Mixtures

Models and simulation-based estimation

Michel Bierlaire

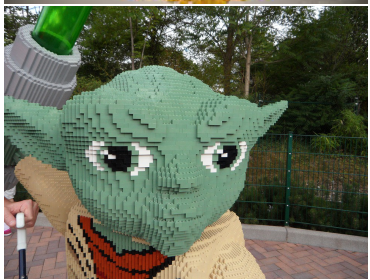
Mathematical Modeling of Behavior

EPFL

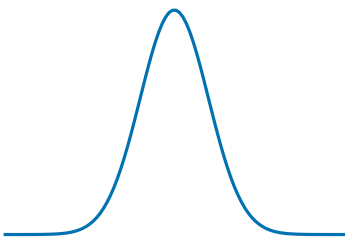
Outline

Motivation

Motivation



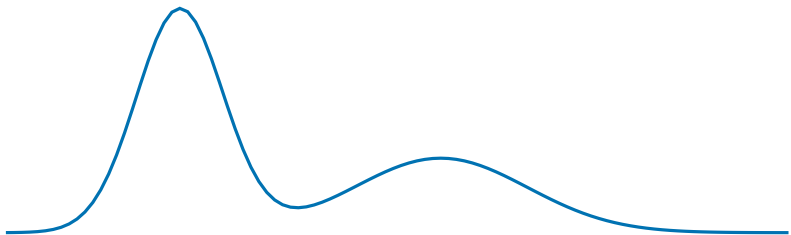
Motivation



Classical distributions

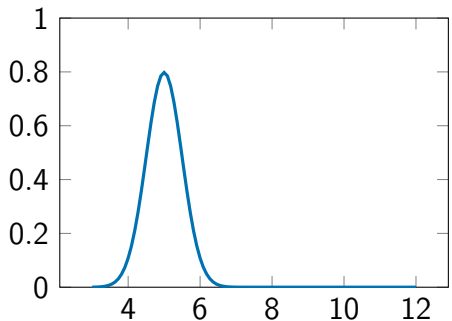
- Simple.
- Usually unimodal.
- May not capture well the complexity of the underlying phenomenon.

Motivation

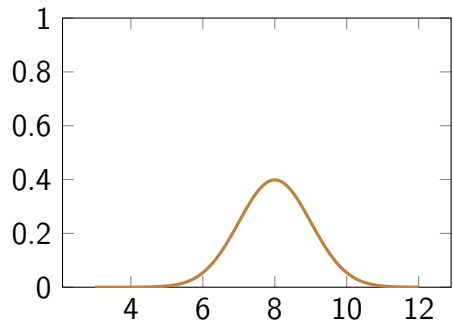


Mixtures: an example

$N(5, 0.5)$

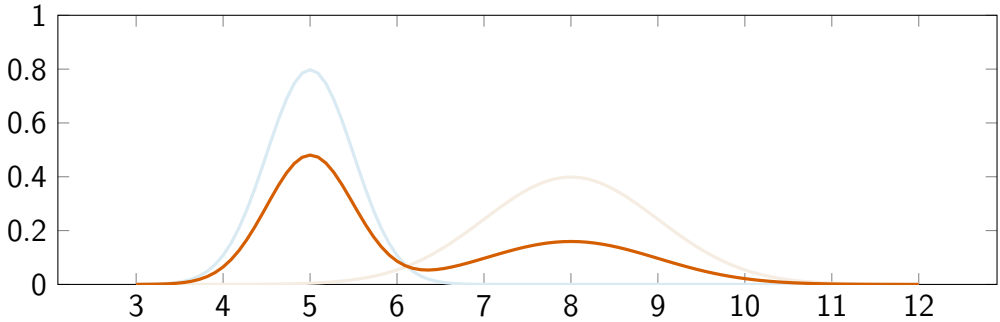


$N(8, 1)$



Mixtures: an example

$0.6\ N(5, 0.5) + 0.4\ N(8, 1)$



Mixtures

Mixture probability distribution function

Convex combination of other probability distribution functions.

Building blocks — kernel

Consider $f(\varepsilon, \theta)$ a parametrized family of distribution functions.

Mixtures

Discrete mixtures

If w_i, $i = 1, \ldots, n$ are non negative weights such that

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

Then

$$g(\varepsilon; \theta_1, \ldots, \theta_n) = \sum_{i=1}^{n} w_i f(\varepsilon, \theta_i)$$

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.

Mixtures

Continuous mixtures

- Let $w(\theta)$ be 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 continuous w-mixture of f.

Mixtures: an example

Weights: pdf of exponential distribution

$$w(\theta) = e^{-\theta}.$$

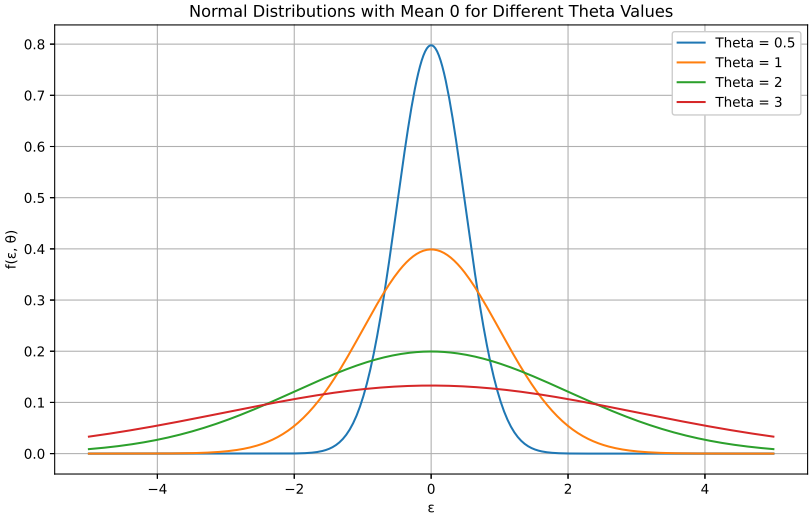
Kernel: normal distribution with mean 0

$$f(\varepsilon, \theta) = \frac{1}{\sqrt{2\pi\theta^2}} e^{-\frac{\varepsilon^2}{2\theta^2}}.$$

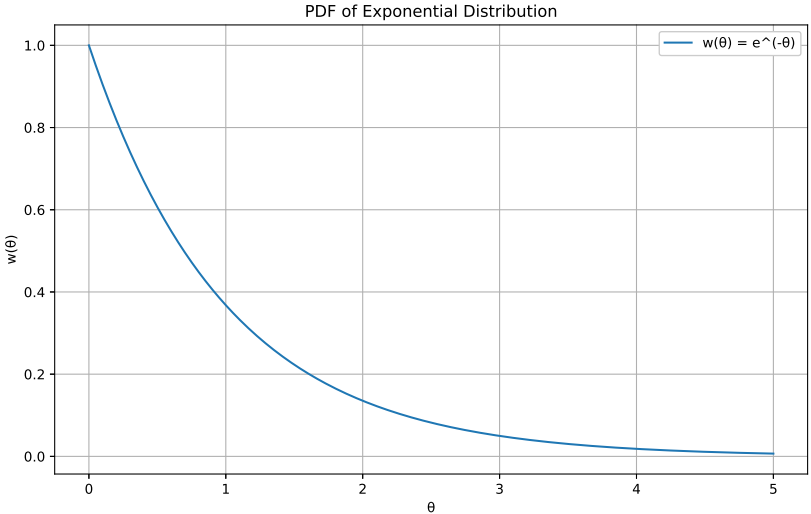
Mixture

$$g(\varepsilon) = \int_\theta e^{-\theta} \frac{1}{\sqrt{2\pi\theta^2}} e^{-\frac{\varepsilon^2}{2\theta^2}} d\theta.$$

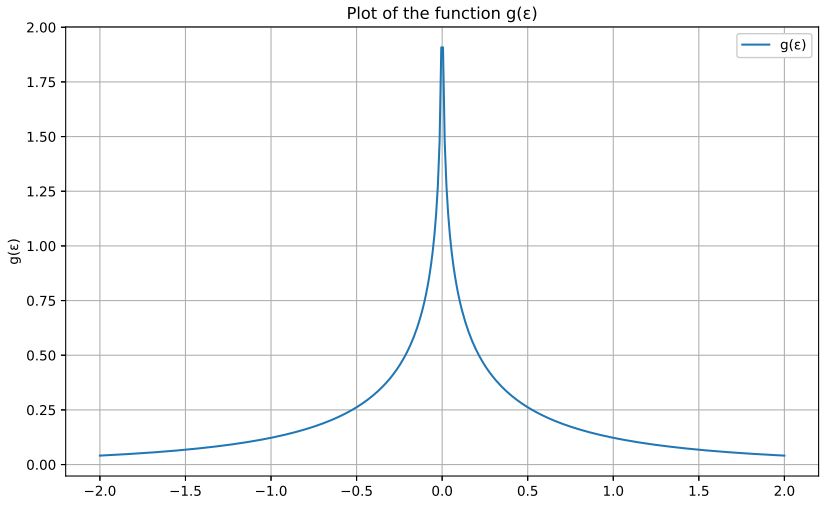
Mixtures: an example



Mixtures: an example



Mixtures: an example



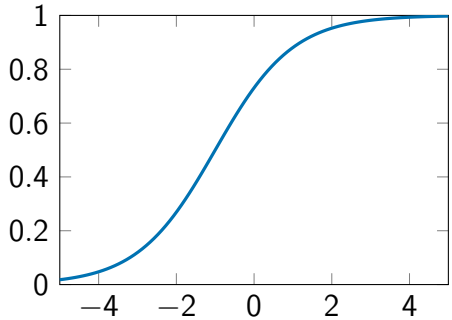
Mixtures of logit

Building blocks

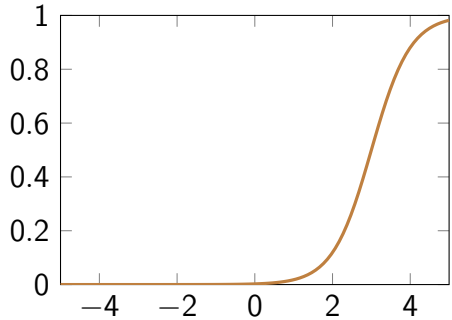
$$P_n(i|x_n, \mathcal{C}_n; \theta) = \frac{e^{V_{in}(x_n;\theta)}}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}(x_n;\theta)}}.$$

Mixtures of logit: an example

$P(i|s=1,x) = 1/(1+\exp(-(1+x)))$

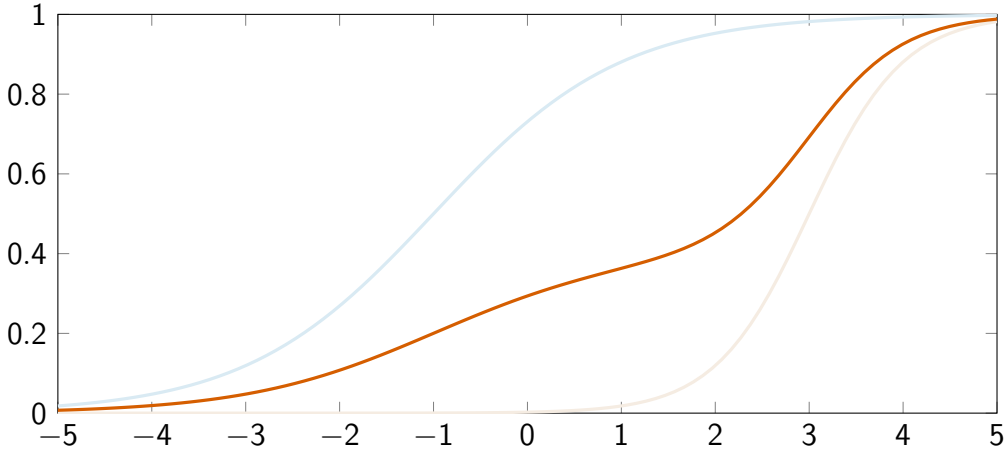


$P(i|s=2,x) = 1/(1+\exp(-(-6+2x)))$



Mixtures of logit: an example

$0.4P(i|s=1,x) + 0.6P(i|s=2,x)$



Combining probit and logit

Error components

$$U_{in} = V_{in} + \xi_{in} + \nu_{in}.$$

i.i.d EV (logit): tractability.

Normal distribution (probit): flexibility.

Combining probit and logit

The choice model

$$\begin{aligned} U_{\text{car},n} &= \beta X_{\text{car},n} + \xi_{\text{car},n} + \nu_{\text{car},n}, \\ U_{\text{train},n} &= \beta X_{\text{train},n} + \xi_{\text{train},n} + \nu_{\text{train},n}, \\ U_{\text{Swissmetro},n} &= \beta X_{\text{Swissmetro},n} + \xi_{\text{Swissmetro},n} + \nu_{\text{Swissmetro},n}. \end{aligned}$$

Distributional assumptions

- $\nu_{i,n}$ i.i.d. extreme value.
- $\xi_n = \begin{pmatrix} \xi_{\text{car},n} \\ \xi_{\text{train},n} \\ \xi_{\text{Swissmetro},n} \end{pmatrix} \sim N(0, \Sigma)$.

Suppose that ξ_n is known...

Combining probit and logit

Suppose that ξ_n is known...

$$\Pr(\text{car}|X_n, \xi_n) = \frac{e^{\beta X_{\text{car},n} + \xi_{\text{car},n}}}{e^{\beta X_{\text{car},n} + \xi_{\text{car},n}} + e^{\beta X_{\text{train},n} + \xi_{\text{train},n}} + e^{\beta X_{\text{Swissmetro},n} + \xi_{\text{Swissmetro},n}}}.$$

Combining probit and logit

Choice model

$$\Pr(\text{car}|X_n, \xi_n) = \frac{e^{\beta X_{\text{car},n} + \xi_{\text{car},n}}}{e^{\beta X_{\text{car},n} + \xi_{\text{car},n}} + e^{\beta X_{\text{train},n} + \xi_{\text{train},n}} + e^{\beta X_{\text{Swissmetro},n} + \xi_{\text{Swissmetro},n}}}.$$

But ξ_n is not known...

$$P(\text{car}|X_n) = \int_{\xi} \Pr(\text{car}|X_n, \xi) f(\xi) d\xi.$$

Integration

$$P(\text{car}|X_n) = \int_{\xi} \Pr(\text{car}|X_n, \xi) f(\xi) d\xi.$$

Problem...

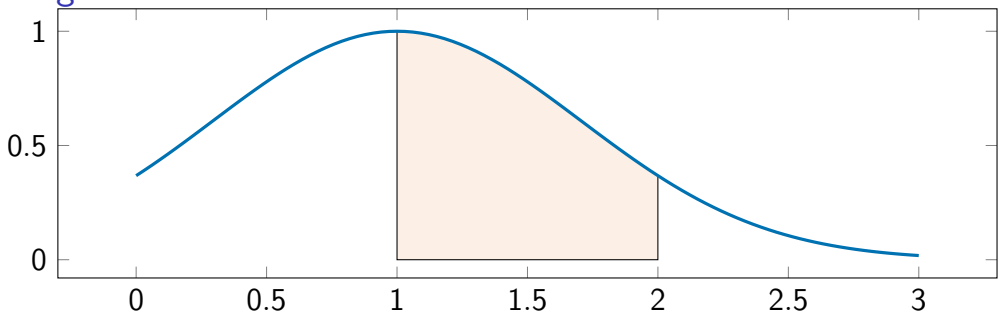
- How to calculate this complicated integral?
- It does not have a closed form.
- Solution: rely on Monte-Carlo integration.

Outline


CASINO
MONTE-CARLO

Monte-Carlo integration

Integral



Random variables

Expectation

- X r.v. with pdf $f_X(x)$.
- Domain of X: $[a, b]$, $a \in \mathbb{R} \cup \{-\infty\}$,$b \in \mathbb{R} \cup \{+\infty\}$.
- Expectation:

$$\mathrm{E}[X] = \int_a^b x\, f_X(x) dx.$$

- If $g : \mathbb{R} \to \mathbb{R}$ is a function, then

$$\mathrm{E}[g(X)] = \int_a^b g(x)\, f_X(x) dx.$$

Simulation

Expectation

$$\mathrm{E}[g(X)] = \int_a^b g(x)\, f_X(x) dx.$$

Draws from X

$$x_r,\ r = 1, \ldots, R.$$

Approximation

$$\mathrm{E}[g(X)] \approx \frac{1}{R} \sum_{r=1}^{R} g(x_r).$$

Simulation: example

Simulation

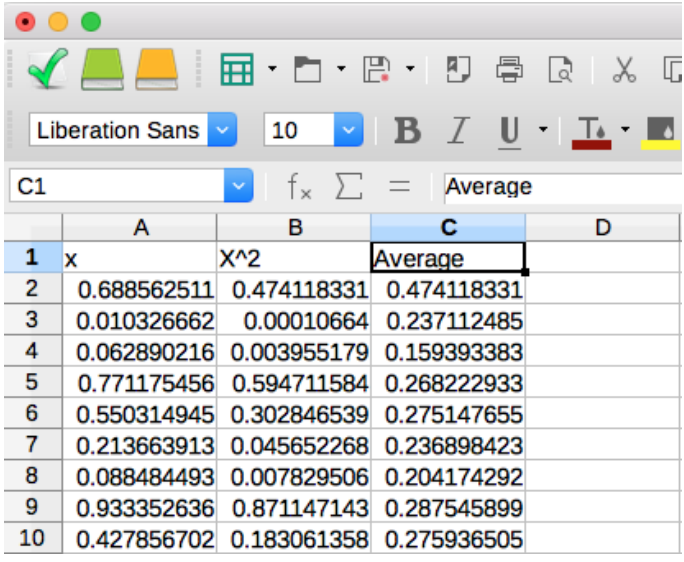
$$\int_a^b g(x)\, f_X(x) dx \approx \frac{1}{R} \sum_{r=1}^{R} g(x_r).$$

Example

$$\int_0^1 x^2 dx = \frac{1}{3}.$$

- $a = 0,\ b = 1.$
- $X \sim U[0,1],\ f_X(x) = 1$ if $0 \leq x \leq 1.$
- $g(x) = x^2.$

Calculate the integral in a spreadsheet

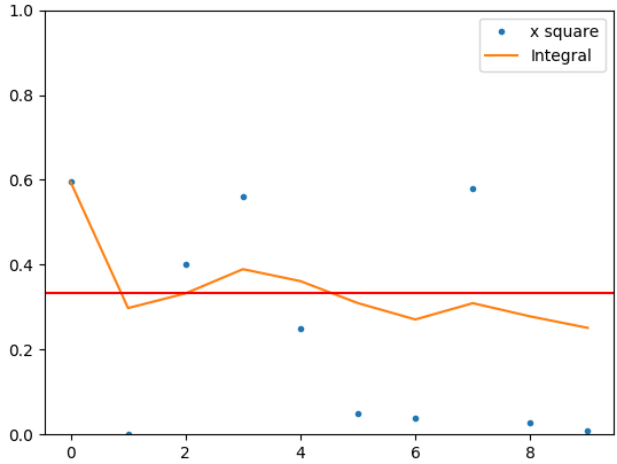


	A	B	C	D
1	x	X^2	Average	
2	0.688562511	0.474118331	0.474118331	
3	0.010326662	0.00010664	0.237112485	
4	0.062890216	0.003955179	0.159393383	
5	0.771175456	0.594711584	0.268222933	
6	0.550314945	0.302846539	0.275147655	
7	0.213663913	0.045652268	0.236898423	
8	0.088484493	0.007829506	0.204174292	
9	0.933352636	0.871147143	0.287545899	
10	0.427856702	0.183061358	0.275936505	

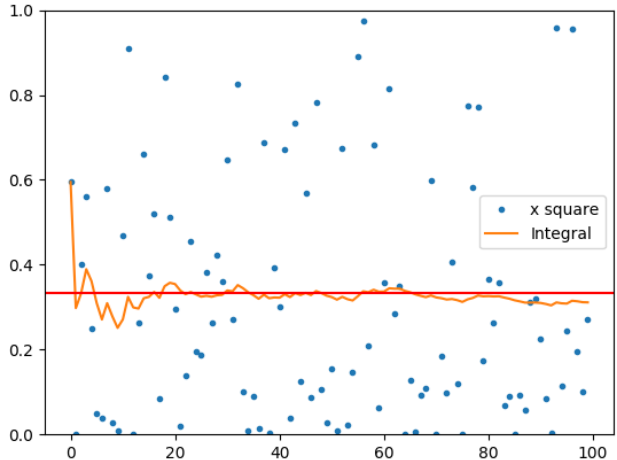
Calculate the integral in a spreadsheet

9980	0.646594694	0.418084699	0.334571277
9981	0.549007291	0.301409006	0.334567954
9982	0.079910337	0.006385662	0.334535074
9983	0.320106529	0.10246819	0.334511825
9984	0.15934821	0.025391852	0.33448086
9985	0.90318574	0.81574448	0.334529064
9986	0.050870209	0.002587778	0.33449582
9987	0.689559852	0.47549279	0.334509939
9988	0.344382493	0.118599302	0.33448832
9989	0.821720381	0.675224384	0.334522435
9990	0.834962081	0.697161677	0.334558739
9991	0.819882215	0.672206846	0.334592537
9992	0.766543325	0.58758867	0.33461786
9993	0.582796288	0.339651513	0.334618363
9994	0.750785318	0.563678594	0.334641285
9995	0.145241915	0.021095214	0.334609912
9996	0.146900665	0.021579805	0.334578593
9997	0.634764026	0.402925369	0.334585431
9998	0.565748431	0.320071287	0.334583979
9999	0.25163396	0.06331965	0.334556847
10000	0.688588606	0.474154268	0.334570808

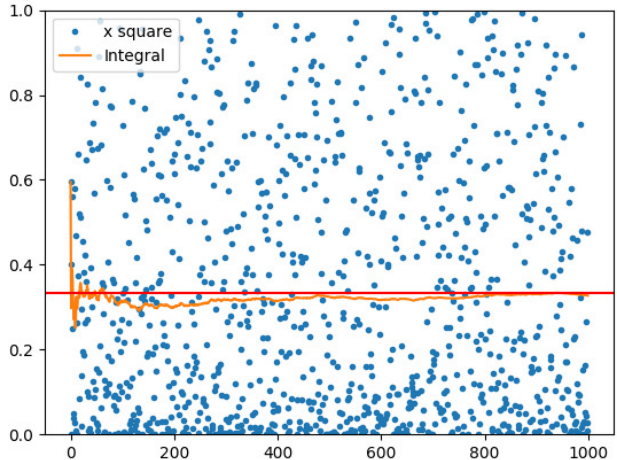
Calculate the integral with Python: $R = 10$



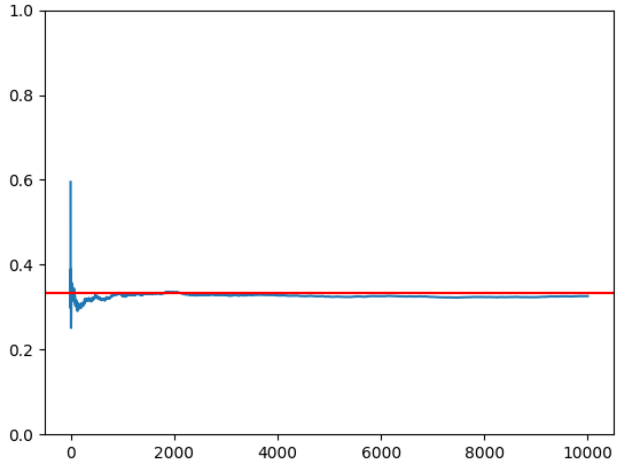
Calculate the integral with Python: $R = 100$



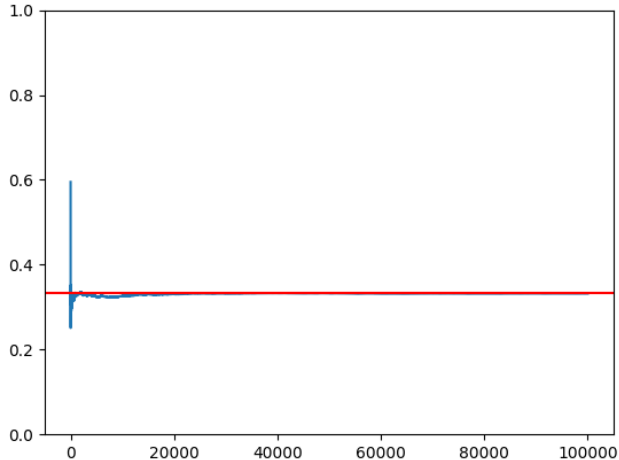
Calculate the integral with Python: $R = 1000$



Calculate the integral with Python: $R = 10000$



Calculate the integral with Python: $R = 100000$



Draws

Methodology

- Write the integral as the expectation of a function of a random variable.
- Draw from that random variable.
- Apply the function to the draws.
- Calculate the mean.

Draws

How to generate draws from a random variable?

Draws from uniform $U[0, 1]$

Matlab

Description

`X = rand` returns a single uniformly distributed random number in the interval (0,1).

Excel

Rand

The RAND function generates a random decimal number between 0 and 1.

1. Select cell A1.

2. Type RAND() and press Enter. The RAND function takes no arguments.

Draws from uniform $U[0, 1]$

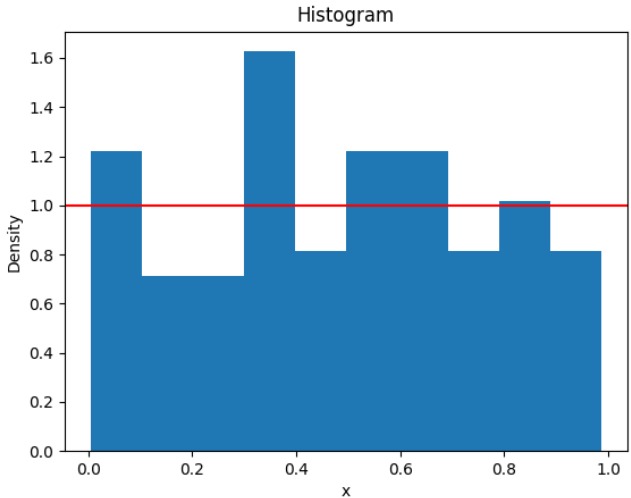
Python

`random.` **`random()`**

Return the next random floating point number in the range [0.0, 1.0).

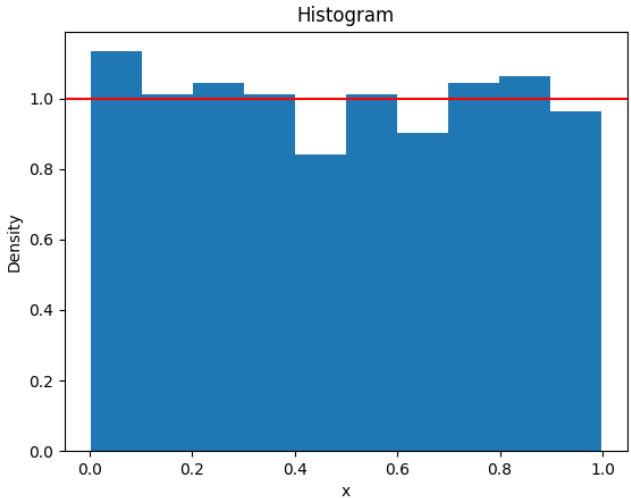
Draws: uniform distribution: $U[0,1], f_X(\varepsilon) = 1$

$R = 100$



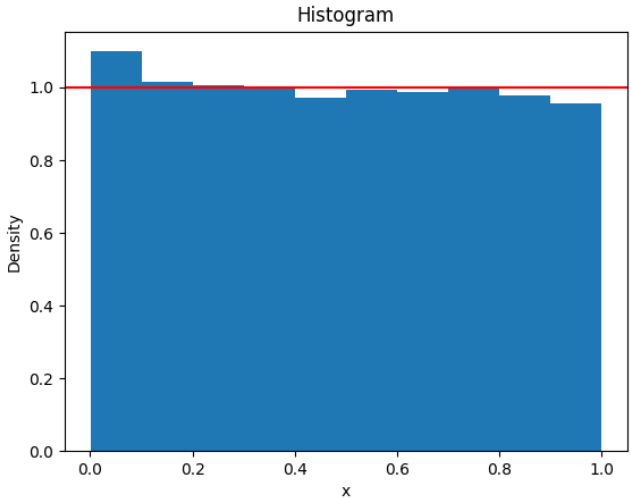
Draws: uniform distribution: $U[0,1], f_X(\varepsilon) = 1$

$R = 1000$



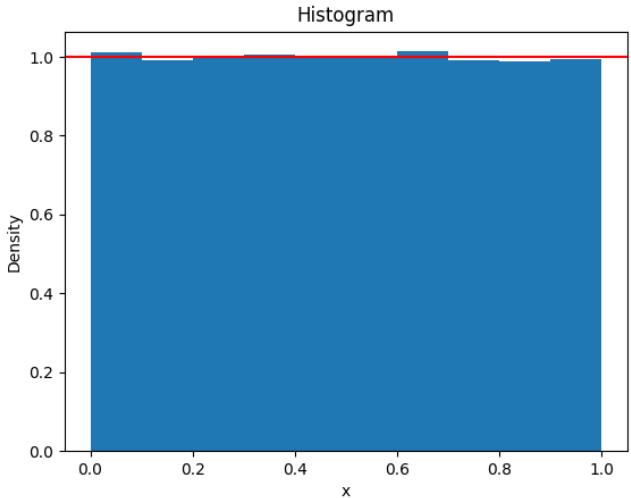
Draws: uniform distribution: $U[0,1], f_X(\varepsilon) = 1$

$R = 10000$



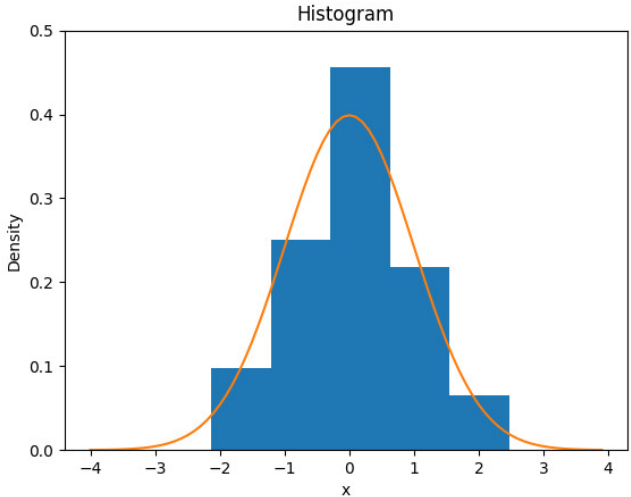
Draws: uniform distribution: $U[0,1], f_X(\varepsilon) = 1$

$R = 100000$



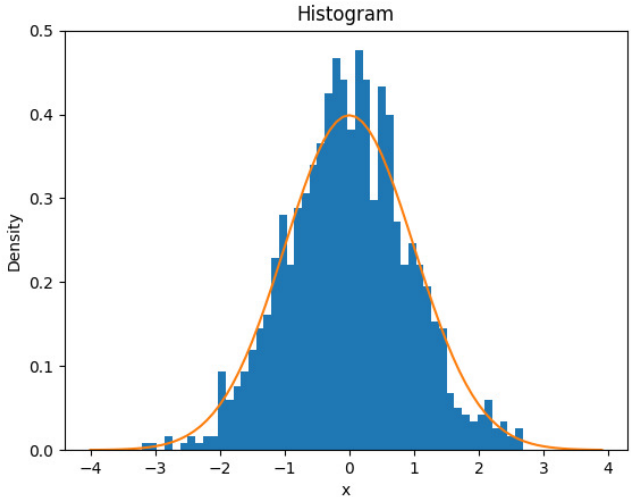
Draws from normal $N(0, 1)$

$R = 100$



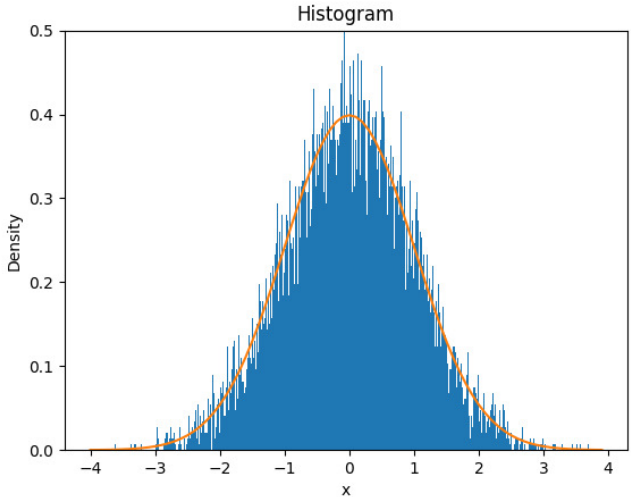
Draws from normal $N(0, 1)$

$R = 1000$



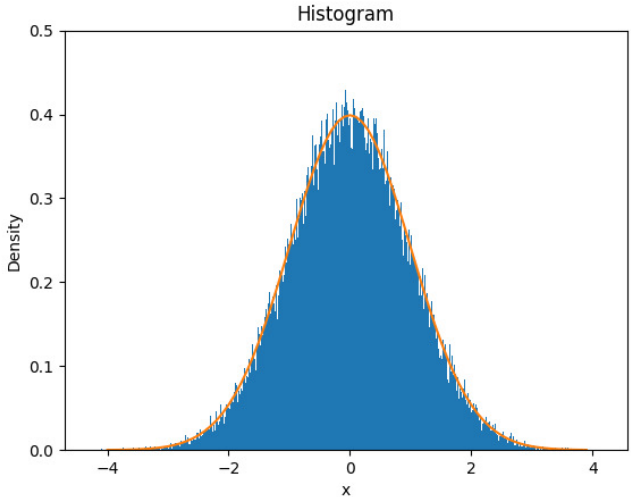
Draws from normal $N(0, 1)$

$R = 10000$



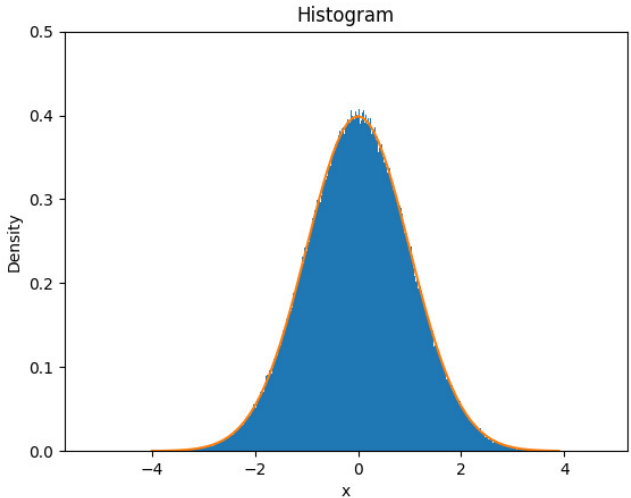
Draws from normal $N(0, 1)$

$R = 100000$



Draws from normal $N(0, 1)$

$R = 1000000$



Draws from normal $N(0, 1)$

Matlab

Description

`r = normrnd(mu,sigma)` generates a random number from the normal distribution with mean parameter `mu` and standard deviation parameter `sigma`.

Python

numpy.random.normal¶

`numpy.random.` **normal** (*loc=0.0, scale=1.0, size=None*)

Draw random samples from a normal (Gaussian) distribution.

Draws from normal $N(0, 1)$

Excel

The Excel NORMINV function calculates the inverse of the Cumulative Normal Distribution Function for a supplied value of x, and a supplied distribution mean & standard deviation.

The syntax of the function is:

NORMINV(probability, mean, standard_dev)

Where the function arguments are:

probability	-	The value at which you want to evaluate the inverse function.
mean	-	The arithmetic mean of the distribution.
standard_dev	-	The standard deviation of the distribution.

Biogeme

Error components

$$U_{in} = V_{in} + \xi_{in} + \nu_{in}.$$

Conditional choice model

$$\Pr(\text{car}|X_n, \xi_n) = \frac{e^{\beta X_{\text{car},n} + \xi_{\text{car},n}}}{e^{\beta X_{\text{car},n} + \xi_{\text{car},n}} + e^{\beta X_{\text{train},n} + \xi_{\text{train},n}} + e^{\beta X_{\text{Swissmetro},n} + \xi_{\text{Swissmetro},n}}}.$$

Choice model

$$P(\text{car}|X_n) = \mathrm{E}_{\xi_n}[\Pr(\text{car}|X_n, \xi_n)] = \int_{\xi} \Pr(\text{car}|X_n, \xi) f(\xi) d\xi.$$

Biogeme

Draws

```
XI = SCALE * bioDraws('XI', 'NORMAL')
```

Utility function

```
Vi = beta * x + XI
```

Logit model

```
prob = models.logit(V, av, CHOICE)
```

Mixtures

```
logprob = log(MonteCarlo(prob))
```

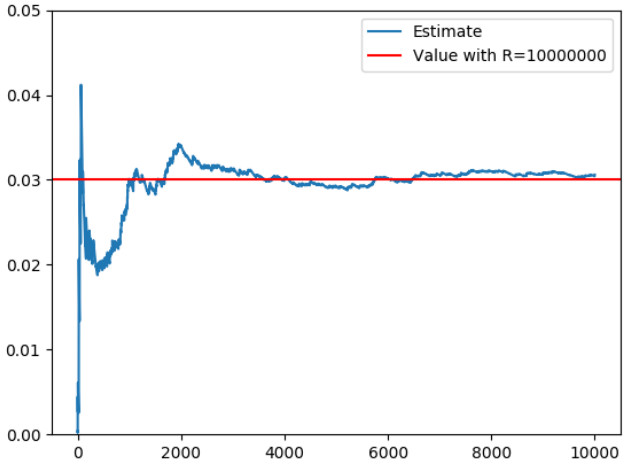
A simple example

$\xi_i \sim N(0,1)$, ν_i, ν_j i.i.d. $EV(0,1)$

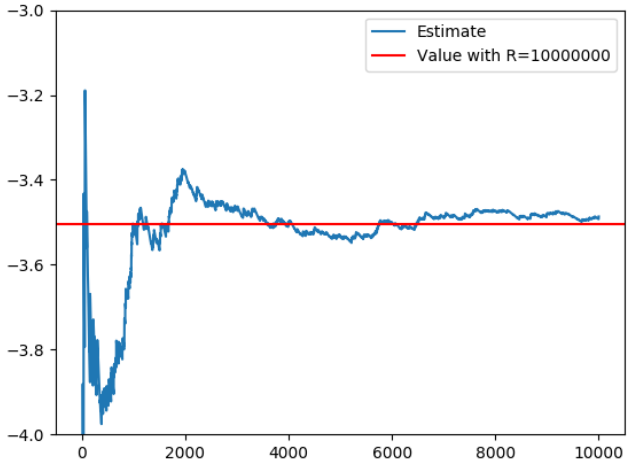
$$U_i = -10 + 5\xi_i + \nu_i,$$
$$U_j = \nu_j.$$

$$P(i) = \int_{\xi_i=-\infty}^{+\infty} P(i|\xi_i) \approx 3\%.$$

Probability



Log of probability



Parameters estimation

Maximum likelihood: intractable

$$\max_{\beta} \sum_{n} \ln P_n(\beta) = \sum_{n} \ln \int_{\xi} P_n(\beta|\xi) f(\xi) d\xi.$$

Maximum simulated likelihood

$$\max_{\beta} \sum_{n} \ln \widehat{P}_n(\beta) = \sum_{n} \ln \frac{1}{R} \sum_{r=1}^{R} P_n(\beta|\xi^r).$$

Bias

- $\widehat{P}_n(\beta)$ is an unbiased estimate of $P_n(\beta)$: $\mathrm{E}[\widehat{P}_n(\beta)] = P_n(\beta)$.
- $\ln \widehat{P}_n(\beta)$ is **not** an unbiased estimate of $\ln P_n(\beta)$: $\mathrm{E}[\ln \widehat{P}_n(\beta)] \neq \ln P_n(\beta)$.

What do we do with mixtures?

Outline

Relaxing the i.i.d. assumption

Logit model

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

ε_{in} are EV, i.i.d. across i and n.

Relaxing the assumptions

- Relaxing independence across i.
- Relaxing the identical distribution assumption.

Example

Residential telephone

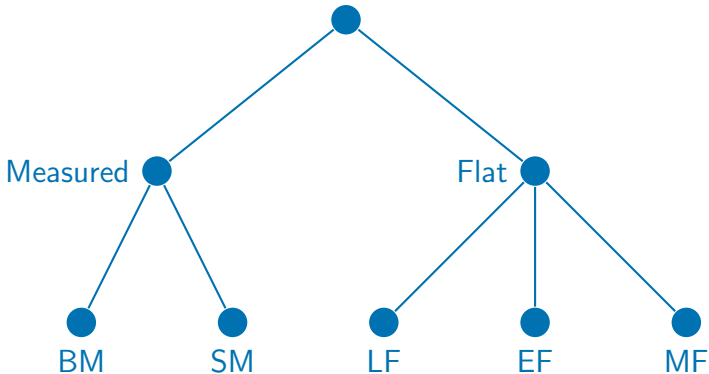
	Ct. BM	Ct. SM	Ct. LF	Ct. EF	β_C
BM	1	0	0	0	ln(cost(BM))
SM	0	1	0	0	ln(cost(SM))
LF	0	0	1	0	ln(cost(LF))
EF	0	0	0	1	ln(cost(EF))
MF	0	0	0	0	ln(cost(MF))

Example: logit

	Value	Rob. Std err	Rob. t-test	Rob. p-value
Cte BM	-2.46	0.301	-8.17	2.22e-16
Cte SM	-1.74	0.267	-6.51	7.28e-11
Cte LF	-0.535	0.199	-2.69	0.00709
Cte EF	-0.737	0.713	-1.03	0.301
β_{cost}	-2.03	0.212	-9.55	0.0

Number of estimated parameters	5
Sample size	434
Null log likelihood	-560.2496
Final log likelihood	-477.5569
Akaike Information Criterion	965.1
Bayesian Information Criterion	985.5

Capturing correlations: nesting



Capturing correlations: nesting with MEV

	Value	Rob. Std err	Rob. t-test	Rob. p-value
Cte BM	-1.78	0.313	-5.69	1.25e-08
Cte SM	-1.41	0.238	-5.9	3.71e-09
Cte LF	-0.512	0.124	-4.13	3.64e-05
Cte EF	-0.558	0.388	-1.44	0.15
β_{cost}	-1.49	0.243	-6.12	9.09e-10
μ_{meas}	2.06	0.573	3.6	0.000317
μ_{flat}	2.29	0.764	3.0	0.00268

Number of estimated parameters	7
Sample size	434
Init log likelihood	-560.2496
Final log likelihood	-473.2193
Akaike Information Criterion	960.4387
Bayesian Information Criterion	988.95

Capturing correlations: nesting with mixtures

Utility

$$
\begin{aligned}
U_{\text{BM}} &= c_{\text{BM}} + \beta_c \ln(\text{cost}_{BM}) + \sigma_M \xi_M + \nu_{\text{BM}}, \\
U_{\text{SM}} &= c_{\text{SM}} + \beta_c \ln(\text{cost}_{SM}) + \sigma_M \xi_M + \nu_{\text{SM}}, \\
U_{\text{LF}} &= c_{\text{LF}} + \beta_c \ln(\text{cost}_{LF}) + \sigma_F \xi_F + \nu_{\text{LF}}, \\
U_{\text{EF}} &= c_{\text{EF}} + \beta_c \ln(\text{cost}_{EF}) + \sigma_F \xi_F + \nu_{\text{EF}}, \\
U_{\text{MF}} &= \beta_c \ln(\text{cost}_{MF}) + \sigma_F \xi_F + \nu_{\text{MF}},
\end{aligned}
$$

with $\xi_F, \xi_M \sim N(0,1)$, $\nu_{\text{MF}} \sim EV(0,1)$.

Capturing correlations: nesting with mixtures

Logit kernel

$$\Pr(\text{BM}|X,\xi_F,\xi_M) = \frac{e^{\mu(c_{\text{BM}}+\beta_c \ln(\text{cost}_{BM})+\sigma_M\xi_M)}}{\sum_j e^{\cdots}}.$$

Mixture of logit

$$\begin{aligned} P(\text{BM}|X) &= \int_{\xi_F}\int_{\xi_M} \Pr(\text{BM}|X,\xi_F,\xi_M) f(\xi_M) f(\xi_F) d\xi_M d\xi_F \\ &\approx \frac{1}{R}\sum_{r=1}^{R} \Pr(\text{BM}|X,\xi_F^r,\xi_M^r). \end{aligned}$$

Capturing correlations: nesting with error components

Residential telephone

	Ct. BM	Ct. SM	Ct. LF	Ct. EF	β_C	σ_M	σ_F
BM	1	0	0	0	ln(cost(BM))	η_M	0
SM	0	1	0	0	ln(cost(SM))	η_M	0
LF	0	0	1	0	ln(cost(LF))	0	η_F
EF	0	0	0	1	ln(cost(EF))	0	η_F
MF	0	0	0	0	ln(cost(MF))	0	η_F

Estimation: 5000 draws. Linear-in-parameter specification: $\mu = 1$.

Capturing correlations: nesting with error components

	Value	Rob. Std err	Rob. t-test	Rob. p-value
Cte BM	-3.82	0.661	-5.78	7.55e-09
Cte SM	-3.02	0.616	-4.91	9.08e-07
Cte LF	-1.1	0.307	-3.58	0.000344
Cte EF	-1.21	0.877	-1.37	0.169
$\beta_{\texttt{cost}}$	-3.27	0.535	-6.12	9.57e-10
σ_M	1.61	0.611	2.63	0.00856
σ_F	2.64	0.837	3.16	0.00158

Number of estimated parameters	7
Sample size	434
Init log likelihood	-560.2496
Final log likelihood	-472.9221
Akaike Information Criterion	959.8442
Bayesian Information Criterion	988.3555

Capturing correlations

Identification issues [Walker, 2001]

- If there are two nests, only one σ is identified.
- If there are more than two nests, all σ's are identified.
- The reason is that only difference in utility matters.

Possible normalizations

- $\sigma_M = 0$,
- $\sigma_F = 0$,
- $\sigma_M = \sigma_F$.

Results with 5000 draws

	logit	nested	mixture	mixture $\sigma_M = 0$	mixture $\sigma_F = 0$	mixture $\sigma_M = \sigma_F$
K	5	7	7	6	6	6
$\mathcal{L}(\widehat{\beta})$	-477.56	-473.22	-472.92	-472.91	-472.94	-473.00
Cte BM	-2.46	-1.78	-3.82	-3.81	-3.81	-3.8
Cte SM	-1.74	-1.41	-3.02	-3.01	-3.01	-3.0
Cte LF	-0.535	-0.512	-1.1	-1.1	-1.1	-1.09
Cte EF	-0.737	-0.558	-1.21	-1.19	-1.2	-1.19
β_{cost}	-2.03	-1.49	-3.27	-3.26	-3.26	-3.25
μ_M		2.06				
μ_F		2.29				
σ_M			1.61		3.06	2.16
σ_F			2.64	3.06		2.16
$\sqrt{\sigma_M^2 + \sigma_F^2}$			3.09	3.06	3.06	3.05

Results with 5000 draws

Comments

- Normalization can be performed in several ways
  - $\sigma_F = 0$,
  - $\sigma_M = 0$,
  - $\sigma_F = \sigma_M$.
- The final log likelihood should be the same.
- But... estimation relies on simulation.
- Only an approximation of the log likelihood is available.
- We re-estimate the models with 1 000 000 draws.

Results with 1 000 000 draws

	logit	nested	mixture	mixture $\sigma_M = 0$	mixture $\sigma_F = 0$	mixture $\sigma_M = \sigma_F$
K	5	7	7	6	6	6
$\mathcal{L}(\widehat{\beta})$	-477.56	-473.22	-473.02	-473.02	-473.02	-473.02
Cte BM	-2.46	-1.78	-3.81	-3.8	-3.8	-3.8
Cte SM	-1.74	-1.41	-3.01	-3.0	-3.01	-3.0
Cte LF	-0.535	-0.512	-1.09	-1.09	-1.09	-1.09
Cte EF	-0.737	-0.558	-1.19	-1.19	-1.19	-1.19
β_{cost}	-2.03	-1.49	-3.26	-3.25	-3.25	-3.25
μ_M		2.06				
μ_F		2.29				
σ_M			2.88		3.05	2.15
σ_F			1.04	3.04		2.15
$\sqrt{\sigma_M^2 + \sigma_F^2}$			3.06	3.04	3.05	3.04

Capturing correlations

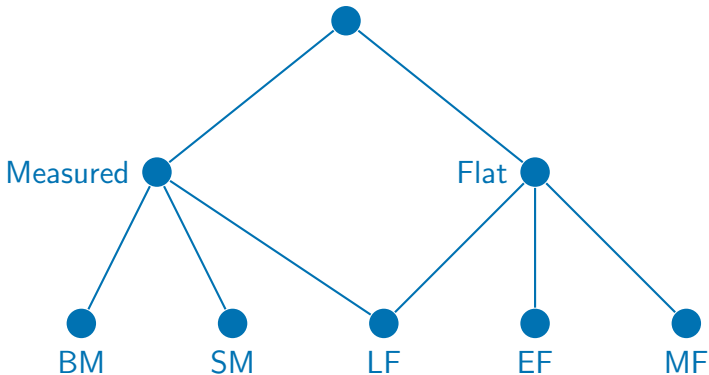
Cost coefficient very different across models

- Nested logit: -1.43.
- Mixture model: -3.26.
- Linear-in-parameter specification.
- Values **cannot** be compared.
- $\mu_{NL}\beta_c = -1.43$.
- $\mu_M\beta_c = -3.26$.
- Re-estimate the models with a "moneymetric" specification: $\beta_c = -1$.
- Or, divide all the coefficients by β_{cost}.

Results with 5000 draws: moneymetric specification

	logit	nested	mixture	mixture $\sigma_M = 0$	mixture $\sigma_F = 0$	mixture $\sigma_M = \sigma_F$
K	5	7	7	6	6	6
$\mathcal{L}(\widehat{\beta})$	-477.56	-473.22	-473.04	-473.06	-473.07	-473.09
Cte BM	-1.21	-1.2	-1.17	-1.17	-1.17	-1.17
Cte SM	-0.857	-0.943	-0.923	-0.925	-0.926	-0.925
Cte LF	-0.264	-0.344	-0.336	-0.336	-0.336	-0.336
Cte EF	-0.364	-0.375	-0.367	-0.367	-0.366	-0.365
μ	2.03	1.49	3.25	3.24	3.25	3.25
μ_M		2.06				
μ_F		2.29				
σ_M			0.648		0.94	0.664
σ_F			0.681	0.932		0.664
$\sigma_M^2 + \sigma_F$			0.94	0.932	0.94	0.939

Capturing correlations: cross-nesting



Capturing correlations: cross-nesting with mixtures

Utility

$$
\begin{aligned}
U_{\mathrm{BM}} &= c_{\mathrm{BM}} + \beta_c \ln(\mathrm{cost}_{BM}) + \sigma_M \xi_M + \nu_{\mathrm{BM}}, \\
U_{\mathrm{SM}} &= c_{\mathrm{SM}} + \beta_c \ln(\mathrm{cost}_{SM}) + \sigma_M \xi_M + \nu_{\mathrm{SM}}, \\
U_{\mathrm{LF}} &= c_{\mathrm{LF}} + \beta_c \ln(\mathrm{cost}_{LF}) + \sigma_M \xi_M + \sigma_F \xi_F + \nu_{\mathrm{LF}}, \\
U_{\mathrm{EF}} &= c_{\mathrm{EF}} + \beta_c \ln(\mathrm{cost}_{EF}) + \sigma_F \xi_F + \nu_{\mathrm{EF}}, \\
U_{\mathrm{MF}} &= \beta_c \ln(\mathrm{cost}_{MF}) + \sigma_F \xi_F + \nu_{\mathrm{MF}},
\end{aligned}
$$

with $\xi_F, \xi_M \sim N(0,1)$, $\nu_{\mathrm{MF}} \sim EV(0,1)$.

Capturing correlations: cross-nesting with mixtures

Logit kernel

$$\Pr(\text{LF}|X,\xi_F,\xi_M) = \frac{e^{\mu(c_{\text{BM}}+\beta_c \ln(\text{cost}_{BM})+\sigma_M\xi_M+\sigma_F\xi_F)}}{\sum_j e^{\cdots}}.$$

Mixture of logit

$$\begin{aligned} P(\text{LF}|X) &= \int_{\xi_F}\int_{\xi_M} \Pr(\text{LF}|X,\xi_F,\xi_M) f(\xi_M) f(\xi_F) d\xi_M d\xi_F \\ &\approx \frac{1}{R}\sum_{r=1}^{R} \Pr(\text{LF}|X,\xi_F^r,\xi_M^r). \end{aligned}$$

What do we do with mixtures?

Outline

Relaxing the i.i.d. assumption

Logit model

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

ε_{in} are EV, i.i.d. across i and n.

Relaxing the assumptions

- Relaxing independence across i.
- Relaxing the identical distribution assumption.

Alternative specific variance

Logit: i.i.d. error terms

- In particular, they have the same variance

$$U_{in} = \beta^T x_{in} + \text{ASC}_i + \nu_{in}.$$

- ν_{in} i.i.d. $EV(0, \mu) \Rightarrow \text{Var}(\nu_{in}) = \pi^2/6\mu^2$.

Relax the identical distribution assumption

$$U_{in} = \beta^T x_{in} + \text{ASC}_i + \sigma_i \xi_i + \nu_{in},$$

where $\xi_i \sim N(0, 1)$.

Alternative specific variance

Error term

$$\varepsilon_{in} = \sigma_i \, \xi_i + \nu_{in}.$$

Unknown distribution.

Variance

Assume ξ_i and ν_{in} independent:

$$\mathrm{Var}(\sigma_i \, \xi_i + \nu_{in}) = \sigma_i^2 + \frac{\pi^2}{6\mu^2}.$$

Identification issue: process

Examine the variance-covariance matrix

1. Specify the model of interest.
2. Take the differences in utilities.
3. Apply the order condition: necessary condition.
4. Apply the rank condition: sufficient condition.
5. Apply the equality condition: verify equivalence.

Identification issue: example

Model

$$\begin{array}{rcccccccl} U_1 & = & \beta x_1 & +\sigma_1\xi_1 & & & & & +\nu_1, \\ U_2 & = & \beta x_2 & & +\sigma_2\xi_2 & & & & +\nu_2, \\ U_3 & = & \beta x_3 & & & +\sigma_3\xi_3 & & & +\nu_3, \\ U_4 & = & \beta x_4 & & & & +\sigma_4\xi_4 & & +\nu_4, \end{array}$$

where $\xi_i \sim N(0,1)$, $\nu_i \sim EV(0,\mu)$.

Covariance matrix ($\gamma = \pi^2/6$)

$$\mathrm{Cov}(U) = \begin{pmatrix} \sigma_1^2 + \gamma/\mu^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 + \gamma/\mu^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 + \gamma/\mu^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 + \gamma/\mu^2 \end{pmatrix}.$$

Identification issue: differences

Utility differences

$$\begin{pmatrix} U_1 - U_4 \\ U_2 - U_4 \\ U_3 - U_4 \end{pmatrix} = \Delta_4 \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix},$$

where

$$\Delta_4 = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

Identification issue: differences

Utility differences

$$\begin{aligned} U_1 - U_4 &= \beta(x_1 - x_4) + (\sigma_1\xi_1 - \sigma_4\xi_4) + (\nu_1 - \nu_4), \\ U_2 - U_4 &= \beta(x_2 - x_4) + (\sigma_2\xi_2 - \sigma_4\xi_4) + (\nu_2 - \nu_4), \\ U_3 - U_4 &= \beta(x_3 - x_4) + (\sigma_3\xi_3 - \sigma_4\xi_4) + (\nu_3 - \nu_4). \end{aligned}$$

Covariance of utility differences

$\mathrm{Cov}(\Delta U) =$

$$\begin{pmatrix} \sigma_1^2 + \sigma_4^2 + 2\gamma/\mu^2 & \sigma_4^2 + \gamma/\mu^2 & \sigma_4^2 + \gamma/\mu^2 \\ \sigma_4^2 + \gamma/\mu^2 & \sigma_2^2 + \sigma_4^2 + 2\gamma/\mu^2 & \sigma_4^2 + \gamma/\mu^2 \\ \sigma_4^2 + \gamma/\mu^2 & \sigma_4^2 + \gamma/\mu^2 & \sigma_3^2 + \sigma_4^2 + 2\gamma/\mu^2 \end{pmatrix}.$$

Identification issue: order condition

Upper bound

- S is the number of estimable parameters.
- J is the number of alternatives.

$$S \leq \frac{J(J-1)}{2} - 1.$$

- It represents the number of entries in the lower part of the (symmetric) var-cov matrix,
- minus 1 for the scale.
- $J = 4$ implies $S \leq 5$.

Identification issue: rank condition

Idea

- Number of estimable parameters =
- number of linearly independent equations
- -1 for the scale

$\mathrm{Cov}(\Delta U) =$

$$\begin{pmatrix} \sigma_1^2 + \sigma_4^2 + 2\gamma/\mu^2 & & \\ \sigma_4^2 + \gamma/\mu^2 & \sigma_2^2 + \sigma_4^2 + 2\gamma/\mu^2 & \\ \sigma_4^2 + \gamma/\mu^2 & \sigma_4^2 + \gamma/\mu^2 & \sigma_3^2 + \sigma_4^2 + 2\gamma/\mu^2 \end{pmatrix}$$

dependent

scale

Identification issue: rank condition

Three parameters out of five can be estimated

Formally...

1. Identify unique elements of $\mathrm{Cov}(\Delta U)$.
2. Compute the Jacobian wrt σ_1^2, σ_2^2, σ_3^2, σ_4^2, γ/μ^2.
3. Compute the rank:

$$\begin{pmatrix} \sigma_1^2 + \sigma_4^2 + 2\gamma/\mu^2 \\ \sigma_2^2 + \sigma_4^2 + 2\gamma/\mu^2 \\ \sigma_3^2 + \sigma_4^2 + 2\gamma/\mu^2 \\ \sigma_4^2 + \gamma/\mu^2 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$S =$ Rank - $1 = 3$.

Identification issue: equality condition

Normalization

- We know how many parameters can be identified.
- There are infinitely many normalizations.
- The normalized model is equivalent to the original one.
- Obvious normalizations, like constraining extra-parameters to 0 or another constant, may not be valid.

Identification issue: equality condition

Error components

$$\begin{array}{rcccccl} U_n & = & \beta^T x_n & + & L_n \xi_n & + & \nu_n, \\ \mathrm{Cov}(U_n) & = & & & L_n L_n^T & + & (\gamma/\mu^2) I, \\ \mathrm{Cov}(\Delta_j U_n) & = & & & \Delta_j L_n L_n^T \Delta_j^T & + & (\gamma/\mu^2) \Delta_j \Delta_j^T. \end{array}$$

Notations

$$\Delta_2 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

$$\begin{array}{rrcccl} \mathrm{Cov}(U_n) = & \Omega_n & = & \Sigma_n & + & \Gamma_n \\ & \Omega_n^{\mathrm{norm}} & = & \Sigma_n^{\mathrm{norm}} & + & \Gamma_n^{\mathrm{norm}} \\ \mathrm{Cov}(\Delta_j U_n) = & \Omega_{n\Delta} & = & \Sigma_{n\Delta} & + & \Gamma_{n\Delta} \\ & \Omega_{n\Delta}^{\mathrm{norm}} & = & \Sigma_{n\Delta}^{\mathrm{norm}} & + & \Gamma_{n\Delta}^{\mathrm{norm}} \end{array}$$

Identification issue: equality condition

The following conditions must hold

- Covariance matrices must be equal

$$\Omega_{n\Delta} = \Omega_{n\Delta}^{\text{norm}}.$$

- Σ_n^{norm} must be positive semi-definite.

Identification issue: equality condition

Example with 3 alternatives (index n dropped)

$$\begin{array}{rcllll} U_1 &=& \beta x_1 & +\sigma_1\xi_1 & & +\nu_1, \\ U_2 &=& \beta x_2 & +\sigma_2\xi_2 & & +\nu_2, \\ U_3 &=& \beta x_3 & & +\sigma_3\xi_3 & +\nu_3. \end{array}$$

$$\mathrm{Cov}(\Delta_3 U) = \Omega_\Delta = \begin{pmatrix} \sigma_1^2 + \sigma_3^2 + 2\gamma/\mu^2 & \\ \sigma_3^2 + \gamma/\mu^2 & \sigma_2^2 + \sigma_3^2 + 2\gamma/\mu^2 \end{pmatrix}.$$

- Parameters: $\{\sigma_1, \sigma_2, \sigma_3, \mu\}$.
- Rank condition: $S = 2$.
- μ is used for the scale.

Identification issue: equality condition

Change of variables

- Denote $\nu_i = \sigma_i^2 \mu^2$ (scaled parameters).
- Normalization condition: $\nu_3 = K$.

$$\Omega_\Delta = \begin{pmatrix} (\nu_1 + \nu_3 + 2\gamma)/\mu^2 & \\ (\nu_3 + \gamma)/\mu^2 & (\nu_2 + \nu_3 + 2\gamma)/\mu^2 \end{pmatrix},$$

$$\Omega_\Delta^{\text{norm}} = \begin{pmatrix} (\nu_1^N + K + 2\gamma)/\mu_N^2 & \\ (K + \gamma)/\mu_N^2 & (\nu_2^N + K + 2\gamma)/\mu_N^2 \end{pmatrix},$$

where index N stands for "normalized"

Identification issue: equality condition

First equality condition: $\Omega_\Delta = \Omega_\Delta^{\text{norm}}$

$$\begin{aligned}(\nu_3+\gamma)/\mu^2 &= (K+\gamma)/\mu_N^2,\\ (\nu_1+\nu_3+2\gamma)/\mu^2 &= (\nu_1^N+K+2\gamma)/\mu_N^2,\\ (\nu_2+\nu_3+2\gamma)/\mu^2 &= (\nu_2^N+K+2\gamma)/\mu_N^2,\end{aligned}$$

that is, writing the normalized parameters as functions of others,

$$\begin{aligned}\mu_N^2 &= \mu^2(K+\gamma)/(\nu_3+\gamma),\\ \nu_1^N &= (K+\gamma)(\nu_1+\nu_3+2\gamma)/(\nu_3+\gamma)-K-2\gamma,\\ \nu_2^N &= (K+\gamma)(\nu_2+\nu_3+2\gamma)/(\nu_3+\gamma)-K-2\gamma.\end{aligned}$$

Identification issue: equality condition

Second equality condition

$$\Sigma^{\text{norm}} = \frac{1}{\mu_N^2} \begin{pmatrix} \nu_1^N & 0 & 0 \\ 0 & \nu_2^N & 0 \\ 0 & 0 & K \end{pmatrix}$$

must be positive semi-definite, that is

$$\mu_N > 0,\ \nu_1^N \geq 0,\ \nu_2^N \geq 0,\ K \geq 0.$$

Putting everything together, we obtain

$$K \geq \frac{(\nu_3 - \nu_i)\gamma}{\nu_i + \gamma},\ i = 1, 2$$

Identification issue: equality condition

Condition to be verified for the normalization to be valid

$$K \geq \frac{(\nu_3 - \nu_i)\gamma}{\nu_i + \gamma}, \ i = 1, 2$$

- If $\nu_3 \leq \nu_i$, $i = 1, 2$, then the rhs is negative, and any $K \geq 0$ would do. Typically, $K = 0$.
- If not, K must be chosen large enough
- In practice, always select the alternative with minimum variance.

Alternative specific variance

Identification issue

- Not all σs are identified.
- One of them must be constrained to zero.
- Not necessarily the one associated with the ASC constrained to zero.
- In theory, the smallest σ^2 must be constrained to zero.
- In practice, we don't know a priori which one it is.
- Solution:
  1. Estimate a model with a full set of σs.
  2. Identify the smallest one and constrain it to zero.

Alternative specific variance

Swissmetro

Param.	Train	Swissmetro	Car
Cte train	1	0	0
Cte car	0	0	1
β_{time}	trav. time/100	trav. time/100	trav. time/100
β_{cost}	trav. cost/100	trav. cost/100	trav. cost/100
β_{headway}	headway/1000	headway/1000	0

$+$ alternative specific variance.

Comparison (using 20000 draws)

	Logit		ASV		ASV norm.	
$\mathcal{L}$	-5315.39		-5240.354		-5239.911	
	Estim.	Scaled	Estim.	Scaled	Estim.	Scaled
Cte car	-0.262	0.241	-0.666	0.375	-0.664	0.374
Cte train	-0.451	0.416	-0.914	0.515	-0.91	0.513
β_{cost}	-1.08	1.0	-1.78	1.0	-1.78	1.0
β_{headway}	-5.35	4.94	-7.8	4.39	-7.82	4.4
β_{time}	-1.28	1.18	-1.71	0.96	-1.71	0.962
σ_{car}			0.00355			
σ_{train}			0.0109		-0.0133	
σ_{SM}			-3.25		-3.24	

Relax the i.i.d. assumption

i.i.d. assumption

- ✓ Same η for all alternatives i: relaxed.
- ✓ Same η for all observations n: relaxed.
- ▶ Same μ for all alternatives i: relaxed in this lecture.
- ✓ Same μ for all observations n: relaxed.
- ✓ Independence across alternatives i: relaxed.
- ▶ Independence across observations n: relaxed in the lecture on panel data.

What do we do with mixtures?

Outline

Taste heterogeneity

Motivation

- So far, we have investigated model specifications that included error components in the utility functions.
- We now investigate how to include random parameters to model taste heterogeneity.

Random parameters

$$\begin{aligned} U_i &= \beta_{nt} T_i + \beta_c C_i + \nu_i, \\ U_j &= \beta_{nt} T_j + \beta_c C_j + \nu_j. \end{aligned}$$

Let $\beta_{nt} \sim N(\bar{\beta}_t, \sigma_t^2)$, or, equivalently,

$$\beta_{nt} = \bar{\beta}_t + \sigma_t \xi_n, \text{ with } \xi_n \sim N(0,1).$$

$$\begin{aligned} U_i &= \bar{\beta}_t T_i + \sigma_t \xi_n T_i + \beta_c C_i + \nu_i, \\ U_j &= \bar{\beta}_t T_j + \sigma_t \xi_n T_j + \beta_c C_j + \nu_j. \end{aligned}$$

Random parameters

$$\begin{aligned} U_i &= \bar{\beta}_t T_i + \sigma_t \xi_n T_i + \beta_c C_i + \nu_i, \\ U_j &= \bar{\beta}_t T_j + \sigma_t \xi_n T_j + \beta_c C_j + \nu_j. \end{aligned}$$

If ν_i and ν_j are i.i.d. EV and ξ_n is given, we have

$$\Pr(i|\xi_n) = \frac{e^{\bar{\beta}_t T_i + \sigma_t \xi_n T_i + \beta_c C_i}}{e^{\bar{\beta}_t T_i + \sigma_t \xi_n T_i + \beta_c C_i} + e^{\bar{\beta}_t T_j + \sigma_t \xi_n T_j + \beta_c C_j}}, \text{ and}$$

$$P(i) = \int_{\xi} P(i|\xi) f(\xi) d\xi.$$

Random parameters

Swissmetro

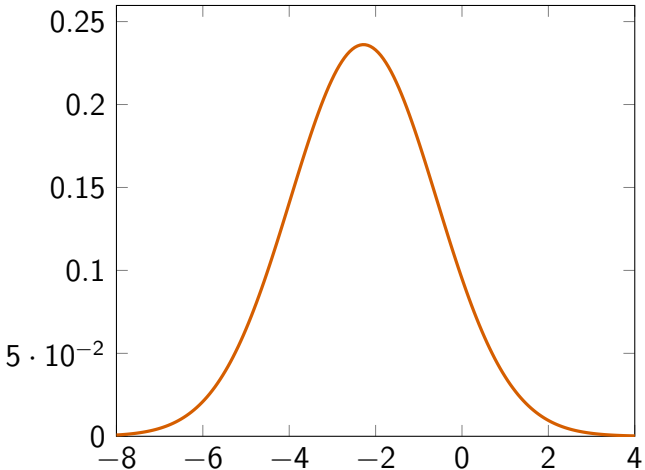
Param.	Train	Swissmetro	Car
Cte train	1	0	0
Cte car	0	0	1
$\beta_{n,\text{time}}$	trav. time/100	trav. time/100	trav. time/100
β_{cost}	trav. cost/100	trav. cost/100	trav. cost/100
β_{headway}	headway/1000	headway/1000	0

$+ \beta_{n,\text{time}}$ randomly distributed across the population, normal distribution (20000 draws).

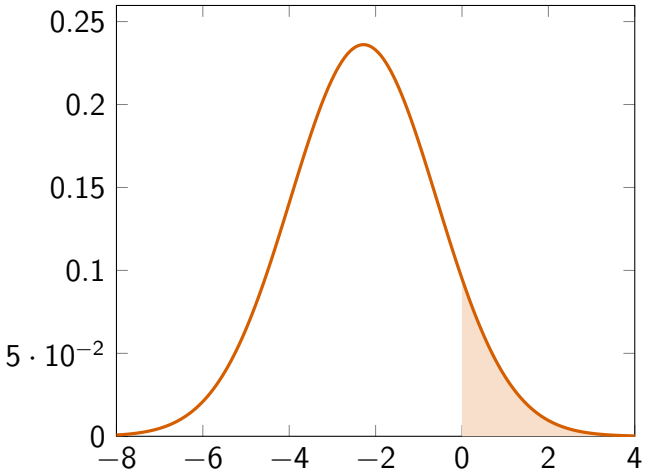
Random parameters

	Logit	Rand. coeff.
Nbr param.	5	6
Sample size	6768	6768
Final $\mathcal{L}$	-5315.39	-5196.84
AIC	10640.77	10405.69
BIC	10674.87	10446.61
Cte car	-0.262	0.0126
Cte train	-0.451	-0.104
$\beta_{\texttt{cost}}$	-1.08	-1.29
$\beta_{\texttt{headway}}$	-5.35	-6.37
$\beta_{\texttt{time}}$	-1.28	-2.28
$\sigma_{\texttt{time}}$		-1.69

Random parameters: time coefficient



Random parameters: time coefficient



Random parameters

	Logit	Rand. coeff.
Nbr param.	5	6
Sample size	6768	6768
Final $\mathcal{L}$	-5315.39	-5196.84
AIC	10640.77	10405.69
BIC	10674.87	10446.61
Cte car	-0.262	0.0126
Cte train	-0.451	-0.104
$\beta_{\texttt{cost}}$	-1.08	-1.29
$\beta_{\texttt{headway}}$	-5.35	-6.37
$\beta_{\texttt{time}}$	-1.28	-2.28
$\sigma_{\texttt{time}}$		-1.69
$\Pr(\beta_{\texttt{time}} \geq 0)$	0.0%	8.91%

Log normal distribution

If $\beta_t \sim N(\bar{\beta}_t, \sigma_t^2)$, then

$$e^{\beta_t} \sim LN(\bar{\beta}_t, \sigma_t^2),$$

Mean

$$\mathrm{E}[e^{\beta_t}] = e^{\mu+(\sigma^2/2)}.$$

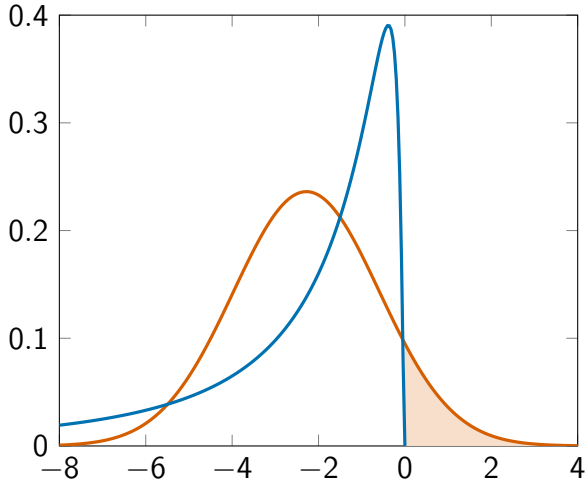
Variance

$$\mathrm{Var}[e^{\beta_t}] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

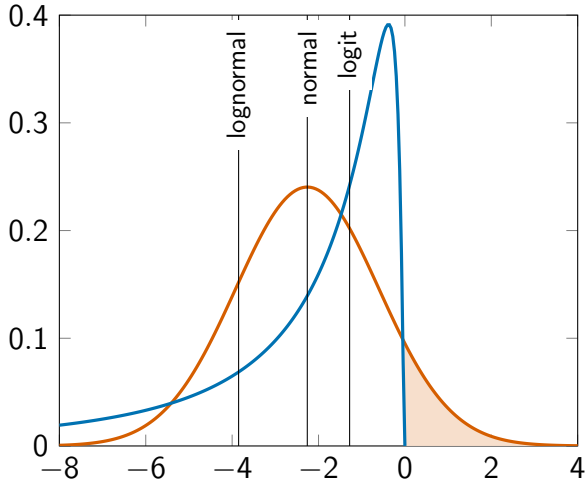
Random parameters

	Logit	RC	RC (log)	
Nbr param.	5	6	6	
Sample size	6768	6768	6768	
Final $\mathcal{L}$	-5315.39	-5196.84	-5215.01	
AIC	10640.77	10405.69	10442.01	
BIC	10674.87	10446.61	10482.93	
Cte car	-0.262	0.0126	0.0553	
Cte train	-0.451	-0.104	-0.0666	
$\beta_{\texttt{cost}}$	-1.08	-1.29	-1.39	
$\beta_{\texttt{headway}}$	-5.35	-6.37	-5.96	
$\beta_{\texttt{time}}$	-1.28	-2.28	0.575	-3.85
$\sigma_{\texttt{time}}$		-1.69	1.24	7.42
$\Pr(\beta_{\texttt{time}} \geq 0)$	0.0%	8.91%	0.0%	

Random parameters: time coefficient



Random parameters: time coefficient



Outline

Latent classes

Swissmetro

- Stated preferences data.
- May be some respondents did not play the game correctly.
- Assumption: some have ignored the time variable to make the choice.
- We should impose $\beta_{\text{time}} = 0$ for those individuals.
- But who are they?

Classes

- Class 1: respondents who considered travel time.
- Class 2: respondents who did not.

Latent classes

Class 1: respondents who considered travel time

Param.	Train	Swissmetro	Car
Cte train	1	0	0
Cte car	0	0	1
β_{time}	trav. time/100	trav. time/100	trav. time/100
β_{cost}	trav. cost/100	trav. cost/100	trav. cost/100
β_{headway}	headway/1000	headway/1000	0

Class 2: respondents who did not

Param.	Train	Swissmetro	Car
Cte train	1	0	0
Cte car	0	0	1
β_{cost}	trav. cost/100	trav. cost/100	trav. cost/100
β_{headway}	headway/1000	headway/1000	0

Latent classes

Class specific models

$P_n(\text{car}|n \in \text{class 1})$. $P_n(\text{car}|n \in \text{class 2})$.

Choice model: discrete mixture

$$P_n(\text{car}) = P(\text{car}|n \in \text{class 1})\,\Pr(n \in \text{class 1}) + P(\text{car}|n \in \text{class 2})\,\Pr(n \in \text{class 2}).$$

Class membership model: simple example

$$\Pr(n \in \text{class 1}) = \omega, \quad \Pr(n \in \text{class 2}) = 1 - \omega.$$

Random parameters – Latent classes

	ASV	RC-Normal	RC-Lognormal		Latent class
Nbr param.	5	6	6		6
Sample size	6768	6768	6768		6768
Final $\mathcal{L}$	-5315.39	-5196.84	-5215.01		-5191.09
AIC	10640.77	10405.69	10442.01		10394.18
BIC	10674.87	10446.61	10482.93		10435.10
Cte car	-0.262	0.0126	0.0553		0.00281
Cte train	-0.451	-0.104	-0.0666		-0.108
β_{cost}	-1.08	-1.29	-1.39		-1.27
β_{headway}	-5.35	-6.37	-5.96		-6.13
β_{time}	-1.28	-2.28	0.575	-3.85	-2.81
σ_{time}		-1.69	1.24	7.42	
Pr(Class 1)					0.749
Pr($\beta_{\text{time}} \geq 0$)	0.0%	8.01%	0.0%		0.0%

Latent classes

Class specific models

$P(\text{car}|n \in \text{class 1})$. $P(\text{car}|n \in \text{class 2})$.

Choice model: discrete mixture

$$P_n(\text{car}) = P(\text{car}|n \in \text{class 1}) \Pr(n \in \text{class 1}) + P(\text{car}|n \in \text{class 2}) \Pr(n \in \text{class 2}).$$

Class membership model: another example

$$\Pr(n \in \text{class 1}) = \frac{1}{1 + \exp(\omega)},$$

where

$$\omega = \gamma_0 + \gamma_{\text{male}}\text{male} + \gamma_{\text{GA}}\text{GA} + \gamma_{\text{business}}\text{business} + \gamma_{\text{lowInc}}\text{lowInc} + \gamma_{\text{first}}\text{first}.$$

Class membership

male	GA	bus.	lowInc	first	Model 1	Model 2
0	0	0	0	0	74.9%	82.1%
0	0	0	0	1	74.9%	92.5%
0	0	0	1	0	74.9%	76.7%
0	1	0	0	0	74.9%	6.36%
0	1	0	0	1	74.9%	15.5%
0	1	0	1	0	74.9%	4.64%
1	0	0	0	0	74.9%	92.2%
1	0	0	0	1	74.9%	96.9%
1	0	0	1	0	74.9%	89.4%
1	1	0	0	0	74.9%	14.8%
1	1	0	0	1	74.9%	31.9%
1	1	0	1	0	74.9%	11.1%

Latent classes

β_{time}: random parameter

$P(\text{car}|n \in \text{class 1})$.

$\beta_{\text{time}} = 0$

$P(\text{car}|n \in \text{class 2})$.

Choice model

$$P_n(\text{car}) = P(\text{car}|n \in \text{class 1}) \Pr(n \in \text{class 1}) + P(\text{car}|n \in \text{class 2}) \Pr(n \in \text{class 2}).$$

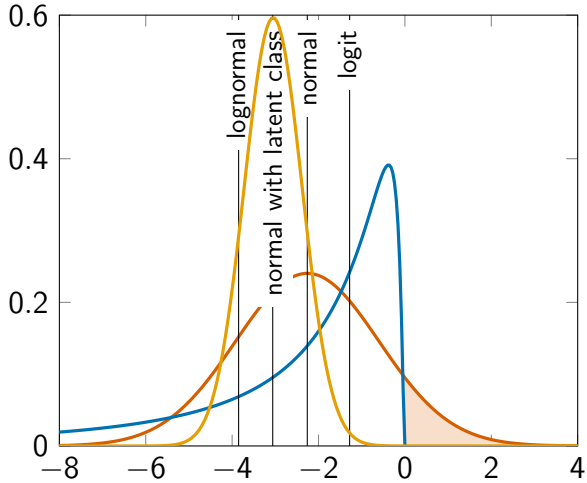
Class membership model

$$\Pr(n \in \text{class 1}) = \frac{1}{1 + \exp(\omega)},$$

where

$$\omega = \gamma_0 + \gamma_{\text{male}}\text{male} + \gamma_{\text{GA}}\text{GA} + \gamma_{\text{business}}\text{business} + \gamma_{\text{lowInc}}\text{lowInc} + \gamma_{\text{first}}\text{first}.$$

Random parameters: time coefficient



Random parameters: summary of the models

	K	$\mathcal{L}(\beta^*)$	AIC	BIC
Logit	5	-5315.39	10640.77	10674.87
RC (normal)	6	-5196.84	10405.69	10446.61
RC (lognormal)	6	-5215.01	10442.01	10482.93
Latent	6	-5191.09	10394.18	10435.10
Latent (socio-eco)	11	-4928.80	9879.60	9954.61
Latent + RC	12	-4927.96	9879.91	9961.75

Conclusion

Latent class (socio-eco), with β_t fixed is preferred.

Latent classes

Capture unobserved heterogeneity

They can represent different model specifications:

- choice sets,
- decision protocols,
- taste parameters,
- model structures,
- etc.

Latent classes

Model structure

$$P_n(i|x_n, z_n, \mathcal{C}_n) = \sum_{s=1}^{S} P_n(i|x_n, z_n, \mathcal{C}_n, s) Q_n(s|z_n).$$

- $P_n(i|x_n, z_n, \mathcal{C}_n, s)$ is the class-specific choice model:
  - probability of choosing i given that the individual n belongs to class s.
- $Q_n(s|z_n)$ is the class membership model:
  - probability of belonging to class s.

Individual level parameters

Motivation

- A person's choice reveals something about her tastes.
- We now describe how to derive the distribution of random parameters of mixtures of logit, conditional on the observed choice.
- This is usually referred to as "individual level parameters".

Individual-level parameters

Mixture of logit

$$P(i|x,\theta) = \int P(i|x,\beta) f(\beta|\theta) d\beta.$$

Discussion

- Taste parameter β is distributed in the population.
- Distribution: $f(\beta|\theta)$.
- Consider individuals choosing a specific alternative i under a context defined by x.
- How is β distributed for them?

[Revelt and Train, 2000]

Derivation

Distribution of β

- In the population: $f(\beta|\theta)$.
- In the sub-population: $h(\beta|i, x, \theta)$.

Bayes theorem

$$h(\beta|i, x, \theta)P(i|x, \theta) = P(i|x, \beta)f(\beta|\theta).$$

Individual-level parameters

$$h(\beta|i, x, \theta) = \frac{P(i|x, \beta)f(\beta|\theta)}{P(i|x, \theta)} = \frac{P(i|x, \beta)f(\beta|\theta)}{\int P(i|x, \beta)f(\beta|\theta)d\beta}.$$

Expected value

Individual-level parameters

$$h(\beta|i,x,\theta) = \frac{P(i|x,\beta)f(\beta|\theta)}{P(i|x,\theta)} = \frac{P(i|x,\beta)f(\beta|\theta)}{\int P(i|x,\beta)f(\beta|\theta)d\beta}.$$

Expected value

$$\bar{\beta} = \int \beta\, h(\beta|i,x,\theta)d\beta = \frac{\int \beta\, P(i|x,\beta)f(\beta|\theta)d\beta}{\int P(i|x,\beta)f(\beta|\theta)d\beta}.$$

Simulation

Procedure

- Each individual n in the sample is associated with a configuration (i_n, x_n).
- For each of them, we calculate $\bar{\beta}_n$.

Example

- Optima data (mode choice in Switzerland).
- Coefficient of waiting time distributed.

Example: Optima

Specification table

	Public transp.	Car	Slow modes
β_1	0	1	0
β_2	0	0	1
β_3	Travel time (min)	0	0
β_4		Travel time (min)	0
β_5	Waiting time (min)	0	0
β_6	Cost if HWH (CHF)	Cost if HWH (CHF)	0
β_7	Cost if not HWH (CHF)	Cost if not HWH (CHF)	0
β_8	0	0	Distance

Random coefficient

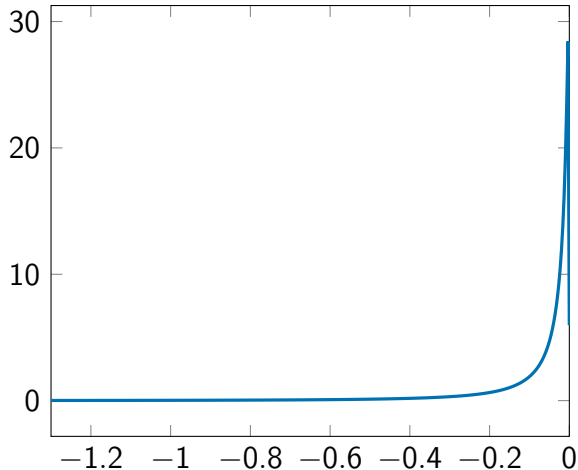
$$\beta_5 = -\exp(N(\bar{\beta}_5, \sigma_5))$$

Example: Optima

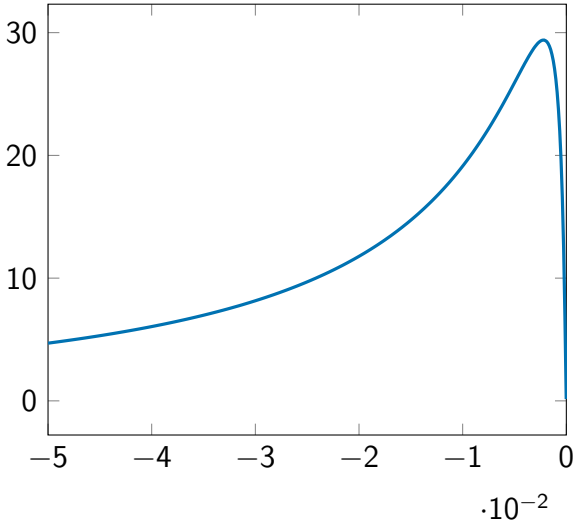
Number of estimated parameters	9
Sample size	1693
Null log likelihood	-1826.702
Final log likelihood	-985.2696
Rho-square-bar for the null model	0.456
Akaike Information Criterion	1988.539
Bayesian Information Criterion	2037.448
Number of draws	20000

	Value	Std err	t-test	p-value
β_1	0.955	0.13	7.32	2.4e-13
β_2	0.197	0.328	0.602	0.547
β_3	-0.0114	0.00453	-2.52	0.0116
β_4	-0.0456	0.0142	-3.22	0.00128
$\bar{\beta}_5$	-3.46	0.495	-7.0	2.55e-12
σ_5	1.63	0.426	3.82	0.000135
β_6	-0.133	0.0366	-3.62	0.000289
β_7	-0.0744	0.0232	-3.2	0.00136
β_8	-1.21	0.309	-3.92	8.96e-05

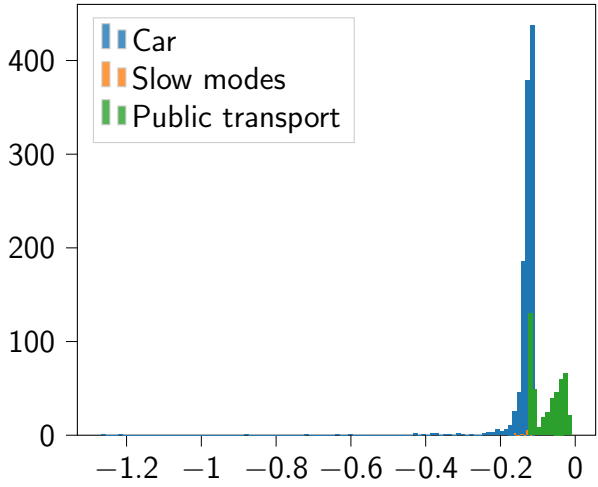
Example: dist. of waiting time coefficient (population)



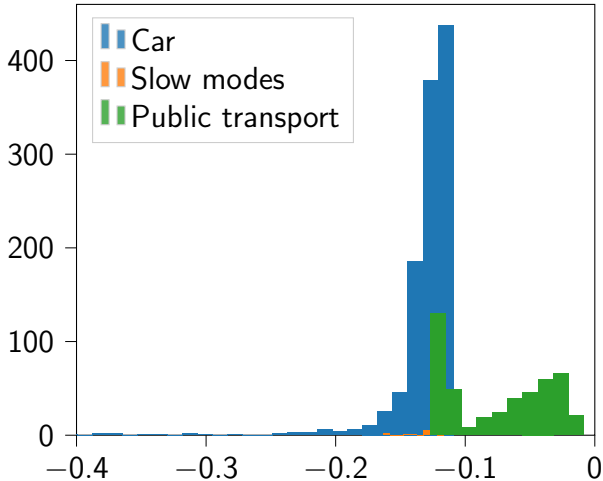
Example: dist. of waiting time coefficient (population, zoom)



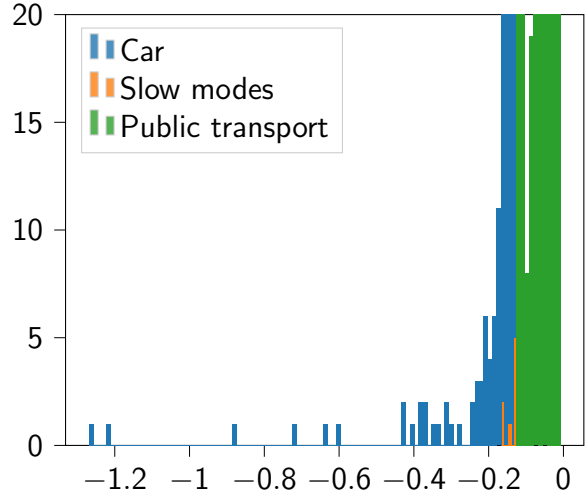
Example: dist. of waiting time coefficient (individuals)



Example: dist. of waiting time coefficient (individuals, zoom)



Example: dist. of waiting time coefficient (individuals, zoom)



Summary

- Mixtures.
- Monte-Carlo integration.
- Relaxing i.i.d.
- Taste heterogeneity.
- Latent classes.
- Individual-level parameters.

Tips for applications

- Be careful: simulation can mask specification and identification issues.
- Do not forget about the systematic portion.

Bibliography I

Revelt, D. and Train, K. (2000).
Customer-specific taste parameters and mixed logit: households' choice of electricity supplier.
Working paper E00-274, Department of Economics, University of California Berkeley.

Walker, J. L. (2001).
Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables.
PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.