# *Estimation of mixed generalized extreme value models*

Michel Bierlaire

michel.bierlaire@epfl.ch

Operations Research Group ROSO

Institute of Mathematics

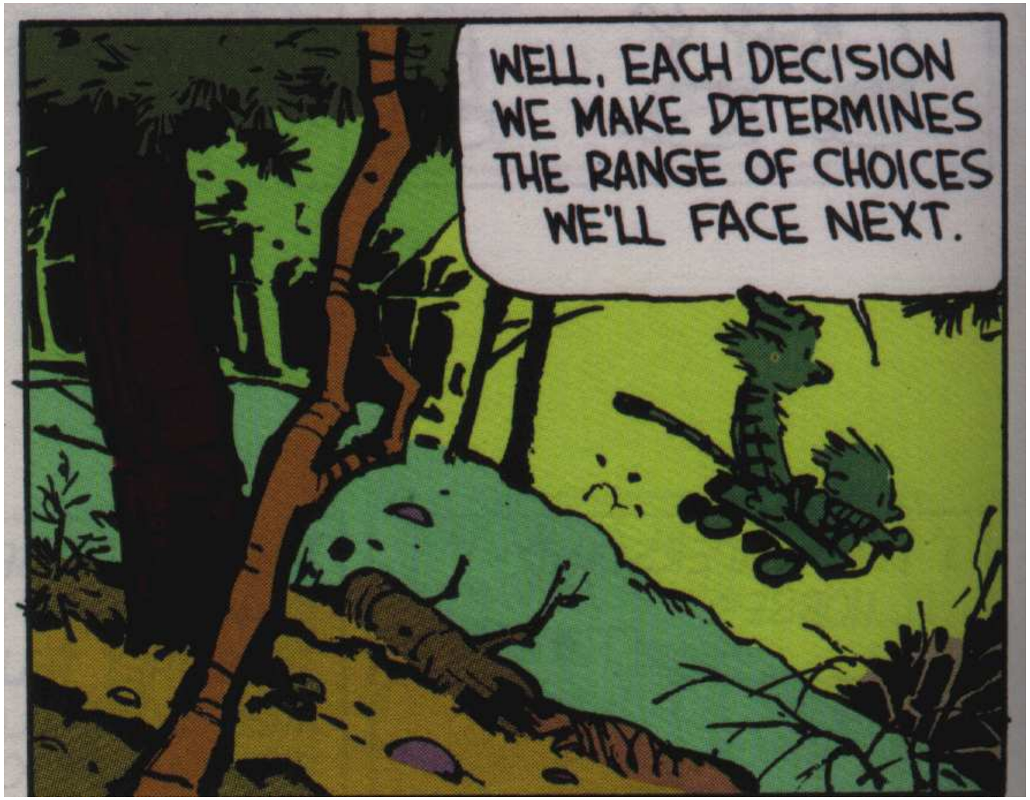
EPFL

# Introduction

*It is our choices that show what we truly are, far more than our abilities*

Albus Dumbledore

# Introduction



# *Introduction*

Nobel Prize 2000 to D. Mc Fadden “for his development of theory and methods for analyzing discrete choice”

# *Introduction*

- Discrete choice models:

$$P(i|\mathcal{C}_n) \text{ where } \mathcal{C}_n = \{1, \ldots, J\}$$

- Random utility models:

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

and

$$P(i|\mathcal{C}_n) = P(U_{in} \geq U_{jn}, j = 1, \ldots, J)$$

- Utility is a latent concept

# *Multinomial Logit Model*

Assumption: $\varepsilon_{in}$ are the maximum of many r.v. capturing unobservable attributes (e.g. mood, experience), measurement and specification errors.

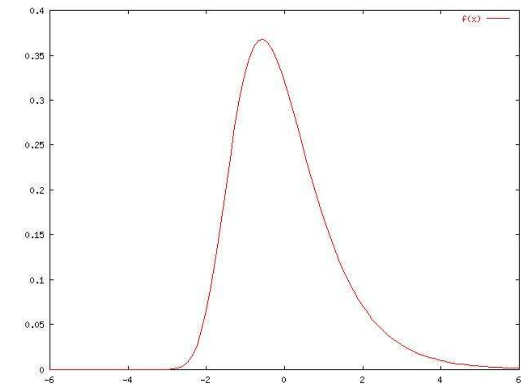
Gumbel theorem: the maximum of many i.i.d. random variables (with a tail) approximately follows a Gumbel distribution.

$$\varepsilon_{in} \sim \text{Gumbel}(0, \mu)$$

# *Multinomial Logit Model*

Gumbel($\eta, \mu$), with $\mu > 0$ :

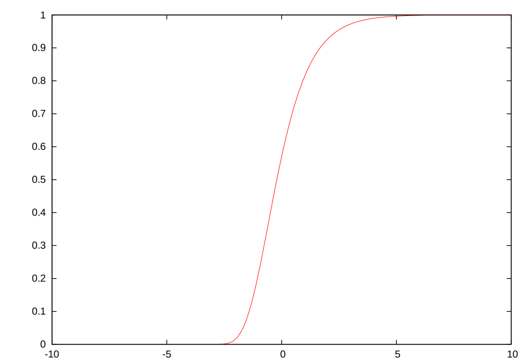
$$f(t) = \mu e^{-\mu(t-\eta)} e^{-e^{-\mu(t-\eta)}}$$




If $\varepsilon \sim$ Gumbel$(\eta, \mu)$, then

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

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



# *Multinomial Logit Model*

If

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

then

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

where

$$\gamma = \lim_{k\to\infty} \sum_{i=1}^{k} \frac{1}{i} - \ln k \approx 0.5772 \quad \text{Euler constant}$$

# Multinomial Logit Model

- The difference of two Gumbel distribution is logistic
- We have

$$P(i|\{i,j\}) = P(V_i + \varepsilon_i \geq V_j + \varepsilon_j) = P(V_i - V_j \geq \varepsilon_j - \varepsilon_i)$$

- We obtain the multinomial logit model

$$P(i|\mathcal{C}_n) = \frac{e^{V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}}}$$

# Multinomial Logit Model

- Multinomial logit model:

$$\varepsilon_{in} \text{ i.i.d. Gumbel}$$

- Gumbel is an Extreme Value distribution
- $\varepsilon_{in}$ is the maximum of many r.v. capturing unobservable attributes, measurement and specification errors.
- Key assumption: Independence

# Relaxing the independence assumption

$$\begin{pmatrix} U_{1n} \\ \vdots \\ U_{Jn} \end{pmatrix} = \begin{pmatrix} V_{1n} \\ \vdots \\ V_{Jn} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1n} \\ \vdots \\ \varepsilon_{Jn} \end{pmatrix}$$

that is

$$U_n = V_n + \varepsilon_n$$

and $\varepsilon_n$ is a vector of random variables.

# *Relaxing the independence assumption*

- $\varepsilon_n \sim N(0, \Sigma)$: multinomial probit model
  - No closed form for the multifold integral
  - Numerical integration is computationally infeasible
- Extensions of multinomial logit model
  - Nested logit model
  - Generalized Extreme Value (GEV) models

# GEV models

Family of models proposed by McFadden (1978)
Idea: a model is generated by a function

$$G : \mathbb{R}^J \to \mathbb{R}$$

From $G$, we can build

- The cumulative distribution function (CDF) of $\varepsilon_n$
- The probability model
- The expected maximum utility

Not equivalent to GEV in statistics

# *GEV models*

1. $G$ is homogeneous of degree $\mu > 0$, that is

$$G(\alpha x) = \alpha^{\mu} G(x)$$

2. $\lim_{x_i \to +\infty} G(x_1, \ldots, x_i, \ldots, x_J) = +\infty$, for each $i = 1, \ldots, J$,

3. the $k$th partial derivative with respect to $k$ distincts $x_i$ is non negative if $k$ is odd and non positive if $k$ is even, i.e., for all (disctincts) indices $i_1, \ldots, i_k \in \{1, \ldots, J\}$, we have

$$(-1)^k \frac{\partial^k G}{\partial x_{i_1} \ldots \partial x_{i_k}}(x) \leq 0, \; \forall x \in \mathbb{R}^J_+.$$

# GEV models

- Density function: $F(\varepsilon_1, \ldots, \varepsilon_J) = e^{-G(e^{-\varepsilon_1}, \ldots, e^{-\varepsilon_J})}$
- Probability: $P(i|C) = \frac{e^{V_i + \ln G_i(e^{V_1}, \ldots, e^{V_J})}}{\sum_{j \in C} e^{V_j + \ln G_j(e^{V_1}, \ldots, e^{V_J})}}$ with $G_i = \frac{\partial G}{\partial x_i}$. This is a closed form
- Expected maximum utility: $V_C = \frac{\ln G(\ldots) + \gamma}{\mu}$ where $\gamma$ is Euler's constant.
- Note: $P(i|C) = \frac{\partial V_C}{\partial V_i}$.

# *GEV models*

Example: $G(e^{V_1}, \ldots, e^{V_J}) = \sum_{i=1}^{J} e^{\mu V_i}$

$$P(i) = \frac{e^{V_i + \ln G_i(e^{V_1}, \ldots, e^{V_J})}}{\sum_{j \in C} e^{V_j + \ln G_j(e^{V_1}, \ldots, e^{V_J})}}$$ with $G_i(x) = \mu x_i^{\mu - 1}$

$$\begin{aligned} e^{V_i + \ln G_i(e^{V_1}, \ldots, e^{V_J})} &= e^{V_i + \ln \mu + (\mu - 1) \ln e^{V_i}} \\ &= e^{\ln \mu + \mu V_i} \end{aligned}$$

$$P(i) = \frac{e^{\ln \mu + \mu V_i}}{\sum_{j \in C} e^{\ln \mu + \mu V_j}} = \frac{e^{\mu V_i}}{\sum_{j \in C} e^{\mu V_j}}$$

Multinomial Logit Model

# *GEV models*

- Multinomial logit model
- Nested logit model
- Cross-nested logit model
- and more...

# *GEV models*

- Closed form probability model
- Provides a great deal of flexibility
- Formulation not in term of correlations
- Require heavy proofs

# Properties of GEV

Let $\mathbb{R}^{J_i}$ be $p$ subspaces spanning $\mathbb{R}^J$. For any vector $y \in \mathbb{R}^J$, $[y]_i$ denotes the projection of $y$ on $\mathbb{R}^{J_i}$. It is assumed that the projection is such that all entries of $[y]_i$ are strictly positive. Let $G^i : \mathbb{R}_+^{J_i} \longrightarrow \mathbb{R}$, $i = 1, \ldots, p$ be $\mu$-GEV functions. Then, the function

$$G : \mathbb{R}_+^J \longrightarrow \mathbb{R} : y \rightsquigarrow G(y) = \sum_{i=1}^{p} \alpha_i G^i([y]_i)$$

is also a $\mu$-GEV function if $\alpha_i > 0$, $i = 1, \ldots, p$.

# Properties of GEV

Let $G : \mathbb{R}_+^J \longrightarrow \mathbb{R}$ be a $\mu$-GEV function. Then $G^\beta$ is a $(\mu\beta)$-GEV function if $0 < \beta \leq 1$.

# Properties of GEV

Let $\mathbb{R}^{J_i}$ be $p$ subspaces spanning $\mathbb{R}^J$. For any vector $y \in \mathbb{R}^J$, denote by $[y]_i$ the projection of $y$ on $\mathbb{R}^{J_i}$. Let $G^i : \mathbb{R}_+^{J_i} \longrightarrow \mathbb{R}$, $i = 1, \ldots, p$ be $\mu_i$-GEV functions. Then, the function

$$G : \mathbb{R}_+^J \longrightarrow \mathbb{R} : y \rightsquigarrow G(y) = \sum_{i=1}^{p} \alpha_i G^i([y]_i)^{\frac{\mu}{\mu_i}}$$

is a $\mu$-GEV function if $\alpha_i > 0$ and $0 < \mu \leq \mu_i$, $i = 1, \ldots, p$.

# Properties of GEV

$$P(i|C) = \frac{e^{V_i + \ln G_i(e^{V_1},\ldots,e^{V_J})}}{\sum_{j\in C} e^{V_j + \ln G_j(e^{V_1},\ldots,e^{V_J})}}$$

Moreover,

$$\alpha_i e^{V_i + \ln G_i(e^{V_1},\ldots,e^{V_J})} = e^{V_i + \ln G_i(e^{V_1 + \ln \alpha_i},\ldots,e^{V_J + \ln \alpha_i})}$$

So,

$$\frac{\alpha_i P(i|\mathcal{C})}{\sum_{j\in B} \alpha_j P(j|\mathcal{C})} = \frac{e^{V_i + \ln G_i(e^{V_1 + \ln \alpha_i},\ldots,e^{V_J + \ln \alpha_i})}}{\sum_{j\in B} e^{V_j + \ln G_j(e^{V_1 + \ln \alpha_j},\ldots,e^{V_J + \ln \alpha_j})}}$$

# Applications

These properties have practical consequences

- Network GEV
- Sampling strategy

# ***Network GEV***

- Extension of the tree representation for Nested Logit
- Investigate new GEV models
- Provide the proof once for all

# *Network GEV*

Let $(V, E)$ be a network with link parameters $\alpha_{(i,j)} \geq 0$

Assumptions:

1. No circuit.
2. One node without predecessor: *root*.
3. $J$ nodes without successor: *alternatives*.
4. For each node $v_i$, there exists at least one path from the root to $v_i$ such that $\prod_{k=1}^{P} \alpha_{(i_{k-1}, i_k)} > 0$.

# Network GEV

For each node $v_i$, we define

- a set of indices $I_i \subseteq \{1,\ldots,J\}$ of $J_i$ relevant alternatives,
- a homogeneous function $G^i : \mathbb{R}^{J_i} \longrightarrow \mathbb{R}$, and
- a parameter $\mu_i$.

Recursive definition of $I_i$:

- $I_i = \{i\}$ for alternatives,
- $I_i = \bigcup_{j \in \text{succ}(i)} I_j$ for all other nodes.

# Network GEV

Recursive definition of $G^i$:

For alternatives:

$$G^i \quad : \quad \mathbb{R} \longrightarrow \mathbb{R} \quad : \quad G^i(x_i) = x_i^{\mu_i} \quad i = 1, \ldots, J$$

For all others:

$$G^i \quad : \quad \mathbb{R}^{J_i} \longrightarrow \mathbb{R} \quad : \quad G^i(x) = \sum_{j \in \textbf{succ}(i)} \alpha_{(i,j)} G^j(x)^{\frac{\mu_i}{\mu_j}}$$

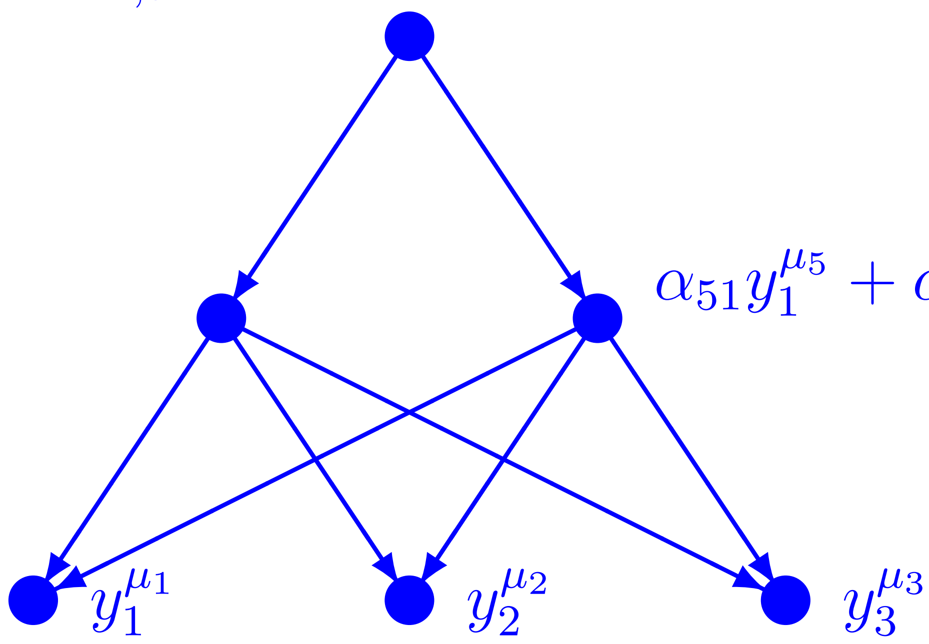
# Network GEV

Example: Cross-Nested Logit

$$G = \sum_m \left( \sum_{j \in C} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$

$$\sum_{i=4,5} \alpha_{0i} \left( \alpha_{i1} y_1^{\mu_i} + \alpha_{i2} y_2^{\mu_i} + \alpha_{i3} y_3^{\mu_i} \right)^{\frac{\mu_0}{\mu_i}}$$

$$\alpha_{51} y_1^{\mu_5} + \alpha_{52} y_2^{\mu_5} + \alpha_{53} y_3^{\mu_5}$$



# *Network GEV*

- Daly & Bierlaire (2003)
- GEV calculus
- Possibility to define new GEV models
- No more proof needed for Network GEV

# Sampling

- Population probability of choice $i \in \mathcal{C}$ and socio-eco char: $P_i(z, \beta^*)p(z)$.
- Probability of being sampled: $R(i, z)$
  - exogeneous sample: $R(i, z) = R(z)$
  - choice-based sample: $R(i, z) = R(i)$
- Sampling of alternative:
  - $A(z) = \{j \in \mathcal{C} | R(j, z) > 0\}$
  - Let's draw $B$, a subset of $A(z)$, with probability $S(B|i, z)$.
  - Analyze choice as if it were limited to $B$.

# *Sampling*

Contribution to the likelihood:

$$P(i|z,B,\beta) = \frac{P_i(z,\beta)R(i,z)S(B|i,z)}{\sum_{j\in B} P_j(z,\beta)R(j,z)S(B|j,z)}$$

If $P_i(z,\beta)$ is given by a GEV model, we obtain

$$P(i|z,B,\beta) = \frac{R(i,z)S(B|i,z)e^{V_i+\ln G_i(e^{V_1},\ldots,e^{V_J})}}{\sum_{j\in B} R(j,z)S(B|j,z)e^{V_j+\ln G_j(e^{V_1},\ldots,e^{V_J})}}$$

Because of the property, let $\alpha(i,z) = \ln R(i,z) + \ln S(B|i,z)$

$$P(i|z,B,\beta) = \frac{e^{V_i+\ln G_i(e^{V_1+\alpha(i,z)},\ldots,e^{V_J+\alpha(i,z)})}}{\sum_{j\in B} e^{V_j+\ln G_j(e^{V_1+\alpha(i,z)},\ldots,e^{V_J+\alpha(i,z)})}}$$

# Sampling

- The model can be estimated as if a pure random sampling strategy was used
- Only the constants are affected.
- Restrictions apply on the sampling of alternatives

# *Mixed GEV*

- GEV models cannot handle all possible correlation structures
- Cannot capture heteroscedasticity and heterogeneity
- Necessity of mixing the model

# Mixed GEV

$$U_n = V_n + \varepsilon_n$$

- $\varepsilon_n$ compliant with GEV theory
- $V_n$ contains random parameters.

$$V_n = \beta^T X_n \text{ where } \beta \sim N(\widehat{\beta}, \Sigma)$$

- Using the Cholesky factorization, we have

$$\beta = \widehat{\beta} + P\zeta \text{ where } \Sigma = PP^T$$

and $\zeta$ are i.i.d. standard normal variates.

# *Mixed GEV*

- McFadden & Train(2000)

  *"Under mild regularity conditions, any discrete choice model derived from random utility maximization has choice probabilities that can be approximated as closely as one pleases by a Mixed MNL model."*

- Why bother with Mixed GEV?

# *Mixed GEV*

- GEV has closed form formulation
- Mixed models require simulated maximum likelihood estimation
- Capture as much as possible of the correlation using GEV
- Use the mixing distribution for the rest
- Issue: estimation

Motivations

- GEV family must be explored
- Complicated implementation
- No appropriate software package
- Most researchers use commercial packages: LIMDEP, ALOGIT, HieLoW or Gauss, Matlab, SAS
- Freeware: Kenneth Train (but based on Gauss)

## Objectives

- Maximum likelihood estimation of a wide variety of GEV models
- Use various nonlinear optimization algorithms
- Open source
- Designed for researchers
- Flexible and easily extensible

# BIOGEME

BIerlaire's Optimization toolbox for GEV Models Estimation

Development :

- Version 0.0: July 2, 2001
- ...
- Version 0.7: December 15, 2003
- Version 0.8: March 19, 2004
- Version 1.0: September 17, 2004

# BIOGEME

Input files

- `mymodel.mod`: model specification
- `sample.dat`: sample data
- `mymodel.par`: general control of the package

Output files

- `mymodel.html` : estimated parameters + statistics
- `mymodel.sta` : sample statistics
- technical and debugging reports

# GEV models

Available in BIOGEME

- Multinomial logit model
- Nested logit model
- Cross-nested logit model
- Network GEV model

# *Heterogeneity*

- GEV models are homoscedastic
- Assume there are two different groups such that

$$\begin{aligned} U_{in_1} &= V_{in_1} + \varepsilon_{in_1} \\ U_{in_2} &= V_{in_2} + \varepsilon_{in_2} \end{aligned}$$

and Var($\varepsilon_{in_2}$) = $\alpha^2$ Var($\varepsilon_{in_1}$)

- Then we prefer the model

$$\begin{aligned} \alpha U_{in_1} &= \alpha V_{in_1} + \alpha\varepsilon_{in_1} \\ U_{in_2} &= V_{in_2} + \varepsilon_{in_2} \end{aligned}$$

# *Heterogeneity*

- If $V_{in_1}$ is linear-in-parameters, that is

$$V_{in_1} = \sum_j \beta_j x_{jin_1}$$

then

$$\alpha V_{in_1} = \sum_j \alpha \beta_j x_{jin_1}$$

is nonlinear.

# *Nonlinear utility funtions*

Other types of nonlinearities

- Box-Cox — Box-Tukey transforms

$$\beta\frac{(x+\alpha)^\lambda - 1}{\lambda},$$

  where $\beta$, $\alpha$ and $\lambda$ must be estimated

- Continuous market segmentation. Example: the cost parameter varies with income

$$\beta_{\text{cost}} = \hat{\beta}_{\text{cost}}\left(\frac{\text{inc}}{\text{inc}_{\text{ref}}}\right)^{\lambda} \text{ with } \lambda = \frac{\partial\beta_{\text{cost}}}{\partial\text{inc}}\frac{\text{inc}}{\beta_{\text{cost}}}$$

# Mixed GEV models

- $V_n$ contains random parameters.

$$V_n = f(\beta_f, \beta_N, \beta_U, X_n)$$

where

$$\begin{array}{rcl} \beta_f & & \text{are deterministic} \\ \beta_N & \sim & N(\widehat{\beta}_N, \Sigma) \\ (\beta_U)_i & \sim & U(a_i, b_i) \end{array}$$

- Because $f$ is nonlinear, other distributions than normal and uniform are possible

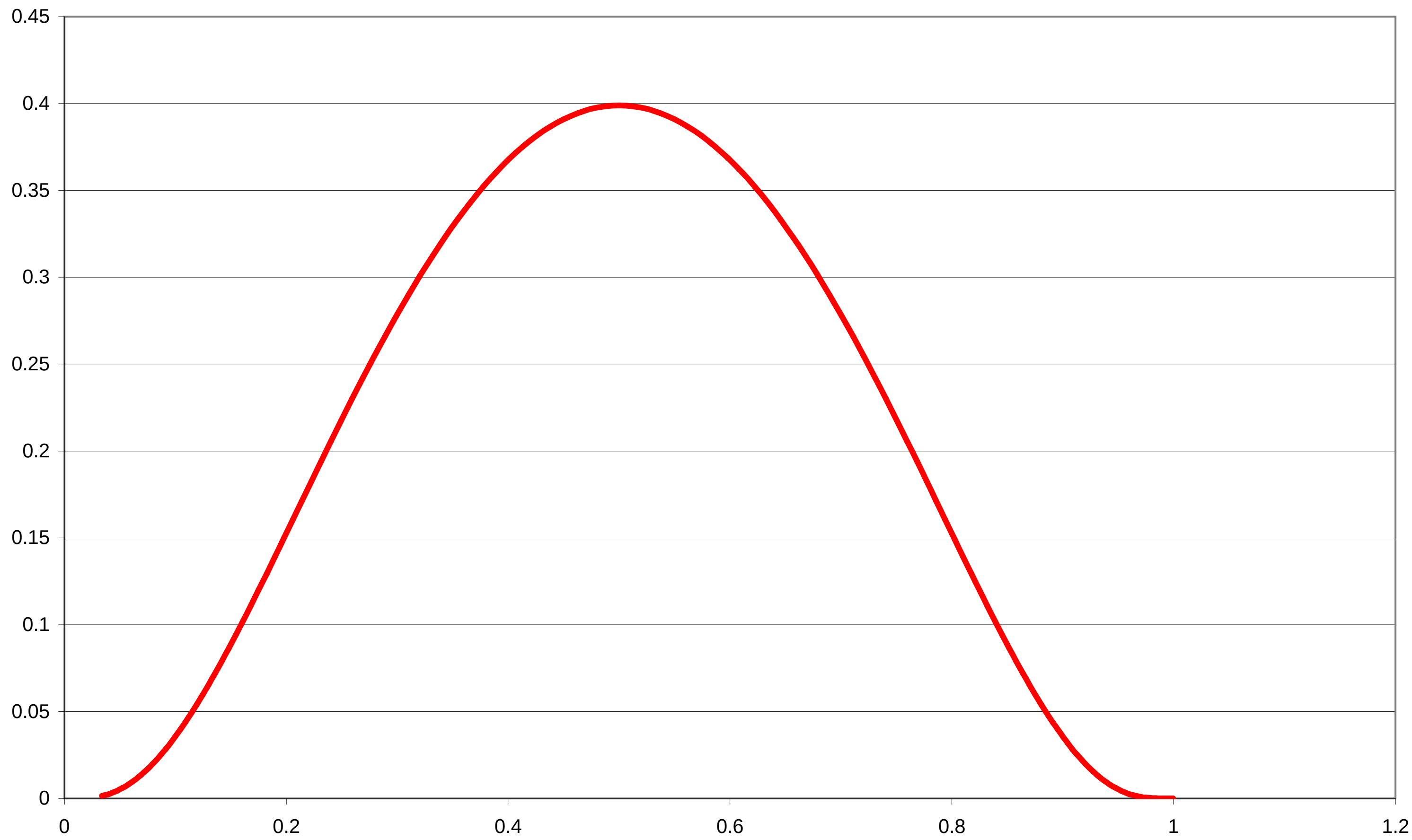
# Mixed GEV models

- Lognormal: if $\beta$ is normal, then $e^{\beta}$ is lognormal
- Triangular: if $\beta_1$ and $\beta_2$ are uniform [0,1], then $\frac{1}{2}(\beta_1 + \beta_2)$ is triangular.
- $S_B$ distribution: if $\beta$ is normal, then

$$\frac{e^{\beta}}{1+e^{\beta}}$$

is a $S_B$ distribution between 0 and 1.

# Mixed GEV models




$S_B$ distribution

# ***Mixed GEV models***

Example: specification with correlated normally and lognormally distributed random coefficients

$$V_{in} = \beta_1 X_{i1n} + \beta_2 X_{i2n}$$

where $\beta_1$ and $\beta_2$ are generated from

$$\begin{pmatrix} \beta_1 \\ \ln \ \beta_2 \end{pmatrix} = \begin{pmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \end{pmatrix} + \begin{bmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

and $\zeta_1$ and $\zeta_2$ are independent $N(0,1)$.

EPFL

# *Panel data*

- Several observations are available for each individual
- Need to capture the individual-specific effects
- At each instance $t$, we have

$$V_{nt} = V(\beta_{nt}, \beta_n, X_{nt})$$

where $\beta_n$ are random parameters constant across $t$ for a given individual $n$.

# ***Miscellaneous features***

- Functionalities can be combined
- Model specification language

```
BETA1 [ BETA1_S ] * x11 + exp( BETA2 [ BETA2_S ] ) * x12
```

- Simulation with Halton draws
- Several optimization packages
- Constrained likelihood estimation

# *Miscellaneous features*

- Robust variance-covariance ("sandwich")
- Output in HTML format

## *And there is more than in BIOGEME...*

**BIOSIM** for forecasting by sample enumeration

**BIOROUTE** for route choice models

**BIOLOOP** to generate large-scale models

**chi2.xls** to perform $\chi^2$ tests

http://roso.epfl.ch/biogeme

EPFL

# *Short course*

Lausanne, March 20-24, 2005

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

http://roso.epfl.ch/DCA