Discrete-continuous maximum likelihood

Discrete Choice Workshop 2017

Anna Fernández Antolín Virginie Lurkin
Meritxell Pacheco Michel Bierlaire

June 22^{nd}, 2017

TRANSP-OR
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

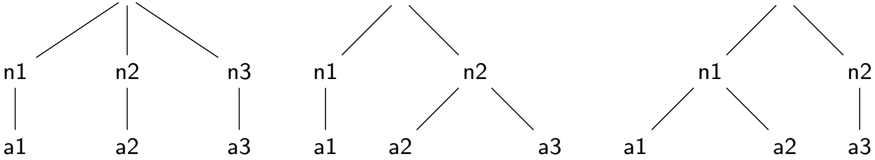
Outline

1. Motivation
2. Mathematical model
3. Case study
4. Conclusions and future work

Discrete choice models

Nested logit

Model unobserved correlations between alternatives.



Nested logit model

Current approach

Try different nesting structures, estimate the parameters of the model and choose the best.

Nested logit model

Current approach

Try different nesting structures, estimate the parameters of the model and choose the best.

Proposed extension

Integrate the selection of the nesting structure in the maximum likelihood framework.

Introducing binary variables to allocate alternatives to nests

$$b_{im} = \begin{cases} 1 & \text{if alternative } i \text{ belongs to nest } m, \\ 0 & \text{otherwise.} \end{cases}$$

$$U_{in} = V_{in} + \sum_{m=1}^{M} b_{im} \left(\tilde{\varepsilon}_{mn} + \frac{1}{\mu_m} (\xi_{imn} - \xi'_{mn}) \right)$$

with

$$V_{in} = f(a_{in}, s_n, \beta), \quad \sum_{m=1}^{M} b_{im} = 1, \quad b_{im} = 0 \text{ if } m > i.$$

Parameters

- s_n: socioeconomic characteristics of individual n.
- a_{in}: attributes of alternative i for individual n.
- y_{in}: availability of alternative i for individual n. It takes value 1 if it is available and 0 otherwise.
- d_{in}: choice of individual n. It takes value 1 if she chooses alternative i and 0 otherwise.
- ε_{mnr}, ξ_{imnr} and ξ'_{mnr}: draws of iid EV(0,1) distribution.

Decision variables

- β: parameter estimates in the utility function.
- μ_m: scale parameter of nest m in the nested logit model.
- b_{im}: it takes value 1 if alternative i belongs to nest m and zero otherwise

Objective function: maximum likelihood

$$\log\left(\prod_{n=1}^{N}\prod_{i=1}^{I} P_n(i)^{d_{in}}\right) = \sum_{n=1}^{N}\sum_{i=1}^{I} d_{in}\log\left(P_n(i)\right),$$

where

$$P_n(i) = \frac{e^{\mu_m V_{in}}}{\sum_{j\in\mathcal{C}_{mn}} e^{\mu_m V_{jn}}} \frac{\left(\sum_{\ell\in\mathcal{C}_{mn}} e^{\mu_m V_{\ell n}}\right)^{\frac{\mu}{\mu_m}}}{\sum_{p=1}^{M}\left(\sum_{\ell\in\mathcal{C}_{pn}} e^{\mu_p V_{\ell n}}\right)^{\frac{\mu}{\mu_p}}}$$

Linearizing the objective function

Sources of non-linearities

1. Expression of the probabilities
2. Logarithm

Linearizing the expression of the probabilities

Work on utilities instead of probabilities Pacheco *et al* (forthcoming)[1]

- Use simulation for the random terms of the utility function:

$$U_{inr} = V_{in} + \sum_{m=1}^{M} b_{im} \left(\tilde{\varepsilon}_{mnr} + \frac{1}{\mu_m} (\xi_{imnr} - \xi'_{mnr}) \right)$$

- Introduce binary variables w_{inr} such that

$$w_{inr} = 1 \iff U_{inr} > U_{jnr}, \quad \forall i \neq j$$

Linear expression of the probability

$$P_n(i) = \frac{1}{R} \sum_{r=1}^{R} w_{inr}$$

[1] Pacheco, Meritxell, Azadeh, Shadi Sharif, Bierlaire, Michel, & Gendron, Bernard. forthcoming. Integrating advanced discrete choice models in mixed integer linear programming. Tech. rept. Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Linearizing the logarithm

Expression of the objective function

$$\sum_{n=1}^{N} \sum_{i=1}^{I} d_{in} \left(\log \left(\sum_{r=1}^{R} w_{inr} \right) - \log R \right)$$

Introduce binary variables Ω_{inp} such that

$$\text{If} \quad \sum_{r=1}^{R} w_{inr} = p \quad \text{then} \quad \Omega_{inq} = 1, \quad \forall q \leq p, \forall p \in \{0, ..., R\}.$$

Ω_{inr} are a permutation of w_{inr} such that they are ordered.

Linearizing the logarithm

Then

$$\log\left(\sum_{r=1}^{R} w_{inr}\right) = \sum_{p=1}^{R} \Omega_{inp} \Delta L_p,$$

where ΔL_p can be preprocessed as follows:

- $\Delta L_p = \log(p) - \log(p-1), \quad p = 2, ..., R$
- $\Delta L_1 = 0$

Linear objective function

$$\sum_{n=1}^{N} \sum_{i=1}^{I} d_{in} \left(\sum_{p=1}^{R} \Omega_{inp} \Delta L_p - \log R \right)$$

Linearization of constraints

If $\sum_{r=1}^{R} w_{inr} = p$ then $\Omega_{inq} = 1, \quad \forall q \leq p$

$$(R - p + 1)\Omega_{inp} - 1 \geq \sum_{r=1}^{R} w_{inr} - p \qquad \forall i, n, p \tag{1}$$

$$\sum_{r=1}^{R} \Omega_{inr} = \sum_{r=1}^{R} w_{inr} \qquad \forall i, n \tag{2}$$

Linearization of constraints

$$U_{inr} = V_{in} + \sum_{m=1}^{M} b_{im}\tilde{\varepsilon}_{mnr} + \sum_{m=1}^{M} b_{im}\frac{1}{\mu_m}(\xi_{imnr} - \xi'_{mnr}),\ \bar{\mu}_m = \frac{1}{\mu_m}$$

$$\tau_{im} \leq b_{im} \qquad \forall i, m \tag{3}$$

$$\tau_{im} \leq \bar{\mu}_m \qquad \forall i, m \tag{4}$$

$$\tau_{im} \geq \bar{\mu}_m + b_{im} - 1 \qquad \forall i, m \tag{5}$$

$$\tau_{im} > 0 \qquad \forall i, m \tag{6}$$

$$U_{inr} = V_{in} + \sum_{m=1}^{M} b_{im}\tilde{\varepsilon}_{mnr} + \sum_{m=1}^{M} \tau_{im}(\xi_{imnr} - \xi'_{mnr}) \tag{7}$$

Constraints

Each alternative belongs to exactly one nest

$$\sum_{m=1}^{M} b_{im} = 1, \quad \forall i \tag{8}$$

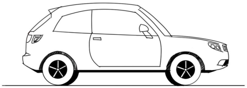
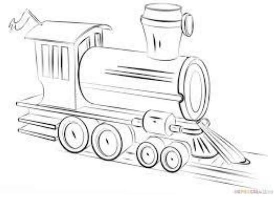
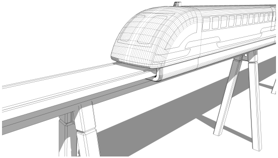
Symmetry breaking constraints

$$b_{im} = 0 \quad \text{if} \quad m > i, \quad \forall i \tag{9}$$

Other constraints

- Each customer chooses one alternative
- An alternative can only be chosen by an individual if it is available to her
- The chosen alternative is the one with highest utility
- Normalization of $\bar{\mu}_m$

Swissmetro: 1998

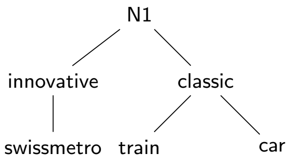
- Analyze the impact of the Swissmetro (mag-lev underground system) compared to car and train.
- Stated preference data

Comparison of log likelihoods: Logit

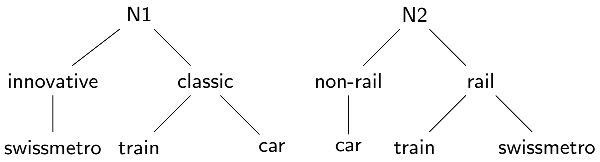
Number of observations: 20
Number of draws MILP: 200
Solving time MILP $\sim$ 3 hours

	Values parameters		Final log likelihood	
	β_{time}	β_{cost}	MILP	Biogeme
Optimal MILP	-0.0126	-0.0530	-19.13	-19.30
Optimal Biogeme	-0.0946	-0.0335	-19.32	-19.27

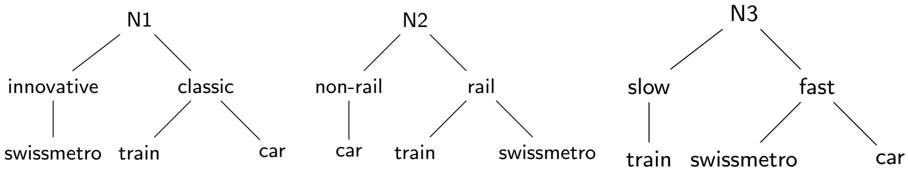
Possible nesting structures



Possible nesting structures



Possible nesting structures



Results

Number of observations: 200

Parameter	N1 Value	N1 *t*-test	N2 Value	N2 *t*-test	N3 Value	N3 *t*-test
ASC_{CAR}	-0.0287	0.63	-1.32	-3.24	-0.545	-2.07
ASC_{SM}	0.574	3.61	0.182	0.79	0.778	3.35
β_{COST}	-0.0581	0.63	-0.261	-1.12	-0.376	-1.67
β_{TIME}	-0.0853	0.62	-0.309	-1.8 6	-0.364	-1.67
$\bar{\mu}_{CLASSIC}$	0.04	0.48	-	-	-	-
$\bar{\mu}_{RAIL}$	-	-	0.36	0.87	-	-
$\bar{\mu}_{FAST}$	-	-	-	-	1.00	0.00
$\mathcal{L}(\hat{\beta})$: Biogeme	-155.77		-161.39		-162.23	
$\mathcal{L}(\hat{\beta})$: MILP (100 draws)	-150.69		-164.43		-170.22	

Conclusions

- New framework to include binary variables to maximum likelihood.
- Convert the problem in an MILP (linearization of the log likelihood).
- Case study: proof of concept.

Future work

- Make the model faster (remove big M constraints)
- Include constraints that insure identification
- Include more symmetry breaking constraints
- Calculate standard errors of the estimated parameters

Thank you for your attention!
Questions?
Suggestions?

Complete model: Logit

$$
\begin{aligned}
\max \quad & \sum_{n=0}^{N-1}\sum_{i=0}^{I-1} d_{in}\left(\sum_{r=0}^{R-1}\Omega_{inr}\Delta L_p - \log R\right) & & \\
\text{subject to} \quad & U_{inr} = f(a_{in}, s_{in}, \beta) + \tilde{\varepsilon_{inr}} & \forall i \in \mathcal{C}_n, n, r \quad & (10)\\
& U_{inr} \leq U_{nr} & \forall i \in \mathcal{C}_n, n, r \quad & (11)\\
& U_{nr} \leq U_{inr} + M_{nr}(1 - w_{inr}) & \forall i \in \mathcal{C}_n, n, r \quad & (12)\\
& w_{inr} \leq y_{in} & \forall i, n, r \quad & (13)\\
& \sum_{i=0}^{I-1} w_{inr} = 1 & \forall n, r \quad & (14)\\
& (R - p + 1)\Omega_{in(p-1)} - 1 \geq \sum_{r=0}^{R-1} w_{inr} - p & \forall i, n, p \quad & (15)\\
& \sum_{r=0}^{R-1}\Omega_{inr} = \sum_{r=0}^{R-1} w_{inr} & \forall i, n \quad & (16)\\
& w_{inr}, \Omega_{inr} \in \{0, 1\} & \forall i, n, r \quad & (17)\\
& \beta \in R^p & & (18)
\end{aligned}
$$

Complete model: Logit

- (10) are the utility functions.
- (11)-(12) sets w_{inr} to one if U_{inr} is largest utility for individual n in scenario r.
- (13)-(14) impose that only one alternative can be chosen and that it has to be an available one.
- (15)-(16) are used to define Ω.