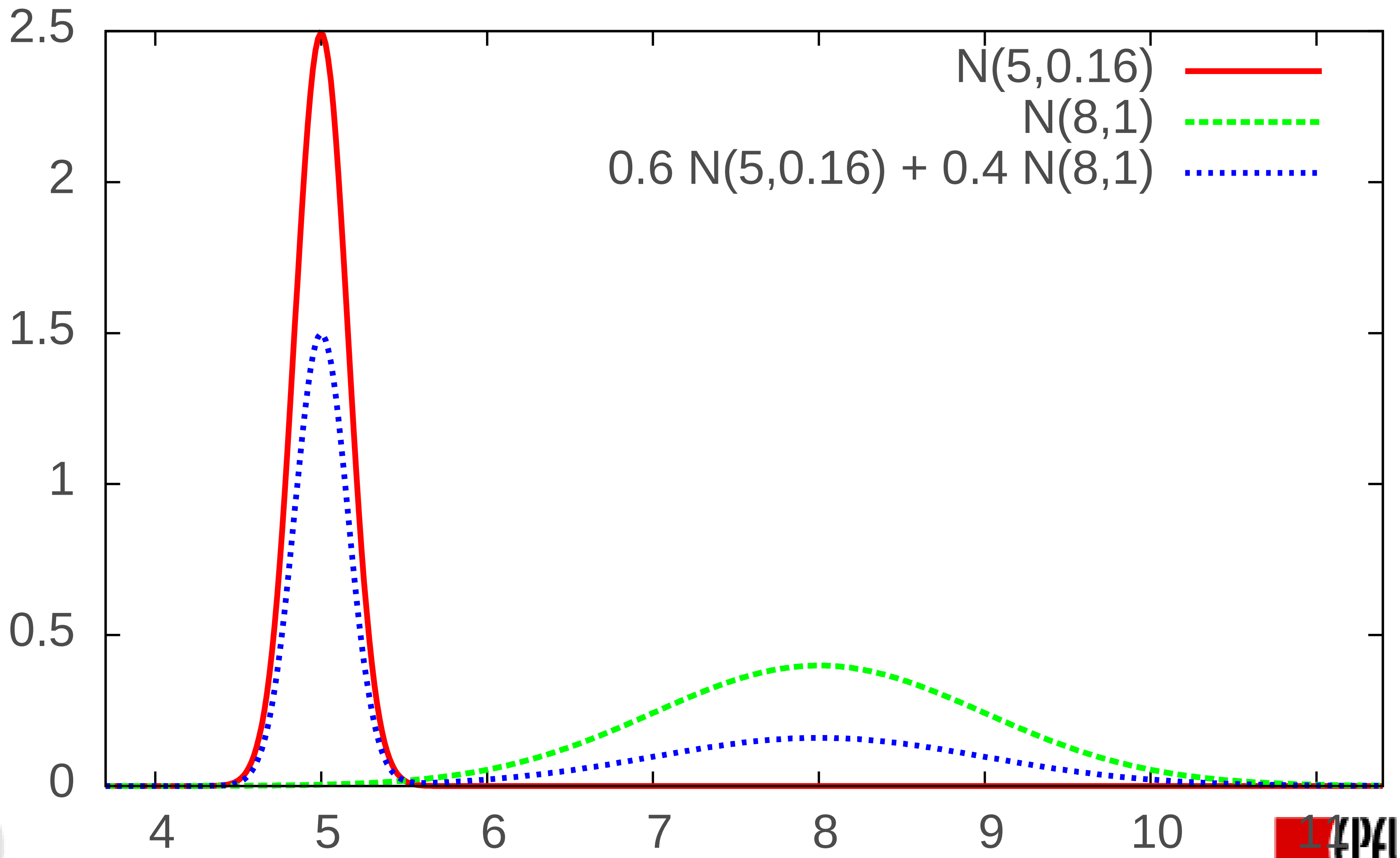
$$\sum_{i=1}^{n} w_i = 1$$

then

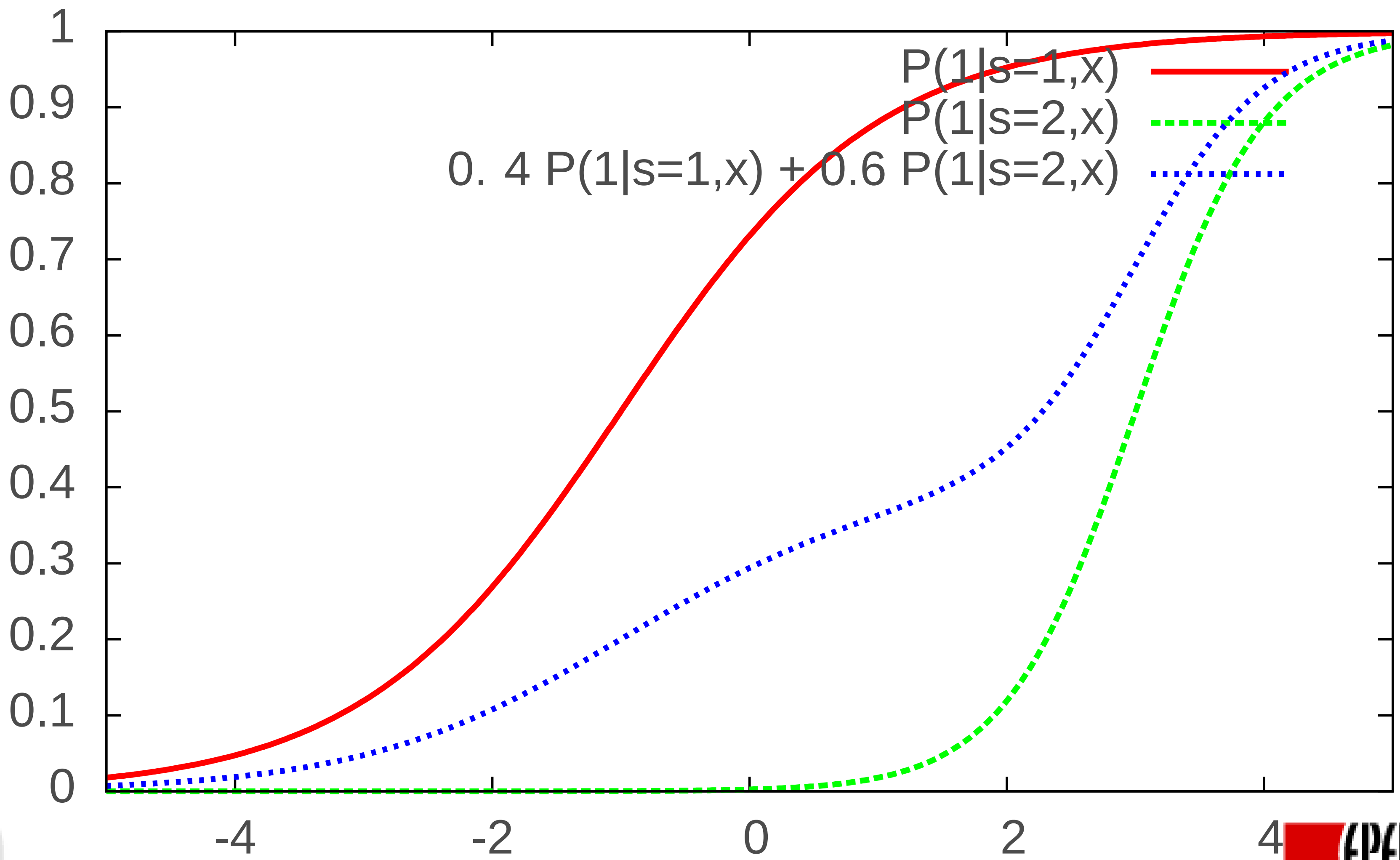
$$g(\varepsilon) = \sum_{i=1}^{n} w_i f(\varepsilon, \theta_i)$$

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

Example: discrete mixture of normal distributions



Example: discrete mixture of binary logit models



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 ξ^1,...,ξ^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 $U(a,b)$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Simulation: standard normal

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

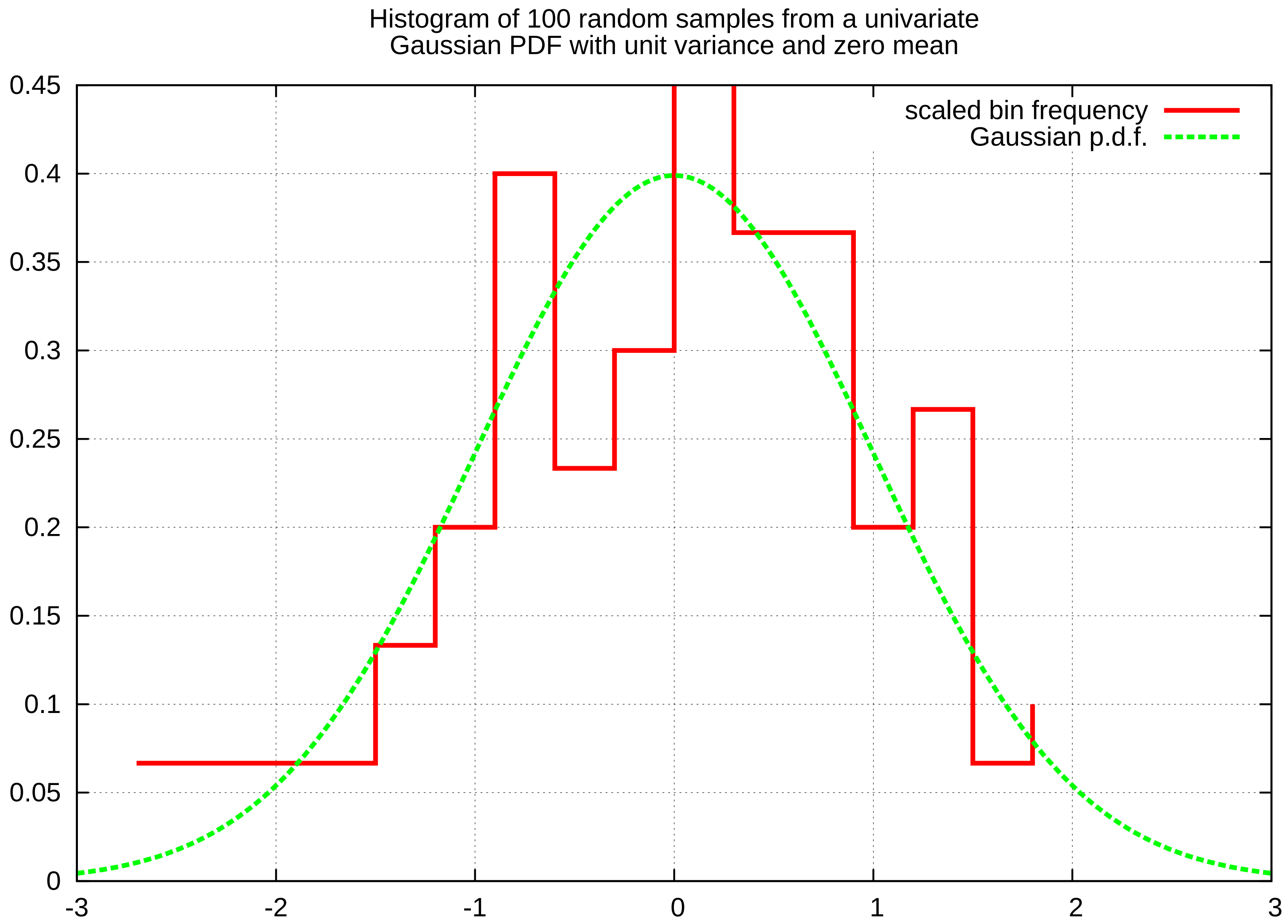
$$s_1 = \sqrt{-2\ln r_1}\,\sin(2\pi r_2)$$
$$s_2 = \sqrt{-2\ln r_1}\,\cos(2\pi r_2)$$

are independent draws from $N(0,1)$

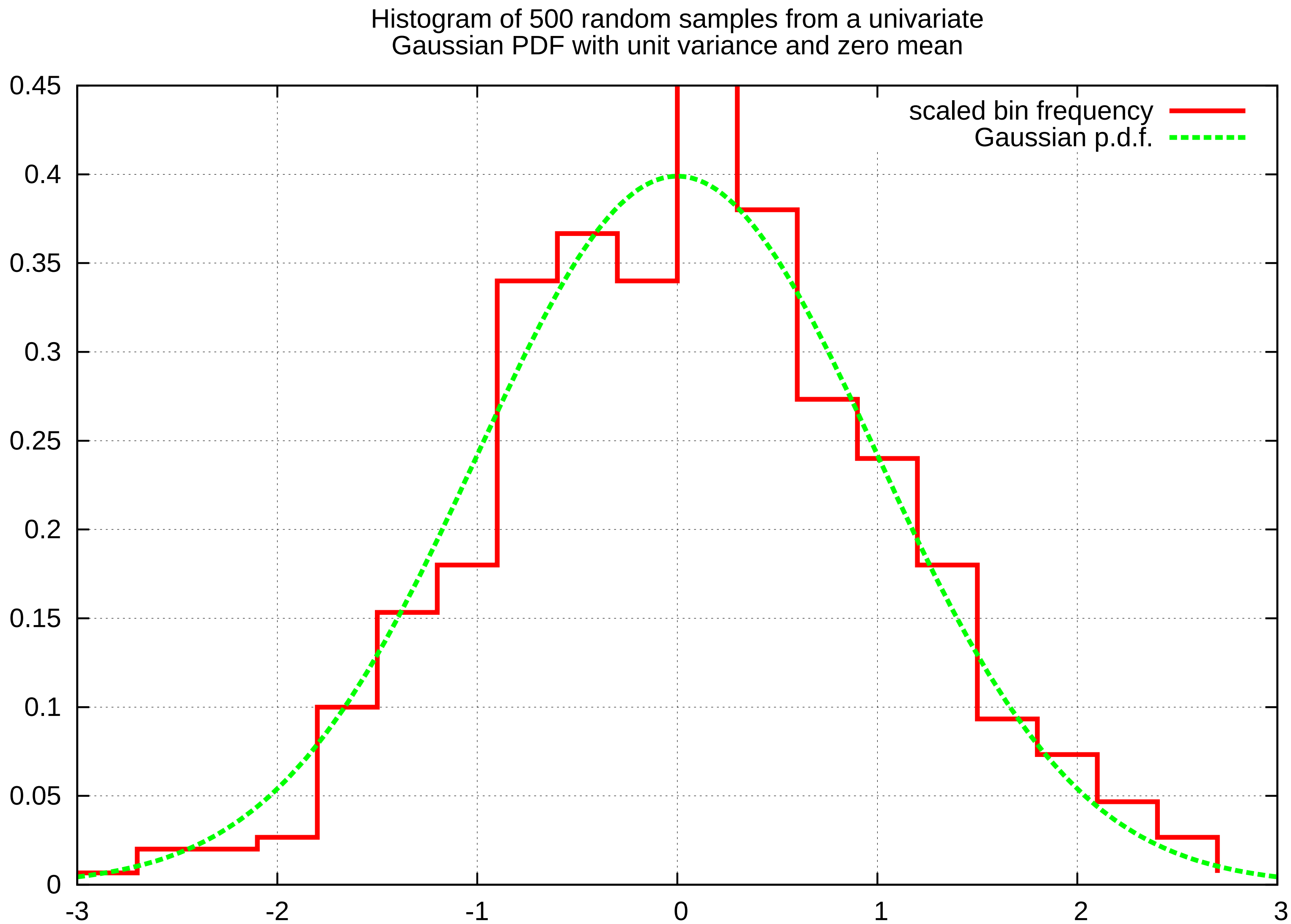
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

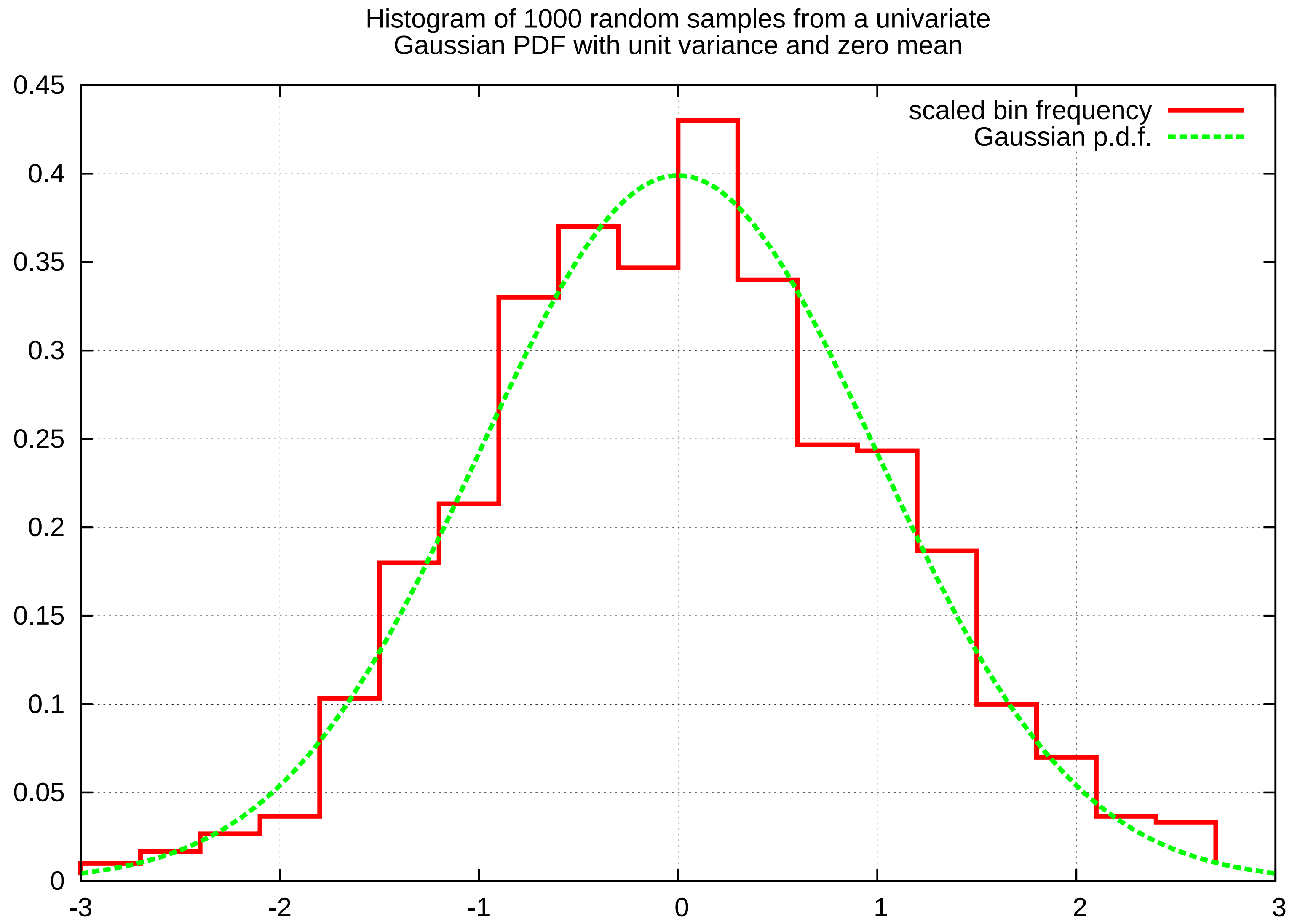
Simulation: standard normal



Simulation: standard normal



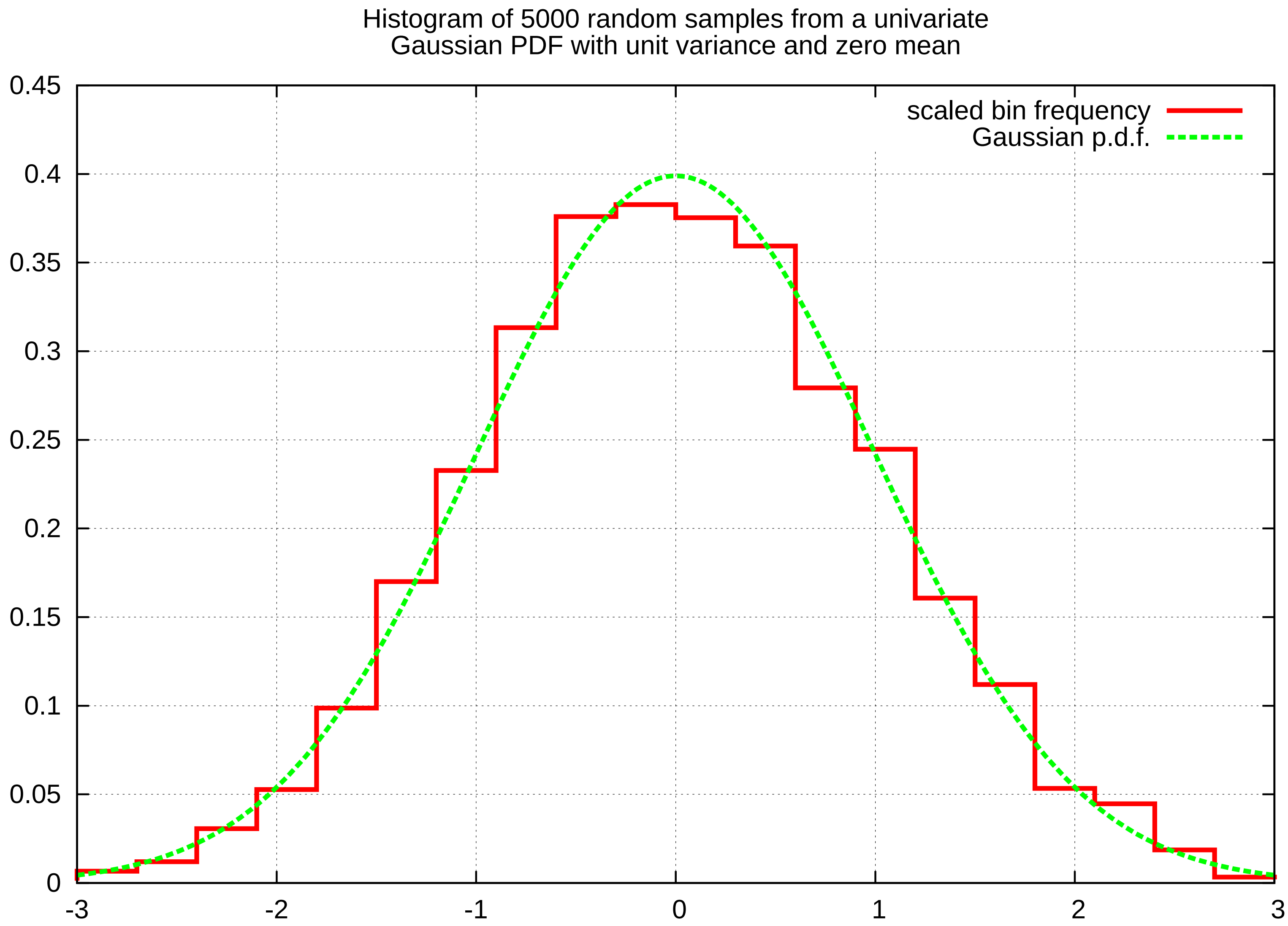
Simulation: standard normal



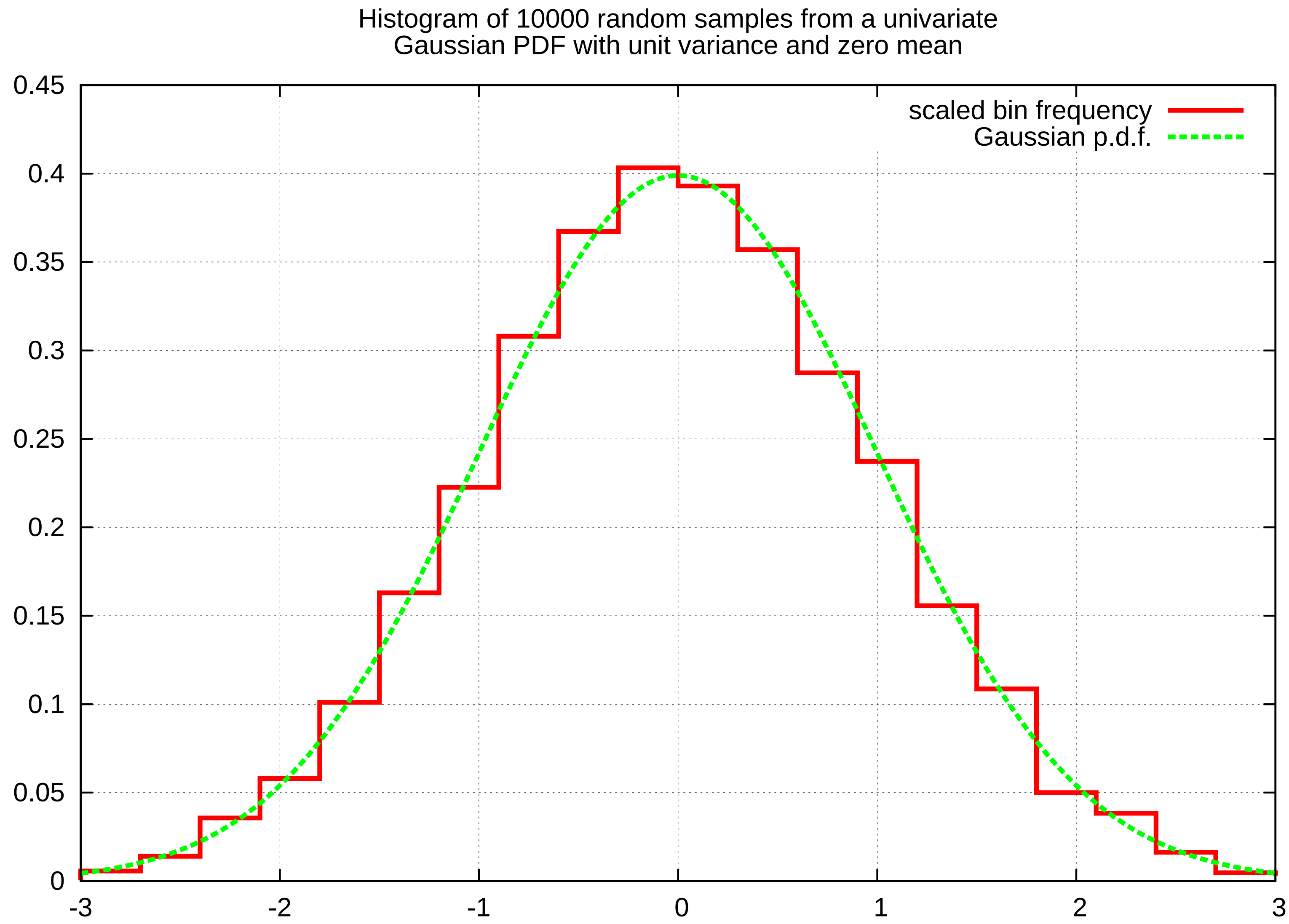
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Simulation: standard normal



Simulation: standard normal



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 θ contains:

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

Short course

Organized at EPFL, Lausanne, March 2011

M. Ben Akiva D. McFadden M. Bierlaire D. Bolduc

http://transp-or.epfl.ch/dca