# Discrete-continuous maximum likelihood for the estimation of error component logit models

Anna Fernández Antolín[1], Virginie Lurkin[1], Matthieu de Lapparent[1] and Michel Bierlaire[1]

[1]Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne

## 1 Introduction

In this paper we aim at integrating the selection of a nesting structure to the maximum likelihood framework of the parameter estimation. Given a finite set of nesting structures, the traditional approach is to estimate them and select a posteriori the most appropriate one based on some fit statistics and informal testing procedures. However, the number of possible nested structures grows as a function of the number of alternatives.

Our approach simultaneously solves the problem of selecting the optimal nesting structure and estimating its corresponding parameters with maximum likelihood. We call this *discrete-continuous maximum likelihood* (DCML). We are able to linearize the logarithm in the objective function so that it results in a mixed integer linear problem. We show that this approach is computationally feasible using an example with stated preference data with three alternatives.

In order to model the nesting structure, we use a mixed logit model as follows:

$$U_{in} = V_{in} + \sum_{m=1}^{M} \sigma_m b_{im} \xi_{im} + \nu_{in}, \tag{1}$$

where $V_{in}$ is the deterministic part of the utility function for alternative $i$ and individual $n$. $V_{in}$ is a linear-in-parameters function of a vector of parameters to be estimated ($\beta$), observed attributes of the alternatives ($a_{in}$) and socioeconomic characteristics of the individual ($s_n$), $V_{in} = f(a_{in}, s_n, \beta)$. $\xi_{im} \overset{iid}{\sim} N(0,1)$ and $\nu_{in} \overset{iid}{\sim}$ Gumbel$(0,1)$. $b_{im}$ are binary variables that indicate if an alternative $i$ belongs to a nest $m$.

## 2 Mathematical model

**Objective function** In order to apply maximum likelihood, we want to maximize the following function

$$\log\left(\prod_{n=1}^{N}\prod_{i=1}^{I} P_n(i)^{d_{in}}\right), \tag{2}$$

where $P_i(n)$ is the probability that individual $n$ chooses alternative $i$, and $d_{in}$ takes value 1 if individual $n$ chooses alternative $i$ and 0 otherwise. $N$ is the number of individuals and $I$ is the number of alternatives. Using the framework developed by Pacheco *et al.* (forthcoming) we can use simulation of the error terms and avoid the non-linearities caused by the expression of the probabilities. This is done by working with the values of the utility functions instead of with the probabilities. Then the objective function becomes

$$\sum_{n=1}^{N} \sum_{i=1}^{I} d_{in} \log \left( \frac{1}{R} \sum_{r=1}^{R} w_{inr} \right), \tag{3}$$

where $r$ is an index that stands for the draw and $R$ is the number of draws or scenarios. $w_{inr}$ takes value 1 if $U_{inr} > U_{jnr}, \quad \forall j \neq i$ and 0 otherwise.

The only remaining non-linearity is the logarithm that appears in the objective function. Since $\sum_{r=1}^{R} w_{inr}$ can only take integer values from 1 to R, we can linearize it by introducing binary variables $\gamma_{inp}$ that take value 1 if $\sum_{r=1}^{R} w_{inr} = p$ and 0 otherwise. Then, Equation (3) is equivalent to

$$\sum_{n=1}^{N} \sum_{i=1}^{I} d_{in} \sum_{p=1}^{R} \gamma_{inp} L_p, \tag{4}$$

where $L_p = \log(p)$, $p = 1, ..., R$ is a pre-processed vector of $R$ components.

The discrete continuous maximum likelihood problem can be formalized as follows:

$$\max \quad \sum_{n=1}^{N}\sum_{i=1}^{I} d_{in} \sum_{p=1}^{R} \gamma_{inp} L_p$$

$$\text{subject to} \quad U_{inr} = f(a_{in}, s_{in}, \beta) + \sum_{m=1}^{M} \tau_{im}\xi_{imr} + \nu_{inr} \quad \forall i, n, r \quad (5)$$

$$\tau_{im} \leq u_m b_{im} \quad \forall i, m \quad (6)$$

$$\tau_{im} \leq \sigma_m \quad \forall i, m \quad (7)$$

$$\tau_{im} \geq \sigma_m - u(1 - b_{im}) \quad \forall i, m \quad (8)$$

$$l_{nr} \leq z_{inr} \quad \forall i, n, r \quad (9)$$

$$z_{inr} \leq l_{nr} + M_{inr} y_{in} \quad \forall i, n, r \quad (10)$$

$$U_{inr} - M_{inr}(1 - y_{in}) \leq z_{inr} \quad \forall i, n, r \quad (11)$$

$$z_{inr} \leq U_{inr} \quad \forall i, n, r \quad (12)$$

$$\sum_{i=1}^{I} w_{inr} = 1 \quad \forall n, r \quad (13)$$

$$w_{inr} \leq y_{in} \quad \forall i, n, r \quad (14)$$

$$z_{inr} \leq U_{nr} \quad \forall i, n, r \quad (15)$$

$$U_{nr} \leq z_{inr} + M_{nr}(1 - w_{inr}) \quad \forall i, n, r \quad (16)$$

$$(R+1)\delta^1_{inr} - 1 \geq \sum_{r=1}^{R} w_{inr} - p \quad \forall i, n, p \quad (17)$$

$$(R+1)\delta^2_{inr} - 1 \geq p - \sum_{r=1}^{R} w_{inr} \quad \forall i, n, p \quad (18)$$

$$\delta^1_{inr} + \delta^2_{inr} - 2\gamma_{inp} \leq 1 \quad (19)$$

$$\sum_{p=1}^{R} \gamma_{inp} = 1 \quad \forall i, n \quad (20)$$

$$\sum_{m=1}^{M} b_{im} = 1 \quad \forall i \quad (21)$$

$$b_{im} = 0 \quad \forall m > i \quad (22)$$

$$b_{im} \in \{0, 1\} \quad \forall i, m \quad (23)$$

$$w_{inr}, \delta^1_{inr}, \delta^2_{inr} \in \{0, 1\} \quad \forall i, n, r \quad (24)$$

$$\gamma_{inp} \in \{0, 1\} \quad \forall i, n, p \quad (25)$$

$$\sigma_m \in R^+ \quad \forall m \quad (26)$$

$$\tau_{im} \in R^+ \quad \forall i, m \quad (27)$$

$$\beta \in R \quad (28)$$

Where the parameters of the model are:

- $s_n$: the socioeconomic characteristics of individual $n$.

- $a_{in}$: the attributes of alternative $i$ for individual $n$.
- $y_{in}$: the availability of alternative $i$ for individual $n$. It takes value 1 if it is available and 0 otherwise.
- $d_{in}$: the choice of individual $n$. It takes value 1 if she chose alternative $i$ and 0 otherwise.
- $\xi_{imr}$ and $\nu_{inr}$: draws of the respective distributions.

The decision variables of the model are:

- $\beta$: the parameter estimates in the utility function.
- $\sigma_m$: it represents the variance of nest $m$ in the error component logit model.
- $w_{inr}$: it takes value 1 if alternative $i$ is chosen by individual $n$ in scenario $r$.
- $b_{im}$: it takes value 1 if alternative $i$ belongs to nest $m$ and zero otherwise

And $M_{inr}$, $M_{nr}$, $l_{nr}$ and $u_m$ are upper and lower bounds to be defined.

**Constraints** Constraints (5) are the utility functions, as explained in Equation (1), where $r$ stands for the draw number and the product $b_{im}\sigma_m$ has been linearized using $\tau_{im}$. Constraints (6)-(8) are used to linearize $b_{im}\sigma_m$. Constraints (9)-(12) are equivalent to $z_{inr} = U_{inr}$ if $y_{in} = 1$, $z_{inr} = l_{nr}$ if $y_{in} = 0$, which sets the utility of a given alternative to a lower bound if the alternative is not available. Constraints (13) express the fact that each customer chooses one alternative. Constraints (14) say that only available alternatives can be selected by individuals. Constraints (15)-(16) are equivalent to $U_{nr} = \max_i z_{inr}$, which means that the chosen alternative is the one with highest utility. Constraints (17)-(20) are equivalent to $\gamma_{inp} = 1 \iff \sum_{r=1}^{R} w_{inr} = p$ and are used to linearize the objective function. Constraints (21) say that each alternative belongs to exactly one nest. Constraints (22) are used as a symmetry-breaking constraints, so that the binary variables are well defined. Finally, constraints (23) to (28) define the space of solutions.

To show that the approach is computationally feasible, we use a stated preferences mode choice case study collected in Switzerland in 1998. The respondents provided information in order to analyze the impact of the model innovation in transportation represented by the Swissmetro, a mag-lev underground system, compared to the usual transport modes of car and train.

# References

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.