Mixture Models — Simulation-based Estimation

Michel Bierlaire

michel.bierlaire@epfl.ch

Transport and Mobility Laboratory

Outline

- Mixtures
- Capturing correlation
- Alternative specific variance
- Taste heterogeneity
- Latent classes
- Simulation-based estimation

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Mixtures

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

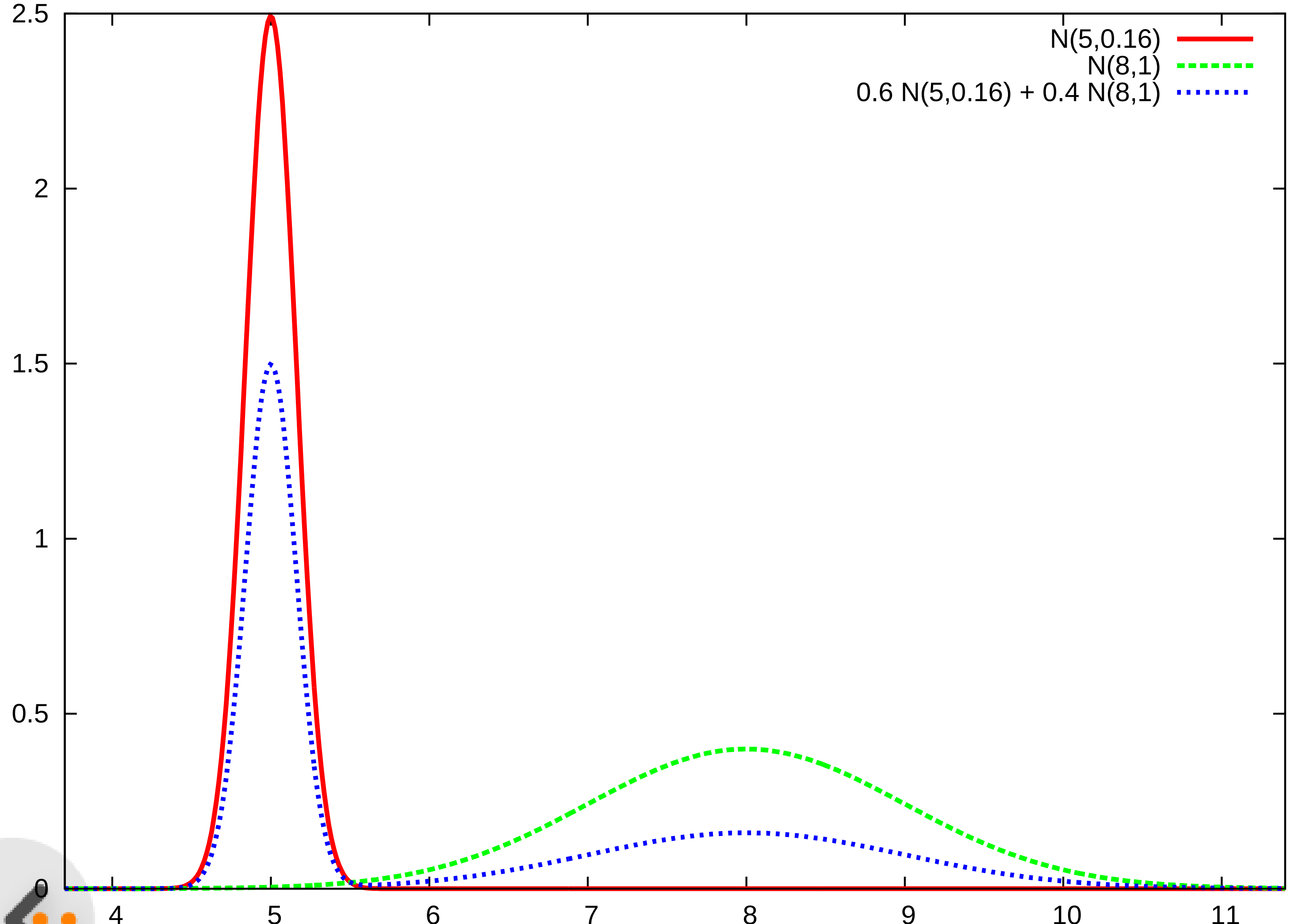
$$\sum_{i=1}^{n} w_i = 1$$

then

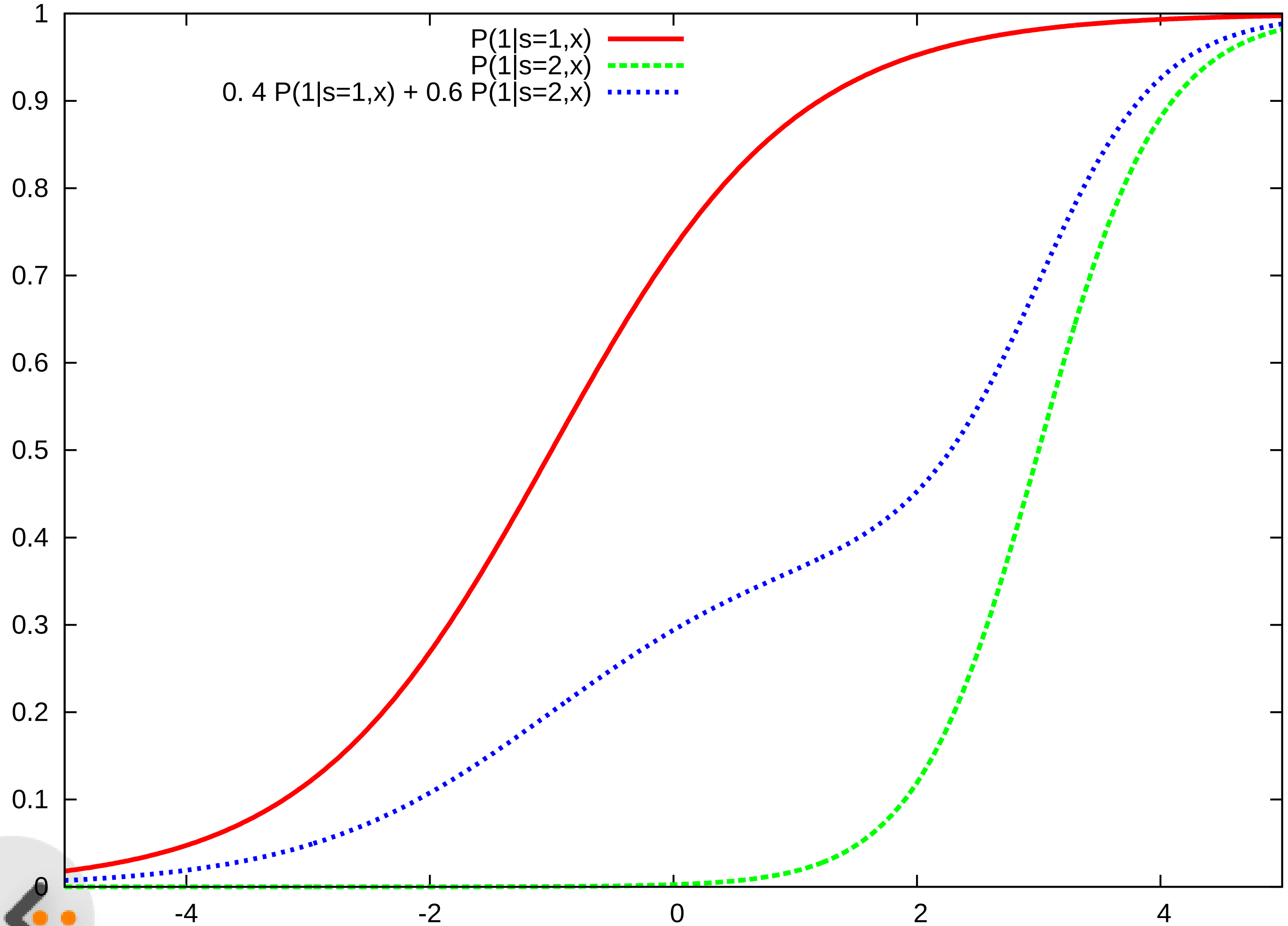
$$g(\varepsilon) = \sum_{i=1}^{n} w_i f(\varepsilon, \theta_i)$$

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



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Mixtures

- General motivation: generate flexible distributional forms
- For discrete choice:
  - correlation across alternatives
  - alternative specific variances
  - taste heterogeneity
  - ...

Back to the telephone example

Budget measured:	U_{BM}	$=$	α_{BM}	$+$	βX_{BM}	$+$	ε_{BM}
Standard measured:	U_{SM}	$=$	α_{SM}	$+$	βX_{SM}	$+$	ε_{SM}
Local flat:	U_{LF}	$=$	α_{LF}	$+$	βX_{LF}	$+$	ε_{LF}
Extended area flat:	U_{EF}	$=$	α_{EF}	$+$	βX_{EF}	$+$	ε_{EF}
Metro area flat:	U_{MF}	$=$			βX_{MF}	$+$	ε_{MF}

Distributions for ε: logit, probit, nested logit

Back to the telephone example

Covariance of U

Logit

$$\begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

Probit

$$\begin{pmatrix} \sigma^2_{\text{BM}} & \sigma_{\text{BM,SM}} & \sigma_{\text{BM,LF}} & \sigma_{\text{BM,EF}} & \sigma_{\text{BM,MF}} \\ \sigma_{\text{BM,SM}} & \sigma^2_{\text{SM}} & \sigma_{\text{SM,LF}} & \sigma_{\text{SM,EF}} & \sigma_{\text{SM,MF}} \\ \sigma_{\text{BM,LF}} & \sigma_{\text{SM,LF}} & \sigma^2_{\text{LF}} & \sigma_{\text{LF,EF}} & \sigma_{\text{LF,MF}} \\ \sigma_{\text{BM,EF}} & \sigma_{\text{SM,EF}} & \sigma_{\text{LF,EF}} & \sigma^2_{\text{EF}} & \sigma_{\text{EF,MF}} \\ \sigma_{\text{BM,MF}} & \sigma_{\text{SM,MF}} & \sigma_{\text{LF,MF}} & \sigma_{\text{EF,MF}} & \sigma^2_{\text{MF}} \end{pmatrix}$$

Nested logit

$$\frac{\pi^2}{6\mu^2} \begin{pmatrix} 1 & \rho_M & 0 & 0 & 0 \\ \rho_M & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \rho_F & \rho_F \\ 0 & 0 & \rho_F & 1 & \rho_F \\ 0 & 0 & \rho_F & \rho_F & 1 \end{pmatrix}, \ \rho_i = 1 - \frac{\mu^2}{\mu_i^2}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Continuous Mixtures of logit

- Combining probit and logit
- Error decomposed into two parts

$$U_{in} = V_{in} + \xi + \nu$$

i.i.d EV (logit): tractability

Normal distribution (probit): flexibility

Logit

- Utility:

$$\begin{array}{rcccl} U_{\text{auto}} & = & \beta X_{\text{auto}} & + & \nu_{\text{auto}} \\ U_{\text{bus}} & = & \beta X_{\text{bus}} & + & \nu_{\text{bus}} \\ U_{\text{subway}} & = & \beta X_{\text{subway}} & + & \nu_{\text{subway}} \end{array}$$

- ν i.i.d. extreme value
- Probability:

$$\Lambda(\text{auto}|X) = \frac{e^{\beta X_{\text{auto}}}}{e^{\beta X_{\text{auto}}} + e^{\beta X_{\text{bus}}} + e^{\beta X_{\text{subway}}}}$$

Normal mixture of logit

- Utility:

$$
\begin{array}{rclllll}
U_{\text{auto}} & = & \beta X_{\text{auto}} & + & \xi_{\text{auto}} & + & \nu_{\text{auto}} \\
U_{\text{bus}} & = & \beta X_{\text{bus}} & + & \xi_{\text{bus}} & + & \nu_{\text{bus}} \\
U_{\text{subway}} & = & \beta X_{\text{subway}} & + & \xi_{\text{subway}} & + & \nu_{\text{subway}}
\end{array}
$$

- ν i.i.d. extreme value, $\xi \sim N(0, \Sigma)$
- Probability:

$$
\Lambda(\text{auto}|X, \xi) = \frac{e^{\beta X_{\text{auto}} + \xi_{\text{auto}}}}{e^{\beta X_{\text{auto}} + \xi_{\text{auto}}} + e^{\beta X_{\text{bus}} + \xi_{\text{bus}}} + e^{\beta X_{\text{subway}} + \xi_{\text{subway}}}}
$$

$$
P(\text{auto}|X) = \int_{\xi} \Lambda(\text{auto}|X, \xi) f(\xi) d\xi
$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Simulation

$$P(\text{auto}|X) = \int_\xi \Lambda(\text{auto}|X,\xi) f(\xi) d\xi$$

- Integral has no closed form.
- Monte Carlo simulation must be used.

Simulation

- In order to approximate

$$P(i|X) = \int_{\xi} \Lambda(i|X,\xi) f(\xi) d\xi$$

- Draw from $f(\xi)$ to obtain $r_1, \ldots, r_R$
- Compute

$$\begin{aligned} P(i|X) \approx \tilde{P}(i|X) &= \frac{1}{R} \sum_{k=1}^{R} P(i|X, r_k) \\ &= \frac{1}{R} \sum_{k=1}^{R} \frac{e^{V_{1n}+r_k}}{e^{V_{1n}+r_k} + e^{V_{2n}+r_k} + e^{V_{3n}}} \end{aligned}$$

Capturing correlations: nesting

- Utility:

$$
\begin{array}{rcllll}
U_{\text{auto}} & = & \beta X_{\text{auto}} & & & + & \nu_{\text{auto}} \\
U_{\text{bus}} & = & \beta X_{\text{bus}} & + & \sigma_{\text{transit}}\eta_{\text{transit}} & + & \nu_{\text{bus}} \\
U_{\text{subway}} & = & \beta X_{\text{subway}} & + & \sigma_{\text{transit}}\eta_{\text{transit}} & + & \nu_{\text{subway}}
\end{array}
$$

- ν i.i.d. extreme value, $\eta_{\text{transit}} \sim N(0,1)$, $\sigma^2_{\text{transit}}$ =cov(bus,subway)
- Probability:

$$
\Lambda(\text{auto}|X,\eta_{\text{transit}}) = \frac{e^{\beta X_{\text{auto}}}}{e^{\beta X_{\text{auto}}} + e^{\beta X_{\text{bus}}+\sigma_{\text{transit}}\eta_{\text{transit}}} + e^{\beta X_{\text{subway}}+\sigma_{\text{transit}}\eta_{\text{transit}}}}
$$

$$
P(\text{auto}|X) = \int_{\eta} \Lambda(\text{auto}|X,\xi)f(\eta)d\eta
$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Nesting structure

Example: residential telephone

	ASC_BM	ASC_SM	ASC_LF	ASC_EF	BETA_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

Nesting structure

Identification issues:

- If there are two nests, only one σ is identified
- If there are more than two nests, all σ's are identified

Walker (2001)

Results with 5000 draws..

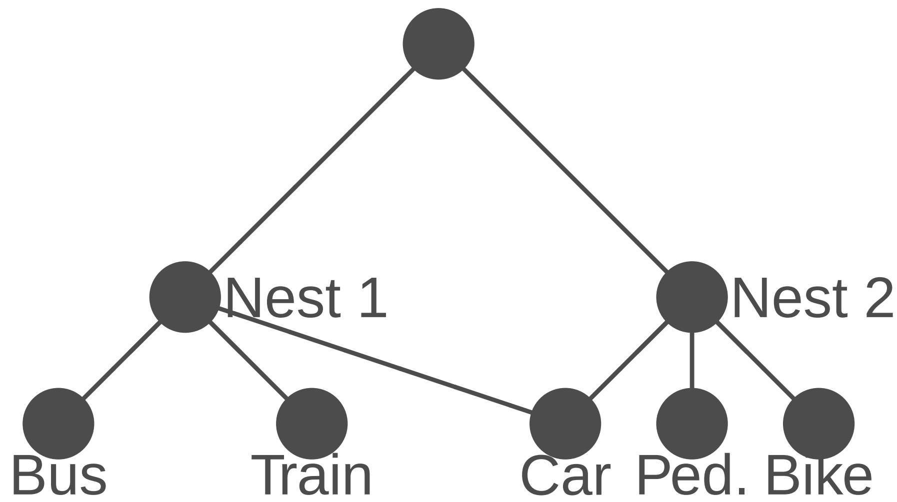
	NL		NML		NML $\sigma_F = 0$		NML $\sigma_M = 0$		NML $\sigma_F = \sigma_M$	
$\mathcal{L}$	-473.219		-472.768		-473.146		-472.779		-472.846	
	Value	Scaled	Value	Scaled	Value	Scaled	Value	Scaled	Value	Scaled
ASC_BM	-1.784	1.000	-3.81247	1.000	-3.79131	1.000	-3.80999	1.000	-3.81327	1.000
ASC_EF	-0.558	0.313	-1.19899	0.314	-1.18549	0.313	-1.19711	0.314	-1.19672	0.314
ASC_LF	-0.512	0.287	-1.09535	0.287	-1.08704	0.287	-1.0942	0.287	-1.0948	0.287
ASC_SM	-1.405	0.788	-3.01659	0.791	-2.9963	0.790	-3.01426	0.791	-3.0171	0.791
B_LOGCOST	-1.490	0.835	-3.25782	0.855	-3.24268	0.855	-3.2558	0.855	-3.25805	0.854
FLAT	2.292									
MEAS	2.063									
σ_F			3.02027		0		3.06144		2.17138	
σ_M			0.52875		3.024833		0		2.17138	
$\sigma_F^2 + \sigma_M^2$			9.402		9.150		9.372		9.430	

Comments

- The scale of the parameters is different between NL and the mixture model
- Normalization can be performed in several ways
  - $\sigma_F = 0$
  - $\sigma_M = 0$
  - $\sigma_F = \sigma_M$
- Final log likelihood should be the same
- But... estimation relies on simulation
- Only an approximation of the log likelihood is available
- Final log likelihood with 50000 draws:

Unnormalized:	-472.872	$\sigma_M = \sigma_F$:	-472.875
$\sigma_F = 0$:	-472.884	$\sigma_M = 0$:	-472.901

Cross nesting



$$
\begin{array}{rclllll}
U_{\text{bus}} & = & V_{\text{bus}} & +\xi_1 & & +\varepsilon_{\text{bus}} \\
U_{\text{train}} & = & V_{\text{train}} & +\xi_1 & & +\varepsilon_{\text{train}} \\
U_{\text{car}} & = & V_{\text{car}} & +\xi_1 & +\xi_2 & +\varepsilon_{\text{car}} \\
U_{\text{ped}} & = & V_{\text{ped}} & & +\xi_2 & +\varepsilon_{\text{ped}} \\
U_{\text{bike}} & = & V_{\text{bike}} & & +\xi_2 & +\varepsilon_{\text{bike}}
\end{array}
$$

$$
P(\text{car}) = \int_{\xi_1}\int_{\xi_2} P(\text{car}|\xi_1,\xi_2) f(\xi_1) f(\xi_2) d\xi_2 d\xi_1
$$

Identification issue

- Not all parameters can be identified
- For logit, one ASC has to be constrained to zero
- Identification of NML is important and tricky
- See Walker, Ben-Akiva & Bolduc (2007) for a detailed analysis

Alternative specific variance

- Error terms in logit are i.i.d. and, in particular, have the same variance

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

- ε_{in} i.i.d. extreme value $\Rightarrow$ $\text{Var}(\varepsilon_{in}) = \pi^2/6\mu^2$
- In order allow for different variances, we use mixtures

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

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

- Variance:

$$\text{Var}(\sigma_i \xi_i + \varepsilon_{in}) = \sigma_i^2 + \frac{\pi^2}{6\mu^2}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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 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.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Alternative specific variance

Example with Swissmetro

	ASC_CAR	ASC_SBB	ASC_SM	B_COST	B_FR	B_TIME
Car	1	0	0	cost	0	time
Train	0	0	0	cost	freq.	time
Swissmetro	0	0	1	cost	freq.	time

+ alternative specific variance

	Logit		ASV		ASV norm.	
$\mathcal{L}$	-5315.39		-5241.01		-5242.10	
	Value	Scaled	Value	Scaled	Value	Scaled
ASC_CAR	0.189	1.000	0.248	1.000	0.241	1.000
ASC_SM	0.451	2.384	0.903	3.637	0.882	3.657
B_COST	-0.011	-0.057	-0.018	-0.072	-0.018	-0.073
B_FR	-0.005	-0.028	-0.008	-0.031	-0.008	-0.032
B_TIME	-0.013	-0.067	-0.017	-0.069	-0.017	-0.071
SIGMA_CAR			0.020			
SIGMA_TRAIN			0.039		0.061	
SIGMA_SM			3.224		3.180	

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

Heteroscedastic: specification

$$
\begin{array}{lclccccc}
U_1 & = & \beta x_1 & +\sigma_1\xi_1 & & & & +\varepsilon_1 \\
U_2 & = & \beta x_2 & & +\sigma_2\xi_2 & & & +\varepsilon_2 \\
U_3 & = & \beta x_3 & & & +\sigma_3\xi_3 & & +\varepsilon_3 \\
U_4 & = & \beta x_4 & & & & +\sigma_4\xi_4 & +\varepsilon_4
\end{array}
$$

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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Heteroscedastic: differences

$$
\begin{aligned}
U_1 - U_4 &= \beta(x_1 - x_4) + (\sigma_1\xi_1 - \sigma_4\xi_4) + (\varepsilon_1 - \varepsilon_4) \\
U_2 - U_4 &= \beta(x_2 - x_4) + (\sigma_2\xi_2 - \sigma_4\xi_4) + (\varepsilon_2 - \varepsilon_4) \\
U_3 - U_4 &= \beta(x_3 - x_4) + (\sigma_3\xi_3 - \sigma_4\xi_4) + (\varepsilon_3 - \varepsilon_4)
\end{aligned}
$$

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

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

Heteroscedastic: order condition

- 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 $S \leq 5$

Heteroscedastic: rank condition

Idea

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

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

Heteroscedastic: rank condition

Three parameters out of five can be estimated
Formally...

1. Identify unique elements of Cov(ΔU)
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

Heteroscedastic: equality condition

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

Heteroscedastic: equality condition

$$
\begin{array}{rcllll}
U_n & = & \beta^T x_n & + & L_n \xi_n & + \; \varepsilon_n \\
\text{Cov}(U_n) & = & & & L_n L_n^T & + \; (\gamma/\mu^2) I \\
\text{Cov}(\Delta_j U_n) & = & & & \Delta_j L_n L_n^T \Delta_j^T & + \; (\gamma/\mu^2) \Delta_j \Delta_j^T
\end{array}
$$

Notations:

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

$$
\begin{array}{rrcll}
\text{Cov}(\Delta_j U_n) = & \Omega_n & = & \Sigma_n & + \; \Gamma_n \\
& \Omega_n^{\text{norm}} & = & \Sigma_n^{\text{norm}} & + \; \Gamma_n^{\text{norm}}
\end{array}
$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Heteroscedastic: equality condition

The following conditions must hold:

- Covariance matrices must be equal

$$\Omega_n = \Omega_n^{\text{norm}}$$

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

Heteroscedastic: equality condition

Example with 3 alternatives:

$$\begin{array}{lllllll} U_1 & = & \beta x_1 & +\sigma_1\xi_1 & & & +\varepsilon_1 \\ U_2 & = & \beta x_2 & & +\sigma_2\xi_2 & & +\varepsilon_2 \\ U_3 & = & \beta x_3 & & & +\sigma_3\xi_3 & +\varepsilon_3 \end{array}$$

$$\text{Cov}(\Delta_3 U) = \Omega = \begin{pmatrix} \sigma_1^2 + \sigma_3^2 + 2\gamma/\mu^2 & \\ \sigma_3^2 + \gamma/\mu^2 & \sigma_2^2 + \sigma_3^2 + 2\gamma/\mu^2 \end{pmatrix}$$

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

Heteroscedastic: equality condition

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

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

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

where index N stands for “normalized”

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Heteroscedastic: equality condition

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

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

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

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

Heteroscedastic: 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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Heteroscedastic: equality condition

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

- 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.

Taste heterogeneity

- Population is heterogeneous
- Taste heterogeneity is captured by segmentation
- Deterministic segmentation is desirable but not always possible
- Distribution of a parameter in the population

Random parameters

$$\begin{aligned} U_i &= \beta_t T_i + \beta_c C_i + \varepsilon_i \\ U_j &= \beta_t T_j + \beta_c C_j + \varepsilon_j \end{aligned}$$

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

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

$$\begin{aligned} U_i &= \bar{\beta}_t T_i + \sigma_t \xi T_i + \beta_c C_i + \varepsilon_i \\ U_j &= \bar{\beta}_t T_j + \sigma_t \xi T_j + \beta_c C_j + \varepsilon_j \end{aligned}$$

If ε_i and ε_j are i.i.d. EV and ξ is given, we have

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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Random parameters

Example with Swissmetro

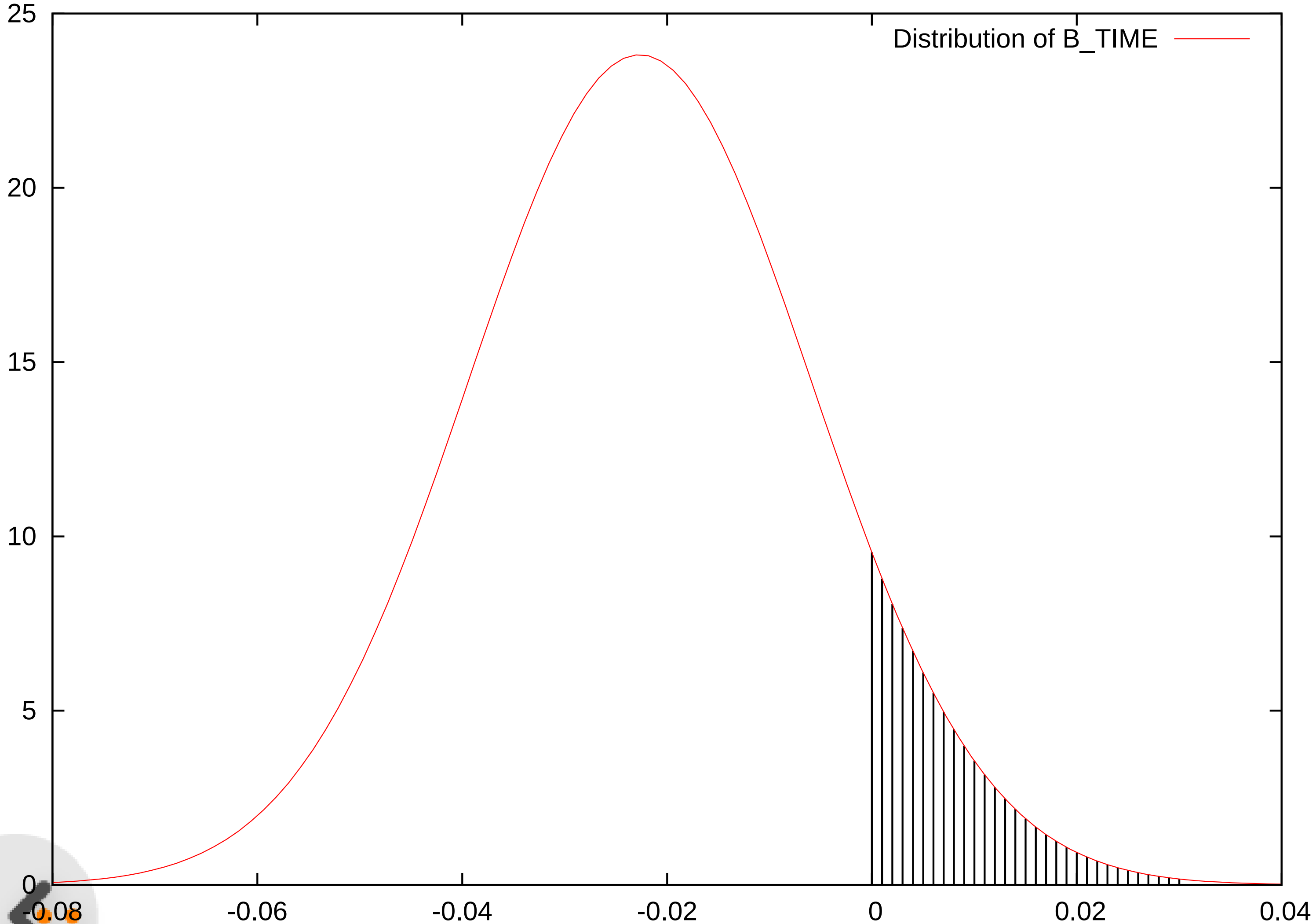
	ASC_CAR	ASC_SBB	ASC_SM	B_COST	B_FR	B_TIME
Car	1	0	0	cost	0	time
Train	0	0	0	cost	freq.	time
Swissmetro	0	0	1	cost	freq.	time

B_TIME randomly distributed across the population, normal distribution

Random parameters

	Logit	RC
$\mathcal{L}$	-5315.4	-5198.0
ASC_CAR_SP	0.189	0.118
ASC_SM_SP	0.451	0.107
B_COST	-0.011	-0.013
B_FR	-0.005	-0.006
B_TIME	-0.013	-0.023
S_TIME		0.017
Prob(B_TIME $\geq$ 0)		8.8%
χ^2		234.84

Random parameters



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Random parameters

Example with Swissmetro

	ASC_CAR	ASC_SBB	ASC_SM	B_COST	B_FR	B_TIME
Car	1	0	0	cost	0	time
Train	0	0	0	cost	freq.	time
Swissmetro	0	0	1	cost	freq.	time

B_TIME randomly distributed across the population, log normal distribution

Random parameters

```
[Utilities]
11 SBB_SP TRAIN_AV_SP ASC_SBB_SP * one        +
                      B_COST     * TRAIN_COST +
                      B_FR       * TRAIN_FR
21 SM_SP SM_AV        ASC_SM_SP  * one        +
                      B_COST     * SM_COST    +
                      B_FR * SM_FR
31 Car_SP CAR_AV_SP   ASC_CAR_SP * one        +
                      B_COST     * CAR_CO

[GeneralizedUtilities]
11 - exp( B_TIME [ S_TIME ] ) * TRAIN_TT
21 - exp( B_TIME [ S_TIME ] ) * SM_TT
31 - exp( B_TIME [ S_TIME ] ) * CAR_TT
```

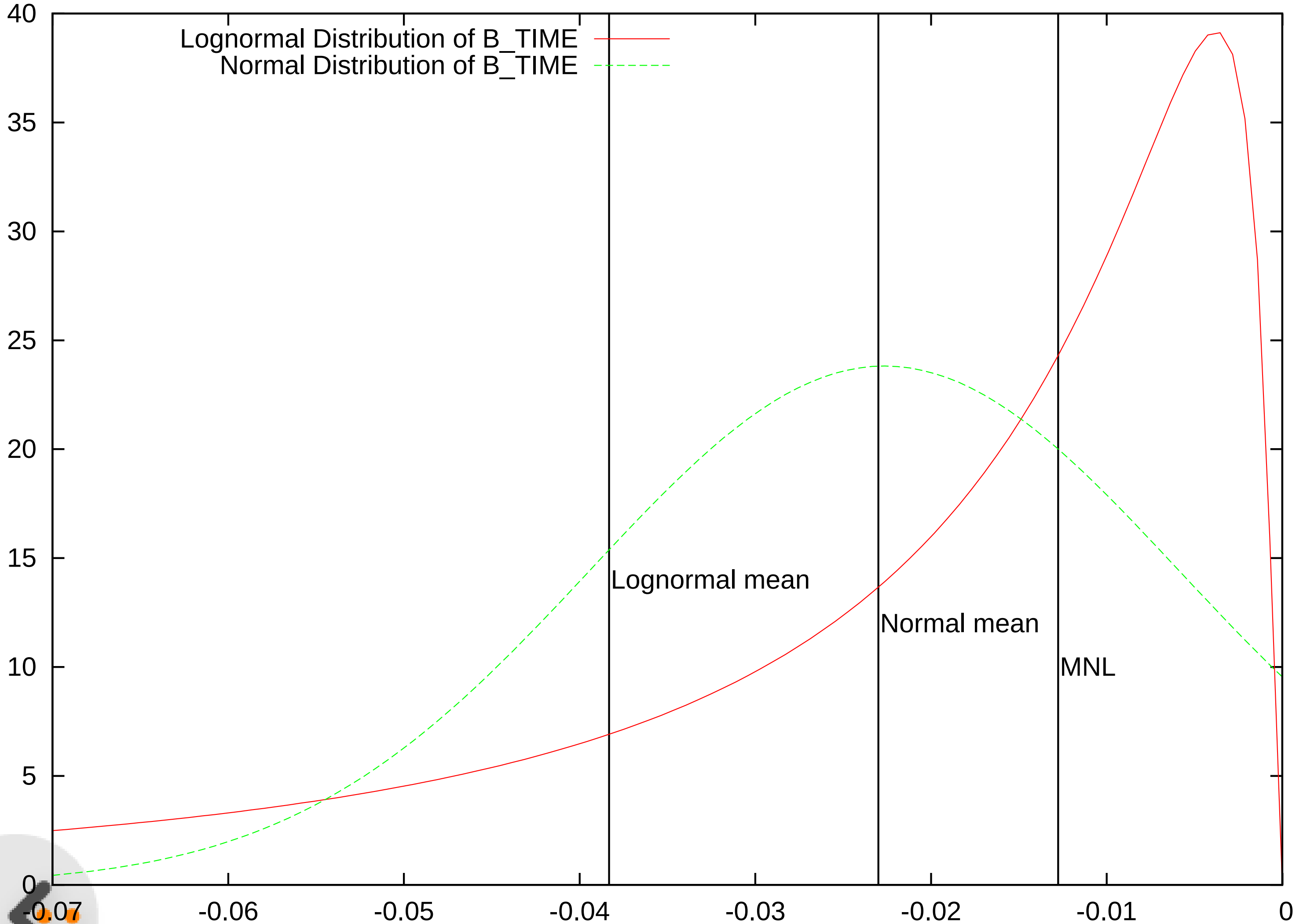
Random parameters

	Logit	RC-norm.	RC-logn.	
	-5315.4	-5198.0	-5215.81	
ASC_CAR_SP	0.189	0.118	0.122	
ASC_SM_SP	0.451	0.107	0.069	
B_COST	-0.011	-0.013	-0.014	
B_FR	-0.005	-0.006	-0.006	
B_TIME	-0.013	-0.023	-4.033	-0.038
S_TIME		0.017	1.242	0.073
Prob($\beta > 0$)		8.8%	0.0%	
χ^2		234.84	199.16	

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Random parameters



Random parameters

Example with Swissmetro

	ASC_CAR	ASC_SBB	ASC_SM	B_COST	B_FR	B_TIME
Car	1	0	0	cost	0	time
Train	0	0	0	cost	freq.	time
Swissmetro	0	0	1	cost	freq.	time

B_TIME randomly distributed across the population, discrete distribution

$$P(\beta_{\text{time}} = \hat{\beta}) = \omega_1 \quad P(\beta_{\text{time}} = 0) = \omega_2 = 1 - \omega_1$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Random parameters

```
[DiscreteDistributions]
B_TIME < B_TIME_1 ( W1 ) B_TIME_2 ( W2 ) >

[LinearConstraints]
W1 + W2 = 1.0
```

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Random parameters

	Logit	RC-norm.	RC-logn.		RC-disc.
	-5315.4	-5198.0	-5215.8		-5191.1
ASC_CAR_SP	0.189	0.118	0.122		0.111
ASC_SM_SP	0.451	0.107	0.069		0.108
B_COST	-0.011	-0.013	-0.014		-0.013
B_FR	-0.005	-0.006	-0.006		-0.006
B_TIME	-0.013	-0.023	-4.033	-0.038	-0.028
					0.000
S_TIME		0.017	1.242	0.073	
W1					0.749
W2					0.251
Prob($\beta > 0$)		8.8%	0.0%		0.0%
χ^2		234.84	199.16		248.6

Latent classes

- Latent classes capture unobserved heterogeneity
- They can represent different:
  - Choice sets
  - Decision protocols
  - Tastes
  - Model structures
  - etc.

Latent classes

$$P(i) = \sum_{s=1}^{S} \Lambda(i|s)Q(s)$$

- $\Lambda(i|s)$ is the class-specific choice model
  - *probability of choosing i given that the individual belongs to class s*
- $Q(s)$ is the class membership model
  - *probability of belonging to class s*

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Summary

- Logit mixtures models
  - Computationally more complex than MEV
  - Allow for more flexibility than MEV
- Continuous mixtures: alternative specific variance, nesting structures, random parameters

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

- Discrete mixtures: well-defined latent classes of decision makers

$$P(i) = \sum_{s=1}^{S} \Lambda(i|s) Q(s).$$

Tips for applications

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

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

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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Simulation

$$P(i) \approx \frac{1}{R} \sum_{r=1}^{R} \Lambda(i|r).$$

The kernel is a logit model, easy to compute.

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

Therefore, it amounts to generating the appropriate draws.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Appendix: Simulation

Pseudo-random numbers generators

Although deterministically generated, numbers exhibit the properties of random draws

- Uniform distribution
- Standard normal distribution
- Transformation of standard normal
- Inverse CDF
- Multivariate normal

Appendix: 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 $U(a,b)$

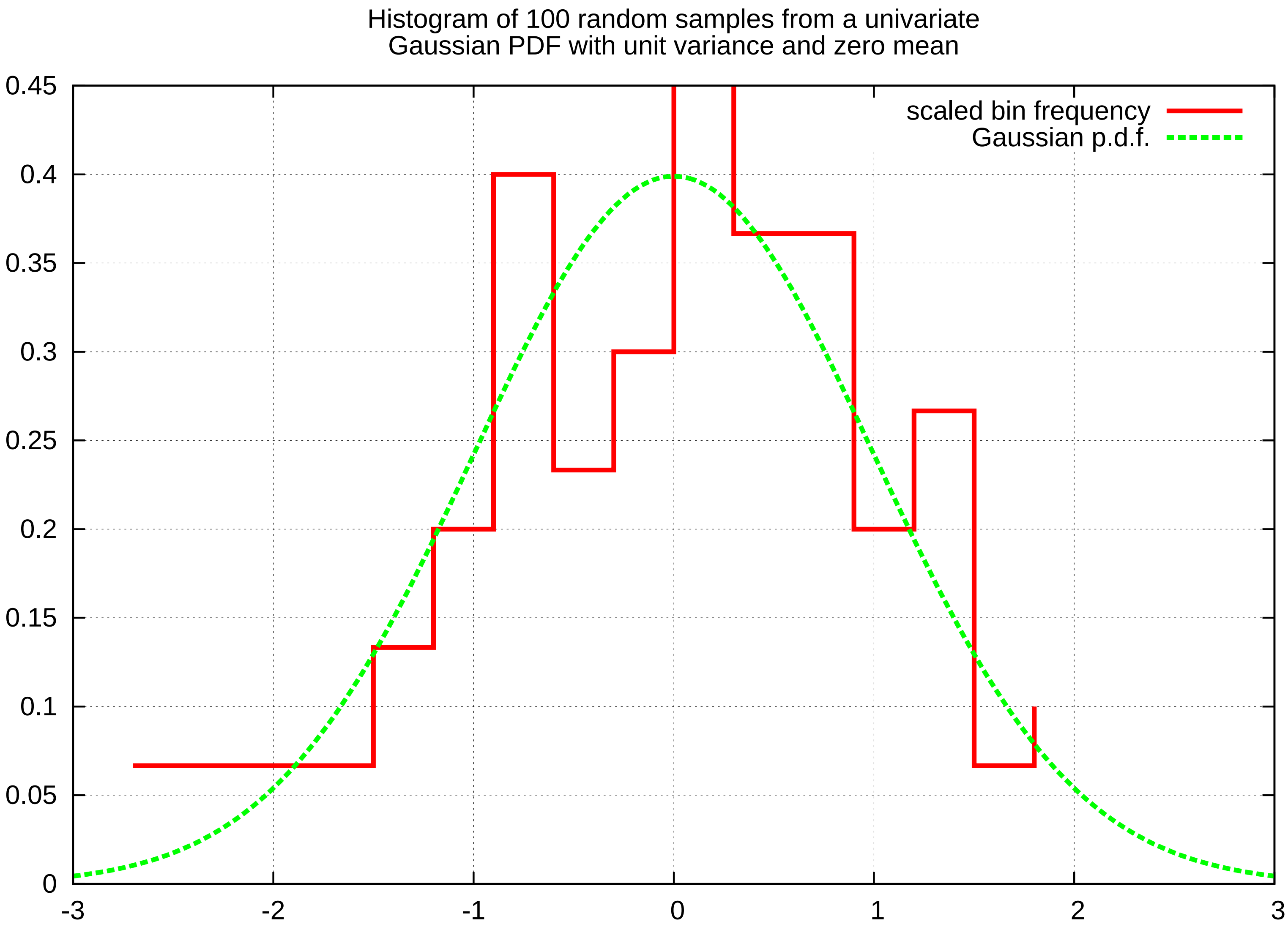
Appendix: Simulation: standard normal

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

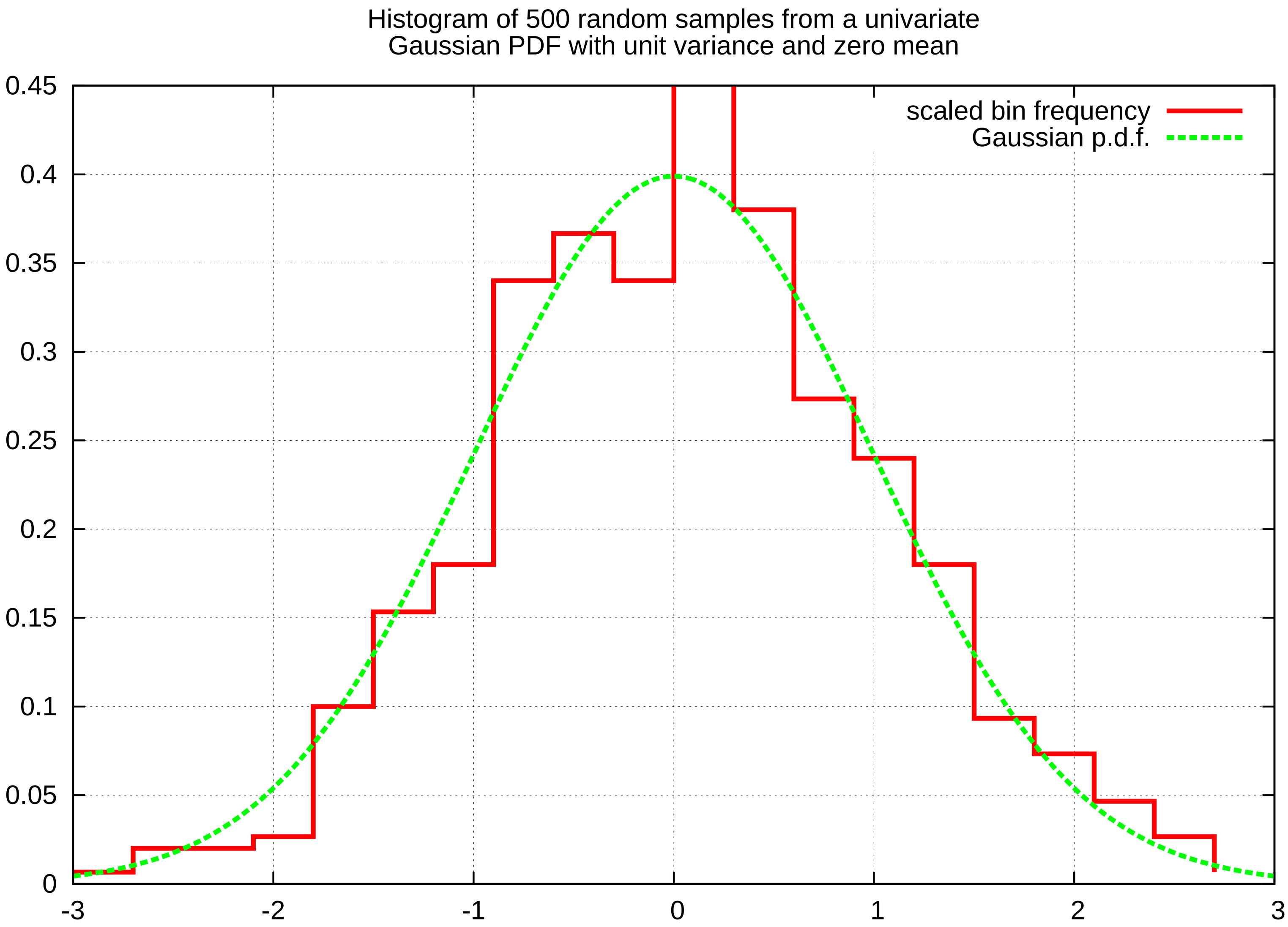
$$s_1 = \sqrt{-2\ln r_1}\,\sin(2\pi r_2)$$
$$s_2 = \sqrt{-2\ln r_1}\,\cos(2\pi r_2)$$

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

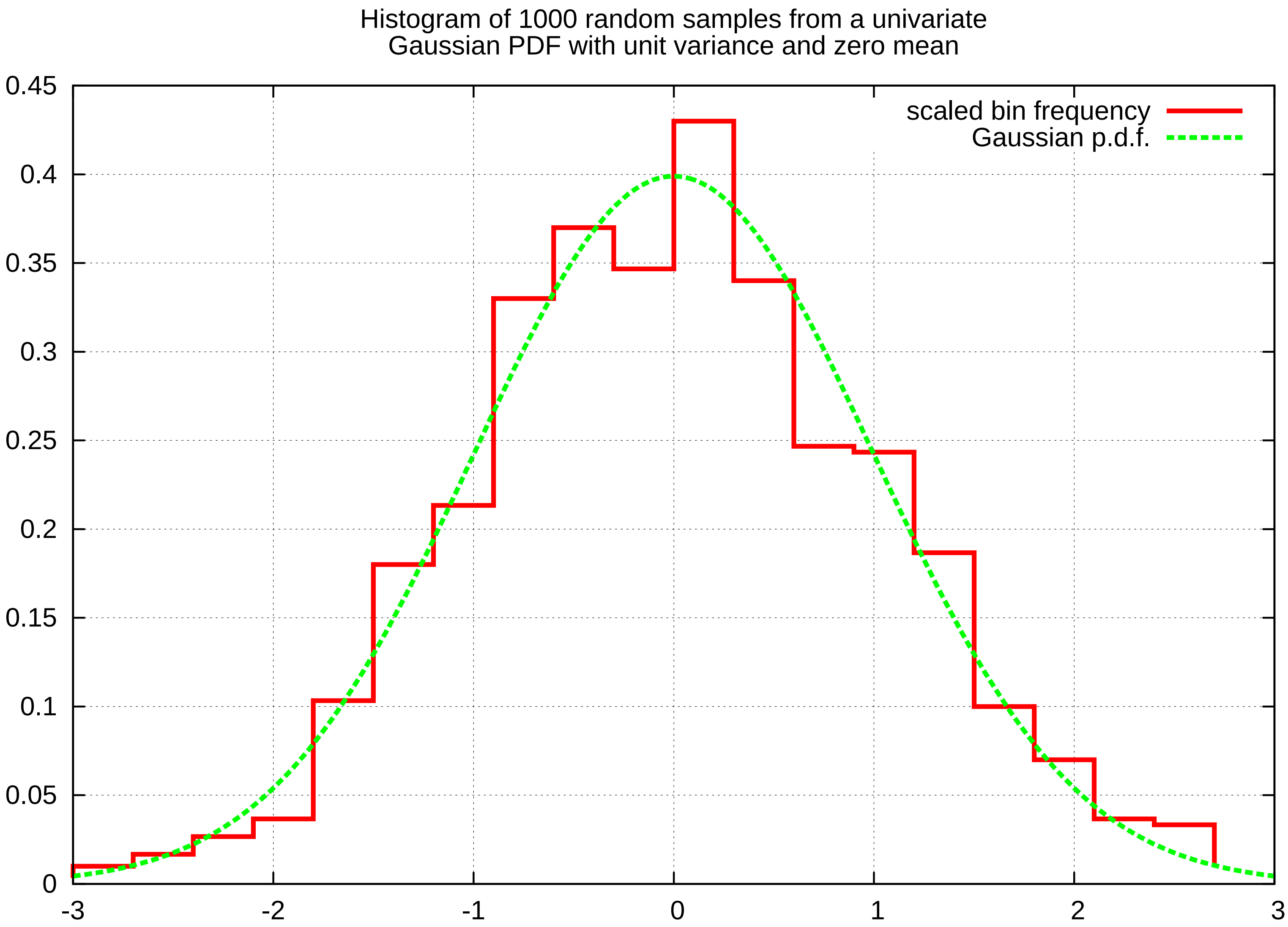
Appendix: Simulation: standard normal



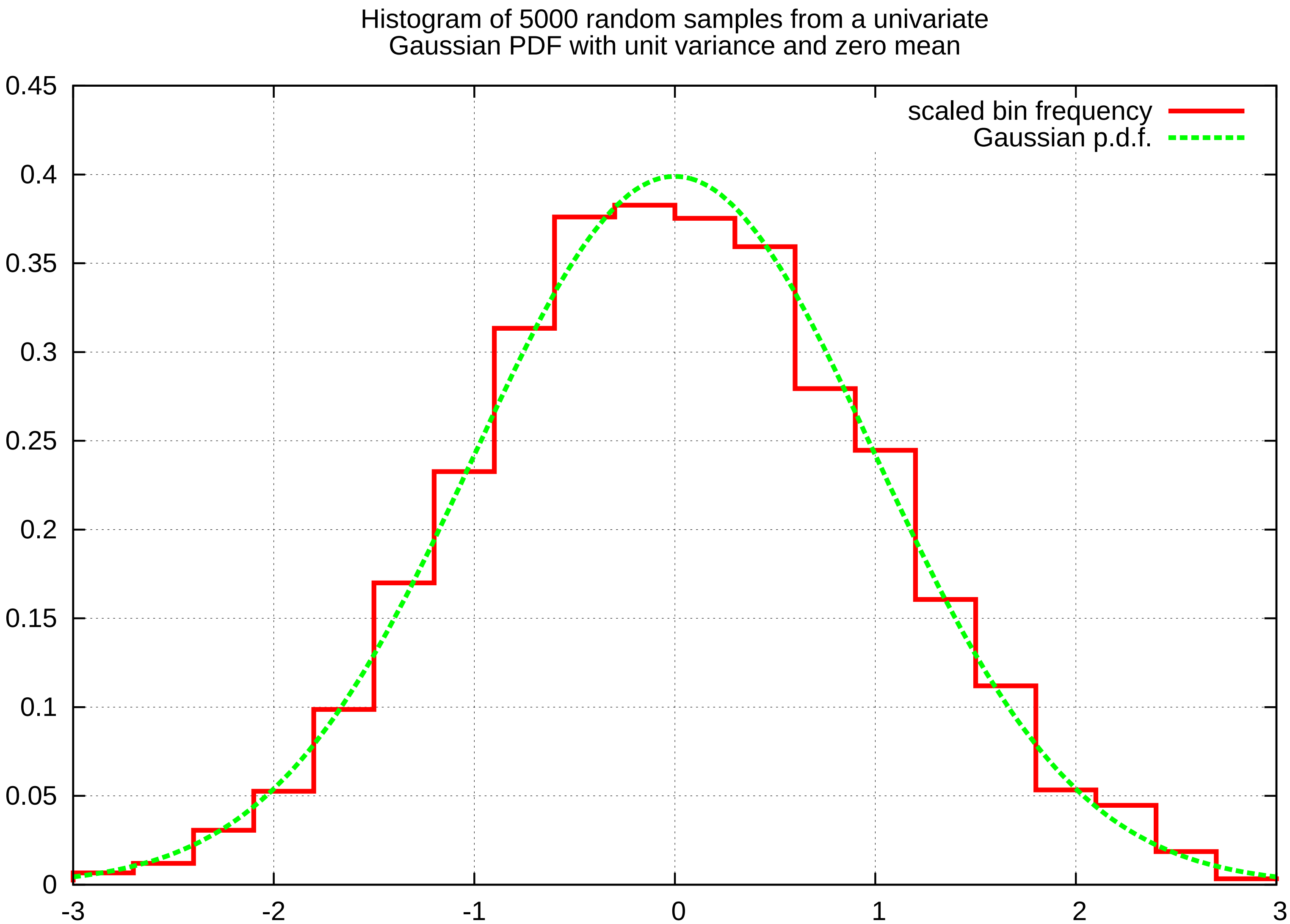
Appendix: Simulation: standard normal



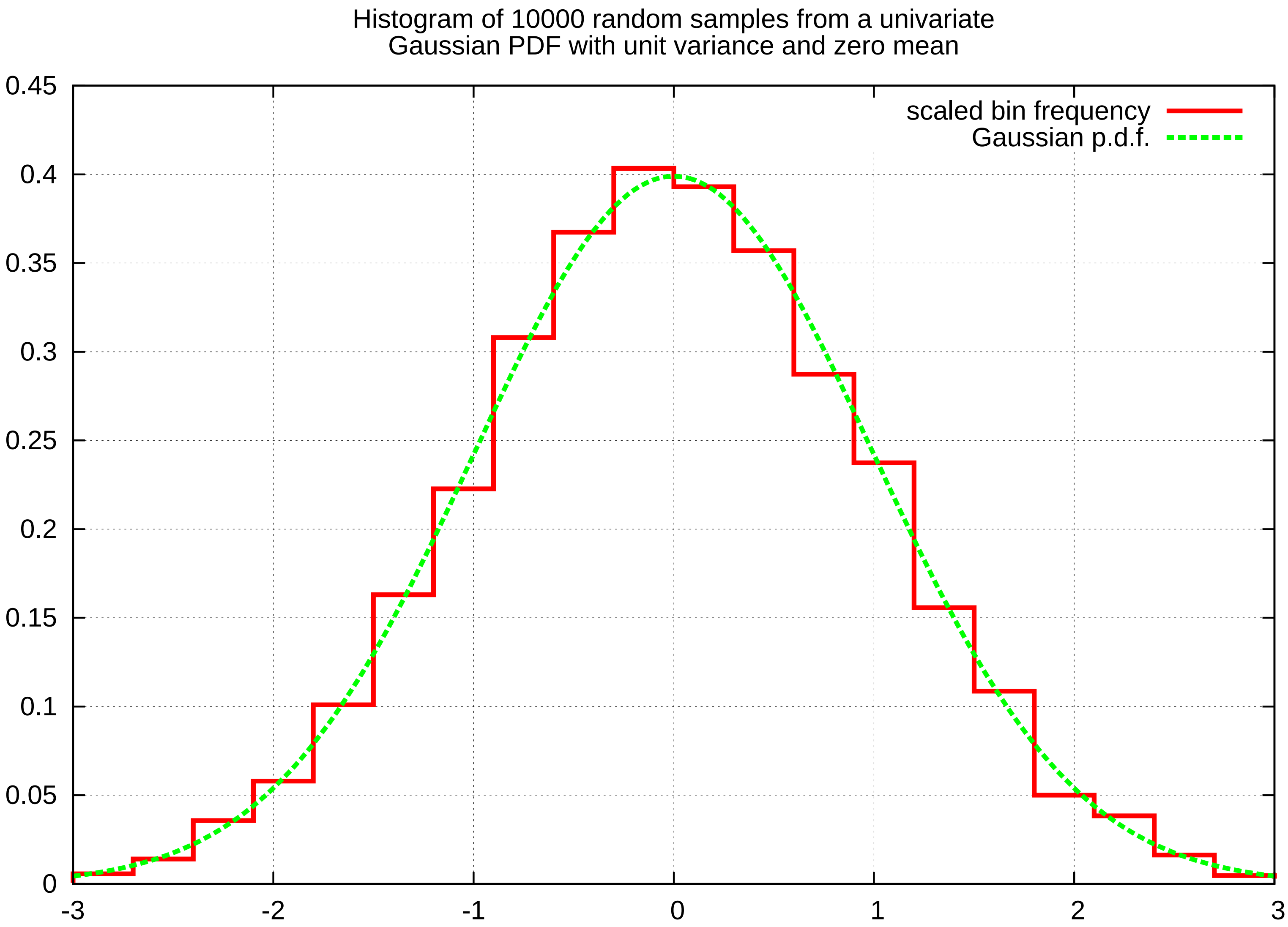
Appendix: Simulation: standard normal



Appendix: Simulation: standard normal



Appendix: Simulation: standard normal



Appendix: Simulation: transformations of standard no

- If r is a draw from $N(0,1)$, then

$$s = br + a$$

is a draw from $N(a, b^2)$

- If r is a draw from $N(a, b^2)$, then

$$e^r$$

is a draw from a log normal $LN(a, b^2)$ with mean

$$e^{a+(b^2/2)}$$

and variance

$$e^{2a+b^2}(e^{b^2} - 1)$$

Appendix: Simulation: inverse CDF

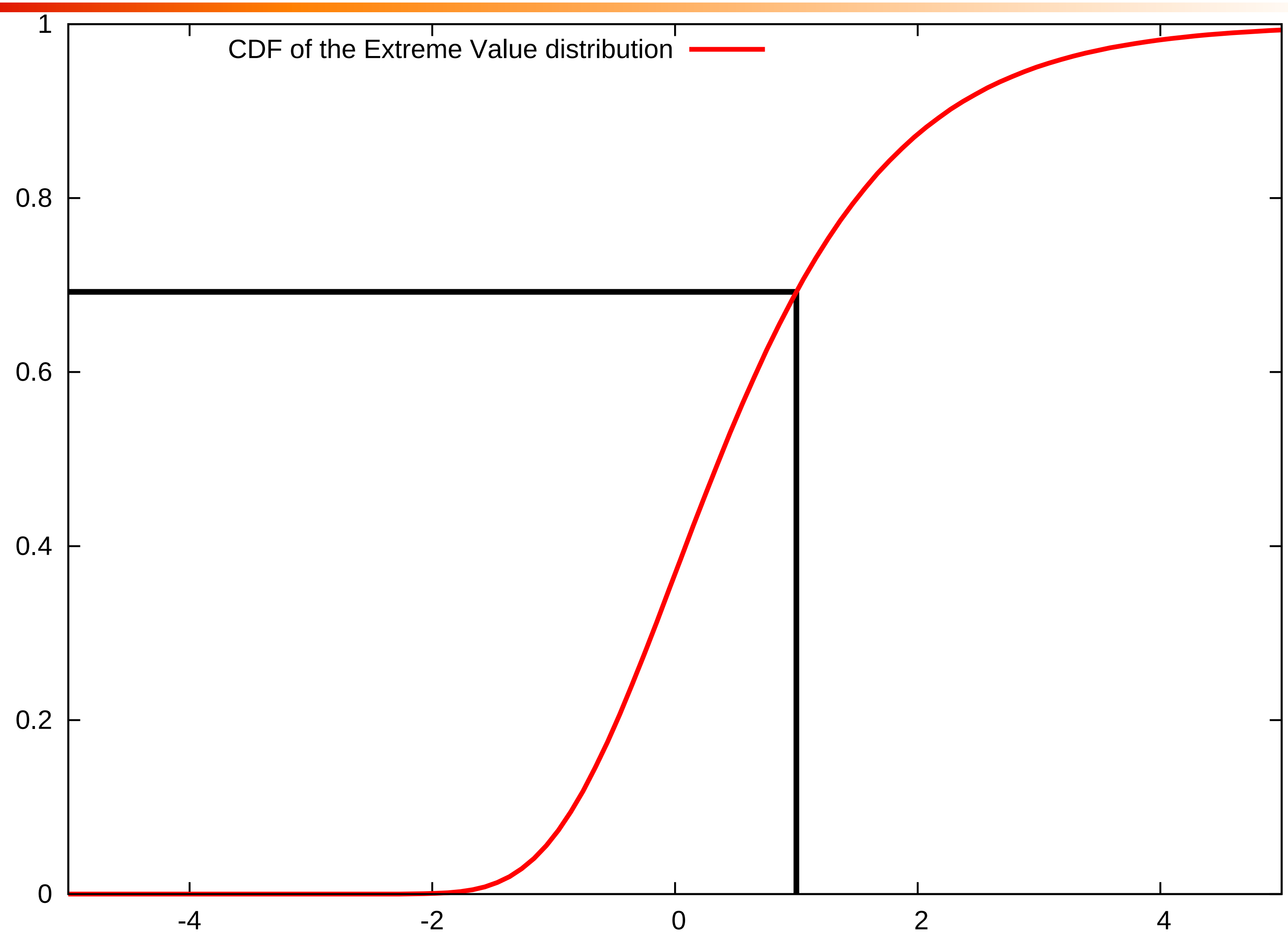
- Consider a univariate r.v. with CDF $F(\varepsilon)$
- If F is invertible and if r is a draw from $U(0,1)$, then

$$s = F^{-1}(r)$$

  is a draw from the given r.v.
- Example: EV with

$$F(\varepsilon) = e^{-e^{-\varepsilon}} \quad F^{-1}(r) = -\ln(-\ln r)$$

Appendix: Simulation: inverse CDF



Appendix: Simulation: multivariate normal

- If r_1,...,r_n are independent draws from $N(0,1)$, and

$$r = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}$$

- then

$$s = a + Lr$$

is a vector of draws from the n-variate normal $N(a, LL^T)$, where

  - L is lower triangular, and
  - LL^T is the Cholesky factorization of the variance-covariance matrix

Appendix: Simulation: multivariate normal

Example:

$$L = \begin{pmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{pmatrix}$$

$$\begin{aligned} s_1 &= \ell_{11} r_1 \\ s_2 &= \ell_{21} r_1 + \ell_{22} r_2 \\ s_3 &= \ell_{31} r_1 + \ell_{32} r_2 + \ell_{33} r_3 \end{aligned}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Appendix: Simulation for mixtures of logit

- In order to approximate

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

- Draw from $f(\xi)$ to obtain $r_1, \ldots, r_R$
- Compute

$$\begin{aligned} P(i) \approx \tilde{P}(i) &= \frac{1}{R} \sum_{k=1}^{R} \Lambda(i|r_k) \\ &= \frac{1}{R} \sum_{k=1}^{R} \frac{e^{V_{1n}+r_k}}{e^{V_{1n}+r_k} + e^{V_{2n}+r_k} + e^{V_{3n}}} \end{aligned}$$

Appendix: Maximum simulated likelihood

$$\max_{\theta} \mathcal{L}(\theta) = \sum_{n=1}^{N} \left(\sum_{j=1}^{J} y_{jn} \ln \tilde{P}(j;\theta) \right)$$

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

Appendix: 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.

Appendix: Maximum simulated likelihood

Properties of MSL:

- If R is fixed, MSL is inconsistent
- If R rises at any rate with N, MSL is consistent
- If R rises faster than $\sqrt{N}$, MSL is asymptotically equivalent to ML.