Estimating mixed logit models with quasi-Monte Carlo sequences allowing practical error estimation.

Fabian Bastin[1], Cinzia Cirillo and Philippe L. Toint

FUNDP – GRT
Facultés Universitaires Notre-Dame de la Paix
Groupe de Recherche sur les Transports
8, Rempart de la Vierge 5000 Namur BELGIUM
tél : ++32.81.72.49.23, fax : ++32.81.72.49.14,
email: ccir@math.fundp.ac.be

Abstract

Mixed Multinomial Logit Models (MMNL) are now a popular and efficient framework in discrete choice theory. However, it is well known that the numerical cost associated to the evaluation of multidimensional integrals in MMNL models remains high even if Monte Carlo (MC) or quasi-Monte Carlo (QMC) techniques are used instead of classical quadrature methods, while no analytical solution can be found. Our current approach, developed in the context of modern trust-region optimization techniques at FUNDP (Facultés Universitaires Notre Dame de la Paix), uses statistical inference of Monte-Carlo approximations to speed up computations. We have shown that numerical efficiency is considerably increased by the exploitation of new results on the accuracy and bias estimates relative to the objective function. The crucial ingredient of our algorithm is that, at each iteration, we are able to use only a subset of the random draws, whose size is adapted from iteration to iteration. Convergence of the algorithm has been also demonstrated, towards points satisfying first- and second-order optimality conditions (Bastin et *al.*, 2004b). The methodolgy has been successfully applied to both simulated and real data sets. The results, even on large-scale model estimation, show that the proposed optimization algorithm is competitive with existing tools, including softwares based on quasi-Monte Carlo techniques using Halton sequences.

In this paper, we propose to extend our study and to compare our variable sample size Monte Carlo algorithm with randomized quasi-Monte Carlo sequences. We use Sobol sequences, that are expected to perform better than Halton ones, as suggested by Garrido (2003). There are different ways to randomize quasi-random sequences; some of them have been already explored by the transportation community. Bhat (2003) has suggested that scrambled Halton sequences avoid the problem of poor coverage of the integration domain in high dimensions, and has used random shifts to evaluate the quality of the sequences in the context of MMNL estimation. Hess et *al.* (2003) have proposed the use of randomly shifted and shuffled uniform vectors and have reported better performances. Garrido (2003) has also proposed to use Owen scrambling technique for Sobol sequence.

Since the sequences used in QMC approaches are deterministic, it is not possible to use the classical analytical tools for error estimation as we have done in the MC variable sample size strategy by using the delta method. It is, therefore, desirable to develop techniques, which combine the potential higher accuracy of QMC approximation with the practical error estimation ability of MC methods. By introducing some randomness in low discrepancy sequences, one can use statistical methods for error analysis. Our objective is to investigate how those techniques can be applied in the mixed logit model estimation. In particular, we are interested in seeing how

[1] Research Fellow of the Belgian National Fund for Scientific Research.

randomized QMC sequences can reduce the variance in comparison with MC methods and how they can improve the performance of the original deterministic sequences, in combination with the variable sample size strategy.

We will apply the methodologies on both simulated and real data sets. In particular our real case study is a mode choice model based on stated preference data, collected in 2003 in the Walloon Region (Belgium).

1. Introduction

This paper gives some new insights into the problem of mixed logit model estimation using both Monte-Carlo and quasi-Monte Carlo methods. Our research group has investigated the problem from the optimisation perspective and found that the choice of the appropriate method can reduce drastically the computation time. We have also developed an adaptive algorithm that uses MC statistical properties to reduce the number of draws needed to estimate the multi-dimensional integrals that are present in the objective function, when a smaller accuracy is sufficient. Transport modelers are however more familiar with QMC methods, in particular Halton sequences (which have been implemented for instance into Gauss by Kenneth Train, 1999) are probably the most used in our environment. The difficulty to extend their validity to high dimensional problems has raised new interests in more recent QMC developments and in their randomizations. Recently Sándor and Train have studied the performance of (t,m,s)-nets in MMNL estimation, while Garrido has proposed the use of Sobol sequences (which are a particular case of (t,m,s)-nets) for their good uniformity in the integration space.
This paper represents a first step towards the adoption of this approach for QMC sequences. In particular, we adopt Sobol sequences and we review three different randomization methods, that have been proposed to provide unbiased estimators of multi-dimensionals integrals and to estimate error (Hong and Hickernell, 2003).

Sobol sequences are indicated as low-discrepancy point sets, this definition implying that those points are more regularly distributed than a typical set of i.i.d. uniform points. For this reason, researchers indicate that QMC usually approximate the integral with a smaller error than MC. However the dimensions of the problem need to be not too large, which is in general true in real transport models. The idea of randomizing QMC point sets has been an important contribution that has extended the practical use of these methods, since this allows us to view them as variance reduction techniques.

We are also interested in the applicability of these methods to practical problems in transport modeling. We see many advantages in the use of statistical inference on the objective function. The accuracy (the error made by solving the approximated problem) and the simulation bias can be implemented into a dynamic algorithm that adapt the sample size to the optimization problem and save computing time. (for more accurate definition of those parameters see Bastin et al, 2004). The analyst can check the quality of the results and decide to stop if log-likelihood improvements are too small to be significant with comparison to the error.

Finally, in order to underline the motivations that have lead this research work we would like to cite L'Ecuyer and Lemieux (2002): "*The success of QMC methods in practice is due to a clever choice of point sets exploiting the features of the functions that are likely to be encountered, rather than to an unexplainable way of breaking the curse of dimensionality*".

The paper is organized as follow: Section 2 briefly introduces the MMNL model specification and estimation; in particular basic concepts to understand Sobol sequences and the

relative randomization techniques are described. We dedicate Section 3 to the algorithm. Results on simulated data are on Section 4, while the application to the SP model developed for the Walloon Region is described in Section 5. Section 6 ends the paper and proposes some further research directions.

2. Mixed Multinomial Logit Model formulation (MMNL)

The mixed logit formulation is nowadays extensively used in transport modeling for its flexibility. In particular, MMNL models estimate taste variation, avoid the problem of restricted substitution pattern in standard logit model and account for state dependency across observations.
Mixed logit probabilities are expressed by means of the integral of standard logit probabilities over a density of parameters:

$$P_{ij} = \int L_{ij}(\beta) f(\beta) d\beta, \tag{1}$$

where:

$i (i = 1,\ldots,I)$ is the individual index,
$j (j = 1,\ldots,J)$ is the alternative index,
$L_{ij}(\beta)$ is the logit probability and,
$f(\beta)$ is a density function.

The mixed logit derivation that we will use in our application is based on random coefficients, with a joint distribution $f(\beta)$ that is usually assumed to be continuous. The choice probability is in this case:

$$P_{ij}(\theta) = \int L_{ij}(\beta) \phi(\beta \mid \theta) d\beta, \tag{2}$$

where $\phi(\beta \mid \theta)$ is the density with parameters vector θ.

In the case an individual i chooses among alternatives $j = 1,\ldots,J,$ in choice situations $t = 1,\ldots,T_i,$ (panel data) its utility can be expressed as:

$$U_{ijt} = \beta_i x_{ijt} + \varepsilon_{ijt}, \tag{3}$$

where ε_{ijt} is iid extreme value, $\beta_i = g(\beta \mid \theta)$ is the vector of parameters randomly distributed in the population and x_{ijt} is the vector independent variables. We observe for each individual the sequence of choices $y_i = (j_{i1},\ldots,j_{iT_i})$. The probability to observe the individuals' choices is given by the product of logit probabilities L_{it} (Train, 2003):

$$P_{iy_i}(y_i \mid x_i, \beta) = \int \left(\prod_{t=1}^{T_i} L_{it}(y_{it} \mid \beta) \right) f(\beta) d\beta \,. \tag{4}$$

2.1 MMNL Model estimation

The vector of unknown parameters is estimated by maximizing the log-likelihood function, i.e. by solving the equation:

$$\max_{\theta} LL(\theta) = \max_{\theta} \frac{1}{I} \sum_{i=1}^{I} \ln P_{iy_i}(\theta), \tag{5}$$

where y_i is the vector of alternative choices made by the individual *i*. This involves the computation of $P_{iy_i}(\theta)$ for each individual $i (i = 1,\ldots,I)$, which is impractical since it requires the evaluation of one multidimensional integral per individual. To approximate the integral of the value $P_{iy_i}(\theta)$ a

frequently used approach is to choose a point set , $S_R = \{u_1, \ldots, u_R\} \subset (0,1)^s$, where *s* is the problem dimension, and then take the average value of the function over S_R . This leads to the simulated probability

$$SP_{iy_i}^R = \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} L_{ij_{it}}(\beta_r, \theta), \tag{6}$$

where R is the number of random draws β_r , taken from the distribution function of β . More precisely, $\beta_r = F^{-1}(u_r)$, where F is the cumulative function corresponding to the distribution $f(\beta)$. As a result, θ is now computed as one solution of the simulated log-likelihood problem

$$\max_{\theta} SLL^R(\theta) = \max_{\theta} \frac{1}{I} \sum_{i=1}^{I} \ln SP_{iy_i}^R(\theta). \tag{7}$$

We will denote by θ_R^* one solution of this last approximation (often called Sample Average Approximation, or SAA), while θ^* denotes the solution of the true problem (5).

Monte Carlo method (MC) is usually appropriate to solve the problem (7), it amounts to choosing S_R as a set of *R* i.i.d. distributed points over $(0,1)^s$. $SLL(\theta)$ is an asymptotically unbiased estimator (bias) of $LL(\theta)$ whose error can be approximated with the delta method. In a previous work we use the error calculation to increase the numerical efficiency of an adaptive algorithm, developed in the context of trust region techniques, to estimate mixed logit model. (Bastin et *al.*, 2004b).

Quasi-Monte Carlo (QMC) methods can be seen as the deterministic counterpart to the MC method. They are based on the idea of using more regularly distributed point sets S_R to construct the approximation (6) than the random point sets associated with MC. In order to measure the accuracy of QMC methods, analysts use a set *A* of functions and a definition of discrepancy *D(S_R)*. Once *A* and $D(S_R)$ are determined, one can usually derive upper bounds on the deterministic error, of the following form:

$$SP - P \le D(S_R) V(f), \qquad f \in A \tag{8}$$

It is clear that a small value of $D(S_R)$ is desirable (low-discrepancy sequence notion).
To give an example, we consider the case in which $D(S_R)$ is the rectangular-star discrepancy. To compute this particular definition of $D(S_R)$ one can consider all rectangular boxes in $(0,1)^s$, aligned with the axes and with a "corner" of origin, and then take the supremum, over all these boxes, of the absolute difference between the volume of a box and the fraction of points that fall in it. In this case $D(S) = O(n^{-1} \log^s n)$ where n is the number of draws and s the number of dimensions. This error bound, for a fixed *s*, is a better asymptotic rate than the $n^{-1/2}$ rate associated to MC.

2.3 Sobol Sequences

Sobol sequences belong to the digital nets family, which has been investigated for application to the mixed logit problem by Sandor and Train (2003). While a more general definition can be found in L'Ecuyer and Lemieux (2002), we will used the more restricted, but more practical, definition given by Hong and Hickernell (2003). Let *b* be a prime number superior or equal to 2, and consider the algebra $Z_b = \{0,1,\ldots,b-1\}$, with the corresponding operations. For any nonnegative integer $i = \sum_{k=1}^{\infty} i_k b^{k-1} = \ldots i_3 i_2 i_1 \,(\text{base}\, b)$ define the $\infty \times 1$ vector $\psi(i)$ as the vectors of its digits, that is

$\psi(i)=(i_1,i_2,\ldots)^T$. For any point $z=\sum_{k=1}^{\infty} i_k b^{-k}=0.z_1z_2$ (base b) $\in[0,1)$, let $\phi(z)=(z_1,z_2,\ldots)^T$ denote the $\infty\times 1$ vector of the digits of *z*. Let also $C_1,\ldots,C_s$ be predetermined $\infty\times\infty$ *generator matrices*. The digital sequence in base *b* is $\{a_0,a_1,a_2,\ldots\}$, where each $a_i=(a_{i1},\ldots,a_{is})^T\in[0,1)^s$ is defined by

$$\phi(a_{ij})=C_j\psi(i),\ j=1,\ldots,s,i=0,1\ldots \tag{9}$$

Here and through the remaining of this section, all arithmetic operations take place modulo *b* and the indices of all vectors and matrices are positive.

For Sobol sequences, we consider that the base is $b = 2$. The specification of each generating matrix C^j is given by a primitive polynomial $f_j(z)$ over Z_2 and integers $m_{j,q}$ for $1\le q\le \deg(f_j(z))$, to initialize a recurrence based on $f_j(z)$ that generates the direction numbers defining C^j. The method specifies that the polynomial $f_j(z)$ should be the *j*th one in the list of primitive polynomials over Z_2 sorted by increasing degree, while within each degree a specific order must be given. In order to define the parameters $m_{j,q}$, assume $f_j(z)=z^k+c_{j,1}z^{k-1}+\ldots+c_{j,k}$, where $c_{j,l}\in Z_2$ for each *j*, *l*. The direction numbers $v_{j,1},v_{j,2},\ldots$ are rationals of the form

$$v_{j,q}=\frac{m_{j,q}}{2^q}=\sum_{l=1}^{q} v_{j,q,l}2^{-l}, \tag{10}$$

where $m_{j,q}$ is an odd integer smaller than 2^q, for $q\ge 1$. For the success of this method, the initial values $v_{j,1},\ldots,v_{j,k}$, for the direction numbers must to carefully chosen. The following ones are obtained through the recurrence

$$v_{j,q}=c_{j,1}v_{j,q-1}\oplus\ldots\oplus c_{j,k-1}v_{j,q-k+1}\oplus v_{j,q-k}\oplus v_{j,q-k}2^{-k}, \tag{11}$$

where $\oplus$ denotes a bit-by-bit exclusive-or operation. These direction numbers are then used to define C^j, whose entry in the *l*th row and *q*th column is given by $v_{j,q,l}$.

For our application we have implemented within AMLET the primitive polynomials proposed by Joe and Kuo (2003), allowing to manage up to 1111 number of dimensions. We also exclude each vector containing a component equal to 0 as it could produce components equal to minus infinite, when using inverse cumulative functions to transform random distributions.

2.4 Randomizations

We now consider the problem of the error estimation in QMC methods. Upper bounds given by (8) are in general not very useful since they are usually much too conservative, in addition to being hard to compute and restricted to a possible small set of functions (L'Ecuyer and Lemieux, 2002). Instead, we randomize the set S_R so that:

1. each point in the randomized point set S_R has a uniform distribution over $(0,1)$;
2. the regularity (or low discrepancy property) of S_R, as measured by a specific quality criterion, is preserved under the randomization.

Randomized QMC can be viewed as variance reduction techniques; in practice the variance of integral estimator can be estimated by generating i.i.d. copies of the estimator through independent replications of the randomization. This estimator can be compared with the estimated variance of the MC estimator to assess the effectiveness of QMC for any particular problem.

Let $\{d_0, d_1, d_2, \ldots,\}$ denote the randomly scrambled version of the original sequence $\{a_0, a_1, a_2, \ldots\}$ as proposed by Owen. Let a_{ijk} denote the *k*th digit of the *j*th component of a_i, and similarly for d_{ijk}. Then

$$d_{ij1} = \pi_j(a_{ij1}), d_{ij1} = \pi_{a_{ij1}}(a_{ij2}), d_{ij1} = \pi_{a_{ij1},a_{ij2}}(a_{ij3}), \ldots, d_{ij1} = \pi_{a_{ij1},a_{ij2},\ldots,a_{ijk-1}}(a_{ijk}), \quad (10)$$

where the $\pi_{z_1 z_2 \ldots}$ are random permutations of the elements in Z_b chosen uniformly and mutually independently. Owen (1995) has shown that a randomized net preserves the property of $(t, m, s)-$nets almost surely. In this paper we use three randomizations methods, that we describe below.

2.4.1 Owen-scrambling

This first sampling technique that we consider is due to Hong and Hickernell (2003), that however called this Owen-technique since it is a particular application of the original Owen proposal (1998). Let $L_1, \ldots, L_s$ be nonsingular lower triangular $\infty \times \infty$ matrices and let $e_1, \ldots, e_s$ be an $\infty \times 1$ vector. The elements of the L_j and e_j are chosen randomly over Z_2. A particular Owen-scrambling $\{d_0, d_1, d_2, \ldots,\}$ of a digital sequence $\{a_0, a_1, a_2, \ldots\}$ is defined as

$$\varphi(x_{ij}) = L_j \varphi(a_{ij}) + e_j = L_j C_j \psi(i) + e_j, \quad j = 1, \ldots, s, i = 0,1, \ldots \quad (11)$$

The scrambled sequence benefits from strong theoretical properties that are analyzed by Hong and Hickernell; we refer the reader to their paper (2003) since this discussion goes beyond the purposes of this paper. We just note here that it can be shown that the scrambled sequence is also a digital (t, s)-sequence with the same *t*-value as the original sequence.

2.4.2 Faure-Tezuka-scrambling

The Faure-Tezuka-scranbling (Faure and Tezuka, 2002) scrambles the digits of *i* before multiplying by the generator matrices. Let *L* be a nonsingular lower triangular $\infty \times \infty$ matrix and let *e* be an $\infty \times 1$ vector with a finite number of nonzero elements. As before, the elements of *L* and *e* are chosen randomly. A particular Faure-Tezuka-scrambling $\{d_0, d_1, d_2, \ldots,\}$ of a digital sequence $\{a_0, a_1, a_2, \ldots\}$ is defined as

$$\varphi(x_{ij}) = C_j \left[L^T \psi(i) + e \right], \quad j = 1, \ldots, s, i = 0,1, \ldots \quad (12)$$

The effect of the Faure-Tezuka-scrambling can be thought as reordering the original sequence, rather than permuting its digits like the Owen-scrambling. For any $m, \lambda = 0,1, \ldots,$ let $i = \lambda b^m, \ldots, (\lambda + 1) b^m - 1$. $\psi(i)$ then takes on all possible values in its top *m* rows, but not necessarily in the same order.

2.4.3 Combined scrambling (Owen and Faure- Tezuka scrambling)

The two previous scrambling techniques are simultaneously applied. Mathematically speaking, a particular Owen-Faure-Tezuka-scrambling $\{d_0, d_1, d_2, \ldots,\}$ of a digital sequence $\{a_0, a_1, a_2, \ldots\}$ is defined as

$$\varphi(x_{ij}) = L_j C_j \left[L^T \psi(i) + e \right] + e_j, \quad j = 1, \ldots, s, i = 0,1, \ldots \quad (13)$$

3. The algorithm

The maximisation of the log-likelihood function can be seen as a generalization of classical stochastic programming problems. A large number of different optimization algorithms can be used in the solution of such problems; classically, these have included Newton-Raphson, BHHH, and BFGS linesearch methods. The BHHH approach can be much faster than other methods, but can be occasionally fail to produce a solution; BFGS on the other hand is usually seen as good compromise between efficiency and robustness.

In this paper, we use basic trust-region (BTR) methods, which have proved to be one of the most powerful approaches in non-linear programming (see Conn, Gould and Toint, 2000, for an exhaustive review of these methods). The main idea of a trust-region algorithm involves the calculation, at iteration k (with current estimate θ_k), of a trial point $\theta_k + s_k$ by maximizing a model m_k of the objective function inside a trust region defined as

$$B_k = \left\{\theta \in R^m \text{ such that } \left\|\theta - \theta_k\right\| \leq \Delta_k\right\}, \tag{10}$$

where Δ_k is called the trust-region radius. The predicted and actual increases in the value of the objective function are then compared. If the ratio between these two values is greater than a certain threshold, the trial point becomes the new iterate, and the trust-region radius is (possibly) enlarged. If the ratio is below the bound, the trust region is shrunk in order to improve the correspondence of the model to the true objective function.

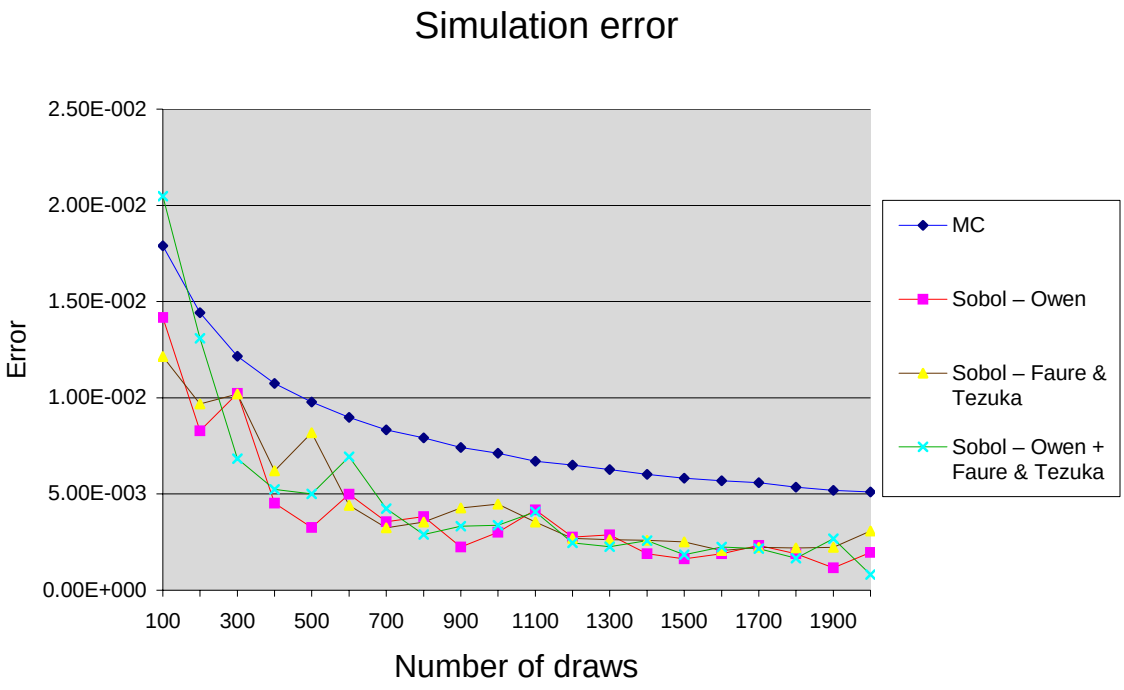
A major advantage of the trust-region approach is that it can easily be adapted to include a variable sample size strategy, as proposed by Bastin et al. (2004a). Such an approach is based on the idea of generating a full set of draws prior to optimisation, but to only use part of it during certain stages of the optimisation process. This is motivated by the understanding that the first steps in an optimisation process are rough steps in the general direction of the optimum, requiring a relatively lower level of precision in simulation. The full set of draws is used during the last few iterations; this not only guarantees maximum simulation-precision at this stage of the optimisation, but also means that the problem used at this stage of the optimisation is identical to that used in methods not based on variable sample size strategies.

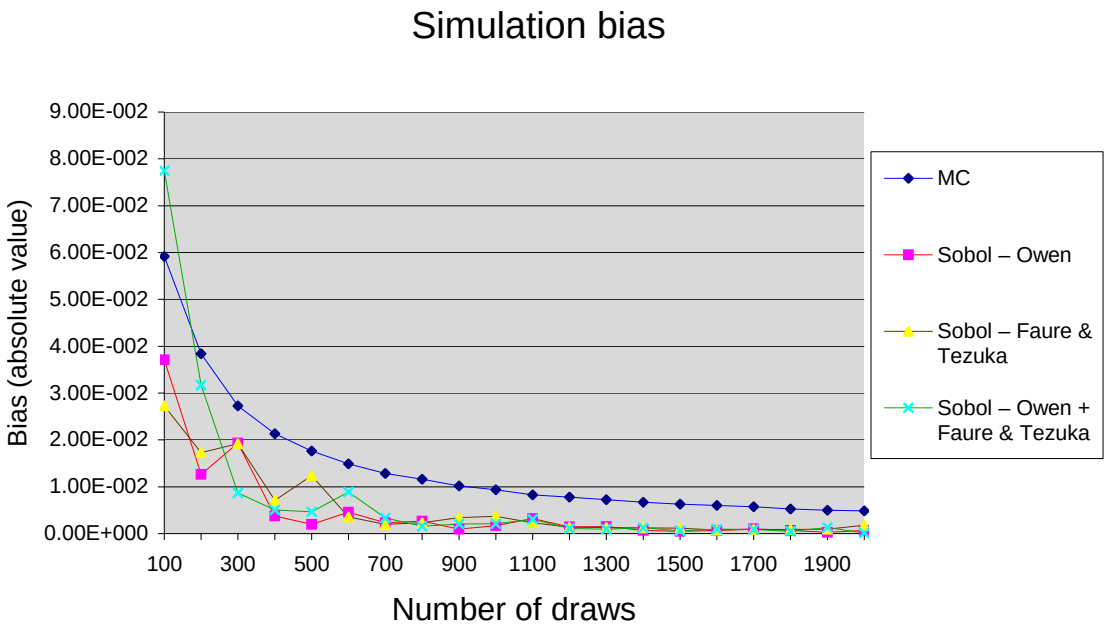
Formally, at each iteration of the trust-region algorithm, the estimation software evaluates whether a significant rising of the objective function at the next iterate is obtained by comparison to the accuracy of the objective function itself, depending on the current number of draws used and on the fitting between the model and the objective function. If the step between successive iterates is large enough by comparison to the error on the objective function, the number of draws is reduced on the basis that an inferior number of draws should be sufficient. If, on the other hand, the step length is significantly smaller than the actual precision of the objective function, the sample size is increased in an attempt to correct this deficiency, and is computed on base of the bias and the variance of the simulated log-likelihood.

4. The simulations

In this section, we compare the effectiveness of the three quasi-Monte-Carlo methods against pure Monte Carlo by means of simulations. The experiment simulates 1000 synthetic individuals giving 5 multiple responses; hypothetic scenarios are built on 5 alternatives, each of them containing 5 continuous variables (drawn from normal distribution N[0,1]). All the parameters are supposed to be normally distributed N[0.5,1].

Monte Carlo accuracy and bias are calculated analytically, as explained in Bastin et al. (2004b). For quasi-Monte Carlo runs we repeat ten times the evaluation of the log-likelihood function at each iteration, and we then deduce the simulation bias and error from the variance observed for the ten i.i.d. runs. Experiments have been conducted varying the number of draws from 100 to 2000, with steps of 100 draws. We plot on Graphs 1 and Graphs 2 the final error on the log-likelihood function and the relative bias. In terms of accuracy and bias quasi-Monte Carlo methods always perform better than Monte Carlo methods, except for Sobol with Owen and Tezuka scrambling methods for very low number of draws (100 and 200). The shape of the curb is irregular: when increasing the number of draws we sometimes get worse results in terms of accuracy and bias. We assume that this is due, at least partially, to the fact that 10 repetitions are not enough to estimate those parameters. It is also difficult to say in this case which is the QMC method that performs better; the effectiveness becomes stable after 1000 draws and half of the equivalent MC draws.





5. Real case study

5.1 The data set: SP survey in the Walloon Region (B)

The data sets used for this study, have been extracted from two stated preference surveys (for more details on the project see Bernard and al., 2003), collected to study individual mode choice for trips on distances longer than 5 km. Both data sets were executed in the Walloon Region (Belgium) in June 2003, during the morning peak hours (from 6:30 to 9:30). The surveys were assisted by personal computer and administrated face-to-face. The first game, called SP1, is a within mode survey (unlabelled experiment), in which we proposed the choice against two different public transport alternatives; the second (SP2 game) gives to private car users the choice between car and public transport service. We collected a total number of 677 responses, of which 286 at home, 207 at the exit of the main train stations of the Region (Namur, Charleroi, Liège, Nivelles, Ottignies) and 184 at the exit of bus stations (same locations).

Those intercepted at the exit of the station were questioned about their trips; then the stated preference games were generated using the revealed trip characteristics. Those travelers evaluated 4 SP1 and 4 SP2 scenarios. People interviewed at home were asked about a trip randomly selected from their displacements longer than 5 km performed during the week before the day of the survey; car users evaluated 8 SP2 scenarios, while public transport users were again submitted to 4 scenarios for both experiments.

The choice experiment contains five variables:

1. in vehicle time,
2. cost,
3. frequency,
4. number of changes,
5. comfort.

The experimental design is a 3^4 x 5^1 profile (cost being the only variable with 5 levels of variation). An orthogonal design was then constructed to reduce to 16 the number of alternatives. Details for the level of variations for each variable are reported in Table 1 for the SP1 game and in Table2 for the SP2 game.

Table 1 : Variables and variation levels in the SP1 game

N.	Variables	**RW-SP1 Within mode experiment**	
		Public Transport	Public Transport
1	In vehicle time	-25, 0, +25 % compared to actual in vehicle time	
2	Fare	-50, -30, -15, 0, +15 % compared to actual fare	
3	Frequency	-1, 0, +1 number of train/bus per hour	
4	Number of changes	-1, 0, +1 changes on the same mode	
5	Comfort	no seats, very crowded no seats, not very crowded seats available	

Table 2 : Variables and variation levels in the SP2 game

N.	Variables	**RW-SP2 Between mode experiment**	
		Public Transport	Car
1	In vehicle time	-30, -15, 0 % compared to actual in vehicle time	0, +15 + 30 % compared to actual in vehicle time
2	Fare (if CAR user) (if PT user)	 -50, -30, -20, -15, 0 % -50, -30, 0, +15,+20% compared to actual fare	 0, +25, +50 % compared to actual car cost
3	Frequency	-1, 0, +1 number of train/bus per hour	
4	Number of changes	-1, 0, +1 changes on the same mode	
5	Comfort	no seats, very crowded no seats, not very crowded seats available	

5.2 Model results

After data cleaning the number of observations available to estimate the model are 3792 belonging to 837 individuals. We estimate a binary mode choice model; variables include one alternative specific constant (public transport is the base), socio-economic characteristics (income, number of cars, size of party and number of season tickets) and level of service variables (time, cost, number of changes, frequency, comfort (on two levels of variation: level 1, crowded no seats, level 2, seats available), access + egress time, distance by car).

We report in Table 3 the estimation results for both multinomial logit and mixed logit formulations. The mixed logit has got 5 dimensions, alternative specific constant is normal distributed, the remaining four parameters are all specified as log-normal. In earlier runs we have

specified all parameters as normal, but although we have estimated them with the right sign, they showed very large standard deviations (compared to the mean values) and consequently a large part of the population was assigned a parameter with the wrong sign.

The model was estimated using AMLET; the same specification did not converge under GAUSS (Kenneth Train's code) although we have been trying different optimization routine (Paul Ruud routine and Maxlik, BHHH and BFGS algorithms) and different starting values. The use of mixed logit model does improve the model fit, the rho-squared adjusted, based on the degrees of freedom, improves from 0.2640 to 0.4256.

In the multinomial logit model we calculate a value of time of about 9.6 Euro per hour. For the mixed logit specification, the VOT is calculated by means of simulations (Hensher and Greene, 2003). We generated 30.000 draws of time and cost parameters from their lognormal distributions and we computed the ratio over the two values. We obtain a VOT median of 6.4 Euro/hour and a median value for access and egress time of about 11.7 Euro/hour. Mean values are very large although we removed the last few percentiles of the distributions (highest and lowest two) as often suggested in the literature.

We believe that the model specification should be improved by adding for example Revealed Preference data, already available from a national and regional travel surveys. Other random distributions rather than lognormals, such as Johnson's S_b could be estimated. This however requires further developments in AMLET, that are currently pursued.

In Table 4 we compare, for the real data set, the final log-likelihood values, the accuracy and the bias obtained with Monte-Carlo methods, Sobol sequences and randomly scrambled Sobol (Owen method). We observe that, as for the simulated data, QMC methods outperform standard Monte-Carlo methods. The estimation error is of great help to identify this superiority. In particular, we observe that the simulation error and bias decrease faster with QMC draws than with pseudo-random draws.

However, the error estimation associated to scrambled Sobol draws suffers from the same weakness than in the simulated data. More precisely, the small number of repetitions does not allow for highly accurate estimation of error and bias, but using a higher number of log-likelihood evaluations increases the associated numerical cost. This directly penalizes the application of variable sample size strategy to QMC algorithms. The first experiment, where simulation is estimated at each iteration takes indeed longer for the same level of accuracy on the final log-likelihood function. However, these experiments have been conducted using the same algorithm as for MC draws. The predicted number of draws for subsequent iterations is then higher than needed. Moreover we argue that it is not necessary to compute the error at each iteration, especially during the last iterations, where the maximum number of draws is used. We are therefore exploring how to adapt such ideas to our dynamic accuracy algorithm in AMLET.

6. Conclusions

In this paper we have studied the performance of Quasi Monte-Carlo methods for mixed logit model estimation. In particular, we see deterministic Sobol sequences as a valid alternative to Halton sequences for their good coverage of the integration domain in high dimension problems. Three randomization techniques have also been the object of our investigation; the randomization allows us to estimate error and bias of the log-likelihood function. We found that this additional information is very helpful to researchers and practitioners, in order to establish the right number of draws required to efficiently solve the optimization problem while ensuring a sufficient accuracy.

The application to both artificial and real data is consistent in results; for the same number of draws, QMC is superior to MC: error and bias are, in fact, found to decrease faster. The three randomizations methods perform equivalently. We also make use of error and bias estimation in an algorithm that adapts the number of quasi-random draws to the function to be estimated. However, our technique, which has been developed for MC simulations, still needs to be refined for QMC methods.

We would like to conclude the paper by saying that QMC methods still have a good potential, but their properties must be well understood in order to be successfully applied to each case.

Table 3: Model results

Parameter	Alternative	Multinomial logit		Mixed logit	
		β	t-stat.	β	t-stat.
ASC Car (mean)	CAR	0.609	3.6	0.314	0.3
ASC Car (s.d.)	CAR	-	-	6.469	7.3
Comfort lev. 1	PT	-0.386	4.2	-0.883	5.2
Comfort lev. 3	PT	0.504	5.8	1.172	7.5
Distance by car	CAR	-0.010	3.4	-0.094	3.5
Access + Egress time	PT	-0.010	2.0	-0.058	1.4
Number of cars	CAR	0.117	3.7	0.764	2.8
Female with child(ren)	CAR	0.634	7.9	3.541	4.5
Size of party	CAR	1.158	7.6	5.531	3.9
Number of season tickets	PT	1.081	8.3	3.980	3.4
Time (mean)	All	-0.024	9.8	-3.475	9.9
Time (s.d.)	All	-	-	1.544	4.8
Cost (mean)	All	-0.150	7.2	-1.258	3.1
Cost (s.d.)	All	-	-	2.632	6.5
Number of changes (mean)	PT	-0.758	14.1	-5.184	1.7
Number of changes (s.d.)	PT	-	-	10.49	2.5
Frequency (mean)	PT	0.142	6.0	-3.179	3.5
Frequency (s.d.)	PT	-	-	2.716	4.0
Number of individuals		837		837	
Number of observations		3792		3792	
Log-likelihood at zero		-2628.41		-2628.41	
Log-likelihood final		-1921.63		-1491.82	
number of parameters		13		18	
rho-squared adjusted		0.2640		0.4256	

Table 4: Comparing MC and QMC

	Number of Draws	*Final log-likelihood*	*Accuracy*	*Bias*
Monte-Carlo	125	-1499.23	8	-11.83
	250	-1501.31	5.94	-6.53
	500	-1508.08	4.63	-3.96
	1000	-1489.53	3.19	-1.89
	2000	-1493.43	2.19	-0.89
Sobol	125	-1499.30	N/A	N/A
	250	-1491.36	N/A	N/A
	500	-1497.81	N/A	N/A
	1000	-1494.05	N/A	N/A
	2000	-1489.52	N/A	N/A
Sobol-Owen	125	-1507.51	6.92	-8.85
	250	-1494.81	5.57	-5.74
	500	-1495.88	2.95	-1.61
	1000	-1490.91	2.47	-1.13
	2000	-1488.15	1.42	-0.37

References

Bastin F., Cirillo C., and Ph.L. Toint (2004a), "A trust region Monte-Carlo algorithm for mixed logit estimation". Submitted to *Transportation Research part B (under revision).*

Fabian Bastin, Cinzia Cirillo et Philippe L. Toint (2004b). *An adaptive Monte Carlo algorithm for computing mixed logit estimators*. Submitted to Computational Management Science.

Bernard P.Y., Cirillo C. Cornelis E. and B. Masquilier (2003) "Integration du transport public au modele de transport de personnes de la Region Wallonne" Rapport final Publications of Mathematics Department N.2003/23 at FUNDP – GRT.

Bhat, C. R. (2001) Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model, *Transportation Research,* **35B** (7) 677-693.

Bratley P. and B.L. Fox. (1988) Algorithm 659: implementing Sobol's quasi-random sequence generator. *ACM Transactions and Mathematical Software*, 14(1): 88-100.

Joe S. and F.Y. Kuo (2003) Remark on algorithm 659: Implementing Sobol's quasirandom sequence generator, ACM Transactions on Mathematical Software (TOMS), 29(1).

Garrido R.A. (2003) Estimation performance of low discrepancy sequences in stated preferences. Paper presented at 10th IATBR, Lucerne

Hensher D.A. and W. L. Greene (2003) The mixed logit model: The state of practice, *Transportation*, **30** (2) 133-176.

Hong H. S. and F. J. Hickernell (2003) Algorithm 823: Implementing Scrambled Digital Sequences, ACM Transactions on Mathematical Software, Vol. 29(2) 95-109

L'Ecuyer P. and C. Lemieux (2002) Recent advances in randomized quasi-Monte Carlo methods, in Modeling Uncertainty: An examination of Stochastic Theory, Methods and Applications. Kluwer Academic Publishers, pp. 419-474.

Owen A.B. (1995) Randomly permuted (t,m,s)-nets and (t,s)-sequences. In Monte-Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Niederreiter and Shiue, Eds. Lecture Notes in Statistics, vol. 106. Springer-Verlag, New-York, NY, 299-317.

Owen A.B. (1998) Scrambling Sobol and Niederreiter-Xing points. Journal of Complexity, 14:466-489

Faure H. and S. Tezuka (2002) Another random scrambling of digital (t,s)-sequences. In Fang, Hickernell and Niederreiter, Eds. 2002. Monte Carlo and Quasi-Monte Carlo Methods 2000. Springer-Verlag, Berlin, Germany.

Train, K. (2003) *Discrete choice analysis with simulations*, Cambridge University Press, New York.