A MILP formulation for the maximum likelihood estimation of continuous and discrete parameters in choice models

Virginie Lurkin, Anna Fernández Antolín, Michel Bierlaire

Transport and Mobility Laboratory (TRANSP-OR),
School of architecture, Civil and Environmental Engineering (ENAC),
École polytechnique fédérale de Lausanne (EPFL)
Email: **virginie.lurkin@epfl.ch**

1 Motivation and contributions

The state-of-the-art for the mathematical modeling of disaggregate demand relies on choice theory. In the estimation of discrete choice models, in general, only continuous parameters are considered, although advanced models include also discrete ones. The most typical example of a discrete parameter that is usually disregarded from the estimation process is the nest allocation parameter in nested logit models. Nesting structures are used in discrete choice models when correlation between alternatives is suspected. They are used in a very broad range of transportation contexts such as airline itinerary choice, car-type choice, route choice, and in mode choice among others. In many contexts, the partition of alternatives into different nests is not obvious and there are several nesting structures that make intuitive sense. In practice, to determine the most appropriate nesting structure, the analyst has two options: (i) to enumerate all the possible values, and estimate the continuous parameters for each combination, and (ii) to make the problem continuous by relaxing the integrality of the discrete parameters. For instance, a membership indicator becomes a continuous variable between 0 and 1 (like in the cross-nested logit model), or by making the membership probabilistic (like in latent class models). In both cases, however, the likelihood function features several local optima, so that classical nonlinear optimization methods may not find the (global) maximum likelihood estimates. Other discrete parameters can be introduced in choice models to replace the discrete decisions that are currently taken on a trial-and-error basis by the modeler, such as the choice of variables that enter the utility specification.

In this work, we propose a new mathematical model designed to find the global maximum likelihood estimates of a choice model involving both discrete and continuous parameters. This is a first attempt towards a MILP formulation of the maximum log likelihood, which results in a problem with high computational complexity. The goal of this presentation is to show under which circumstances our approach is computationally feasible, and to study its strengths and limitations. Our contributions are multiple. First, to the best of our knowledge, we are the first to include discrete parameters estimation in the maximum likelihood framework in the context of discrete choice models. Second, our model is formulated as an MILP. We use simulations and piecewise linear function approximation to dispose of the non-linearity of the log likelihood. We believe that it is the first time that the log likelihood is linearized. Finally, the proposed model is general and can be used with multiple choice models. Our framework is illustrated on the nested logit model in the next section[1].

[1]The mathematical model presented in the next section is for the nested logit model, but the framework remains valid

2 Mathematical model

Following Ben-Akiva and Lerman (1985), the utility associated with each alternative i that belongs to nest m is expressed for individual n as

$$U_{in} = V_{in} + \varepsilon_{mn} + \varepsilon_{imn}, \tag{1}$$

where

- ε_{mn} is such that $\widetilde{\varepsilon}_{mn} = \varepsilon_{mn} + \varepsilon'_{mn}$, with $\widetilde{\varepsilon}_{mn} \overset{iid}{\sim} EV(0,1)$, and $\varepsilon'_{mn} \overset{iid}{\sim} EV(0,\mu_m)$,
- $\varepsilon_{imn} \overset{iid}{\sim} EV(0,\mu_m)$, with $\mu_m \geq 1$.

The utility of Equation (1) can therefore be rewritten as

$$U_{in} = V_{in} + \widetilde{\varepsilon}_{mn} + (\varepsilon_{imn} - \varepsilon'_{mn}), \tag{2}$$

or equivalently as

$$U_{in} = V_{in} + \widetilde{\varepsilon}_{mn} + \frac{1}{\mu_m}(\xi_{imn} - \xi'_{mn}), \tag{3}$$

where $\xi_{imn} \overset{iid}{\sim} EV(0,1)$, and $\xi'_{mn} \overset{iid}{\sim} EV(0,1)$.

As we don't know *a priori* if alternative i belongs to nest m, we introduce the indicator parameters b_{im}, that is equal to 1 if alternative i belongs to nest m, and 0 otherwise. Then, the utility becomes

$$U_{in} = V_{in} + \sum_{m=1}^{M} b_{im}\left(\widetilde{\varepsilon}_{mn} + \frac{1}{\mu_m}(\xi_{imn} - \xi'_{mn})\right), \tag{4}$$

$$= V_{in} + \sum_{m=1}^{M} b_{im}\widetilde{\varepsilon}_{mn} + \sum_{m=1}^{M}\left(\frac{b_{im}}{\mu_m}(\xi_{imn} - \xi'_{mn})\right) \tag{5}$$

Equation (5) is then easily linearized with traditional linearization techniques.

We use maximum likelihood estimation to find the values of our discrete and continuous parameters. However, in order to have a linear formulation, we approximate the choice probability with simulation. That is we maximize

$$\sum_{n=1}^{N}\sum_{i=1}^{I} d_{in} \log\left(\widehat{P}_n(i)\right), \tag{6}$$

where

- d_{in} is observed and takes value 1 if individual n chooses alternative i, and 0 otherwise,
- $\widehat{P}_n(i)$ is the approximate probability that individual n chooses alternative i.

---

for any DCM for which draws can be generated.

To obtain the approximate probability, $\widehat{P}_n(i)$, we follow the framework developed by Pacheco et al. (2017) and we generate R draws for ξ_{imn} and ξ'_{mn} in Equation (5). Then, the choice of an individual n in a particular scenario r[2] is characterized by the binary parameter w_{inr}, that is equal to 1 if $U_{inr} > U_{jnr}$, and 0 otherwise. We then use additional constraints, together with linearization techniques to ensure that w_{inr} is fixed to one for the alternative with the highest utility and to zero otherwise.

This formulation allows to approximate the probability of individual n to choose alternative i as

$$\widehat{P}_n(i) = \frac{1}{R} \sum_{r=1}^{R} w_{inr}, \tag{7}$$

and the objective function becomes

$$\sum_{n=1}^{N} \sum_{i=1}^{I} d_{in} \left(\log \left(\sum_{r=1}^{R} w_{inr} \right) - \log(R) \right). \tag{8}$$

The remaining non-linearity is the logarithm and we use a piecewise linear function to evaluate it. In order to do so, we define $s_{in} = \sum_{r=1}^{R} w_{inr}$ and $z_{in} = \log(s_{in})$. We denote by $PL^p(s_{in})$ the line that passes through points $(p-1, \log(p-1))$ and $(p, \log(p)), \forall p = 1, ..., R$, then

$$PL^p(s_{in}) = \log(p)(s_{in} - (p-1)) + \log(p-1)(p - s_{in}), \forall p = 1, ..., R. \tag{9}$$

Since $\log(0)$ is not defined, and $\lim_{x \to 0} \log(x) = -\infty$, we approximate $-\infty$ by a negative enough number and denote it L_0[3].

Therefore, the maximization of Equation (8) is equivalent to

$$\max \quad \sum_{n=1}^{N} \sum_{i=1}^{I} d_{in} \left(z_{in} - \log R \right) \tag{10}$$

$$\text{s.t.} \quad s_{in} = \sum_{r=1}^{R} w_{inr} \tag{11}$$

$$z_{in} \leq PL^p(s_{in}) \tag{12}$$

Finally, our model contains additional constraints that we only define informally: (1) each customer chooses one alternative, (2) an alternative can only be chosen by an individual if it is available to her, (3) the chosen alternative is the one with highest utility, (4) the scale parameter $\bar{\mu}_m$ is normalized, (5) each alternative belongs to exactly one nest, (6) symmetry is breaking for the nest allocation.

[2] We use the term scenario to describe the realization of a draw.

[3] In practice we consider L_0 = -100.

3 Results

In order to determine the minimum number of draws needed to obtain reliable values of the final log likelihood, we evaluate Equation (10) at the values of the parameters obtained by a continuous estimation software (Biogeme). We do so for the logit model, and for three possible nested logit models (N1,N2 and N3). The results are shown in Table 1, together with the value of the final log likelihood (FLL) obtained with the continuous estimation. The table also shows the relative error between the real FLL and the value obtained with the MILP (using R draws). The relative error is computed as

$$e^R_{FLL} = \frac{FLL - \widehat{FLL}_R}{FLL}.$$

		N1		N2		N3		Logit	
	R	FLL	e_{FLL} [%]	FLL	e_{FLL} [%]	FLL	e_{FLL} [%]	FLL	e_{FLL} [%]
MILP	5	-1648	958	-1560	866	-1558	860	-1344	729
	10	-358.1	130	-678.2	320	-369.8	128	-657.9	306
	20	-152.8	1.93	-180.9	12.1	-172.4	6.28	-160.1	1.29
	50	-153.7	1.32	-169.1	4.78	-171.2	5.54	-159.3	1.79
	100	-154.0	1.12	-168.6	4.46	-170.8	5.31	-161.0	0.757
Continuous estimation	-	-155.8	-	-161.4	-	-162.2	-	-162.2	-

Table 1: Investigating the simulation error

As expected, we see that the relative error decreases with the number of draws. In Table 2, we report the values of the estimated continuous parameters, the FLL, its relative error, the optimal nesting structure, the solution time and the solution gap (in the cases in which a time limit of 48 hours is reached). The last row also shows the values obtained with Biogeme. We can see that thanks to our MILP framework, we can find the values of parameters, together with the best nesting structure (N1), independently of the number of draws used.

	R	β_{TIME}	β_{COST}	$\bar{\mu}$	FLL	e_{FLL} [%]	NS	time [s]	gap [%]
MILP	5	-0.315	-0.405	0.150	-1053	536	N1	3682	0
	10	-0.375	-0.359	0.229	-165.3	0.165	N1	14188	0
	20	-0.291	-0.265	0.135	-155.9	5.84	N1	172808	17
Continuous estimation	-	-0.301	-0.119	0.14	-165.5	-	N1	-	-

Table 2: Finding the best nesting structure

More advanced results and analysis will be presented as part of the presentation.

4 References

- Ben-Akiva, M. and Lerman, S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press series in transportation studies, MIT Press.
- Pacheco, M., Azadeh, S. S., Bierlaire, M. and Gendron, B. (2017). *Integrating advanced discrete choice models in mixed integer linear optimization.* Technical Report TRANSP-OR 170714, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale