Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model

Chandra R. Bhat

Department of Civil Engineering

University of Texas at Austin

Abstract

This paper proposes the use of a quasi-random sequence for the estimation of the mixed multinomial logit model. The mixed multinomial structure is a flexible discrete choice formulation which accommodates general patterns of competitiveness as well as heterogeneity across individuals in sensitivity to exogenous variables. The estimation of this model has been achieved in the past using the pseudo-random maximum simulated likelihood method that evaluates the multi-dimensional integrals in the log-likelihood function by computing the integrand at a sequence of pseudo-random points and taking the average of the resulting integrand values. We suggest and implement an alternative quasi-random maximum simulated likelihood method which uses cleverly crafted non-random but more uniformly distributed sequences in place of the pseudo-random points in the estimation of the mixed logit model. Numerical experiments, in the context of intercity travel mode choice, indicate that the quasi-random method provides considerably better accuracy with much fewer draws and computational time than does the pseudo-random method. This result has the potential to dramatically influence the use of the mixed logit model in practice; specifically, given the flexibility of the mixed logit model, the use of the quasi-random estimation method should facilitate the application of behaviorally rich structures in discrete choice modeling.

Keywords: Mixed multinomial logit model, maximum simulated likelihood estimation, pseudo-random sequences, quasi-random sequences, polynomial-based cubature, discrete choice analysis.

1. Introduction

The econometric field has seen the increasing use of models that involve analytically intractable criterion functions during the estimation and inference process. The criterion function depends on the inference approach adopted and may take the form of a log-likelihood function (for the maximum likelihood inference approach), a pseudo-log likelihood function (for the pseudo-maximum likelihood inference approach), or a conditional moment function (for the Generalized Methods of Moments inference approach). The analytic intractability of the criterion function in all these instances usually arises because of the presence of multi-dimensional integrals.

The increasing use of models with analytically intractable criterion functions may be attributed to two reasons. First, analysts are formulating and estimating models which do not impose *a priori* behavioral restrictions on the mechanism underlying the decision process being examined. Many analytically tractable models, though elegant and simple in structure, maintain restrictions which are difficult to justify. For example, in a discrete choice context, the multinomial logit model has a simple form, but is saddled with the independent from irrelevant alternatives (IIA) property. Relaxing such rigid behavioral restrictions tends to lead to analytically intractable models. Second, the tools available for data processing have seen dramatic improvement over the past few years. This has made possible the implementation of numerical methods to evaluate multi-dimensional integrals in criterion functions.

The numerical methods to evaluate multi-dimensional integrals may be categorized into three broad groups: a) multi-dimensional polynomial-based cubature methods, b) Monte Carlo simulation methods, and c) quasi-Monte Carlo simulation methods. We discuss each of these approaches in the subsequent paragraphs.

The multi-dimensional polynomial-based cubature methods are extensions of the one-dimensional quadrature principle. The theory of integration formulas in cubature methods are closely related to the theory of orthogonal polynomials (see Stroud, 1971). However, the construction of efficient integration formulas based on polynomial interpolation for multi-dimensional integration is substantially more difficult than for single dimensional integration. An important reason is that the theory of orthogonal polynomials is considerably more complex in more than one variable than in just one variable (see Cools, 1992 and Cools and Rabinowitz, 1993). Consequently, polynomial interpolation-based integration is not generally considered for multi-dimensional integration, though it is the method of choice for one-dimensional integration. The one qualification to the above statement is when the region of integration is "nice" (say, the s-cube or the s-sphere) and when the multi-dimensional integral can be transformed into the product of s single integrals for which well-known Gaussian quadrature formulas exist. In such situations, one can construct product formulas from s one-dimensional Gauss formulas each of which use N points. The resulting product formula will contain N^s points. Unfortunately, this number (N^s) increases very rapidly with N and s (for example, with 15 one-dimensional evaluation points and just 3-dimensions for integration, the total number of points for the product formula climbs to 3375!). Consequently, product formulas of one-dimensional Gauss formulas are unable to compute integrals with sufficient precision and speed for optimization of the criterion function in higher than 2 dimensions (see Hajivassiliou and Ruud, 1994; Sloan and Joe, Chapter 1, 1994).

The Monte-Carlo simulation method (or "the method of statistical trials") to evaluating multi-dimensional integrals entails computing the integrand at a sequence of "random" points and computing the average of the integrand values. The basic principle is to replace a continuous average by a discrete average over randomly chosen points. By the strong law of large numbers, convergence is almost sure in this method. The Monte Carlo simulation method has a long history, dating back to the work of Metropolis and Ulam (1949). The concept has been known even before that, though the practical viability of the method had to wait till the turn to the computer era in the early 50s (Halton, 1970 presents an exhaustive review of the history of the method). Of course, in actual implementation, truly random sequences are not available; instead, deterministic pseudo-random sequences which appear random when subjected to simple statistical tests are used (see Niederreiter, 1995 for a discussion of pseudo-random sequence generation). This pseudo-Monte Carlo (or PMC) method has a slow asymptotic convergence rate with the expected integration error of the order of $N^{-0.5}$ (N being the number of pseudo-random points drawn from the s-dimensional integration space). Thus, to obtain an added decimal digit of accuracy, the number of draws needs to be increased hundred fold. Also, the PMC method does not distinguish between nicely behaved smooth integrands and poorly behaved non-smooth integrands. That is, smooth functions are not integrated any more accurately or efficiently than non-smooth functions (this is really the other side of the coin of the oft-touted advantage of the PMC method that it works for any integrand with a finite variance; see Spanier and Maize, 1994).

The quasi-Monte Carlo method is similar to the Monte Carlo method in that it evaluates a multi-dimensional integral by replacing it with an average of values of the integrand computed at discrete points. However, rather than using pseudo-random sequences for the discrete points, the quasi-Monte Carlo approach uses "cleverly" crafted non-random and more uniformly distributed sequences within the domain of integration. The underlying idea of the method is that it is really inconsequential whether the discrete points are truly random; of primary importance is the even distribution (or maximal spread) of the points in the integration space. The sequences used in the quasi-Monte Carlo (or QMC) method are labeled as quasi-random sequences, though this is a misnomer since randomness plays no part in the construction of the sequences. The convergence rate for quasi-random sequences is, in general, faster than for pseudo-random sequences. In particular, the theoretical upper bound for the integration error in the QMC method is of the order of N^{-1}, where N is the number of quasi-random integration points.[1]

To summarize, quasi-MC sequences are more uniformly distributed than random sequences (a uniformly distributed sequence can loosely be described as one which assigns to each sub-interval of the integration domain its "fair" percentage of points). Mathematicians use a measure called "discrepancy" to assess the uniformity of a sequence; lower discrepancy values implies greater uniformity. The discrepancy is directly related to the integration error bound and is of order $N^{-0.5}$ for the random sequences and of order N^{-1} for the quasi-random sequences. The lower discrepancy of the quasi-random sequences results in lower integration error and faster convergence rates. This is especially the case for well behaved and smooth integrands. Thus, for

---

[1]The aforementioned convergence result for the QMC method holds only for integrands with bounded variation in the sense of Hardy and Krause; this requirement of the integrand is more restrictive than the finite variance requirement in the Monte Carlo method. But for most practical purposes, the restriction is not of a serious nature (see Niederreiter, 1995 or Morokoff and Caflisch, 1994 for a more detailed theoretical discussion). Also, the reader should note that the convergence result provided here is an upper bound; in the average case, Wozniakowski (1991) has shown a much faster convergence rate for QMC sequences. Further, Owen (1997, 1998) has shown recently that some scrambled versions of QMC sequences have a convergence rate of the order of N^{-3}.

a given error tolerance level, the quasi-random simulation of integrals for smooth integrands requires significantly less number of simulation points or "draws" relative to the PMC method (see Morokoff and Caflisch, 1995; Press *et al*., 1992, Chapter 7; Brately and Fox, 1988; and Bratley *et al*., 1992).[2]

The appeal of the QMC method should be clear from the above discussion. The generation and application of QMC sequences for multi-dimensional integration has been the subject of intensive research in recent years in mathematics and physics (see, for example, Sloan and Wozniakowski, 1998; Owen, 1998; Mullen *et al*., 1995; Kocis and Whiten, 1997). Krommer and Ueberhuber (1994) provide an extensive review of quasi-random sequences. Against this backdrop of literature extolling the much faster convergence rate and superior accuracy of QMC methods over the PMC method, is the surprising lack of use of these methods in econometric model estimation. While the use of simulation techniques in model estimation has been the focus of much recent econometric work, almost all of this literature is based on the PMC method. Econometricians have refined the PMC method using variation reduction techniques such as stratified sampling, importance sampling, and antithetic variates (see, for example, Hajivassiliou

[2]Theoretically speaking, the discrepancy of a quasi-random sequence also depends on the dimensionality of the integration, with the discrepancy (or integration error) increasing with dimensionality. However, in practice, studies (for example, see Papageorgiou and Traub, 1997 or Paskov, 1997) have shown the integration error to be independent of the number of dimensions. Sloan and Wozniakowski (1998) have recently examined the reasons for this superior empirical performance of the quasi-random sequences for high dimensionality of integration.

et al., 1996); however, these are not specific to the PMC method and are as equally applicable to QMC methods (see Spanier and Maize, 1994).

The objective of this paper is to compare the performance of the polynomial-based cubature method (or simply the cubature method), the PMC method, and the QMC method using numerical experiments. The comparison is conducted in the context of the maximum simulated likelihood estimation of a mixed multinomial logit (MMNL) model. We use the MMNL model form in the numerical analysis for two reasons. First, the MMNL model is a very flexible discrete choice structure. As our review later indicates, it is becoming the method of preference to accommodate unobserved taste variations and general error variance-covariance matrices. Second, the integrands involved in the maximum likelihood estimation of the MMNL model have a multinomial logit structure, which is a very nicely behaved, continuous, and smooth function. Thus, the use of the QMC method would appear to be particularly suitable for the MMNL model. In contrast, all previous applications of the MMNL structure have used the PMC method, except for a very recent use of the QMC method in Bhat (1999a). If our numerical experiments indeed show that the QMC method is very efficient in the MMNL context, it could lead to the more widespread use of the flexible MMNL model.

There are several quasi-random sequences that may be employed in the QMC simulation method. Among these sequences are those that belong to the family of r-*adic* expansion of integers: the Halton, Faure, and Sobol sequences (see Bratley *et al.*, 1992 for a good review). In this paper, we will use the Halton sequence in the QMC simulation because of its conceptual simplicity.

The rest of this paper is structured as follows. The next section discusses the structure of the mixed multinomial logit model. Section 3 presents the methodology to estimate the mixed logit model using the cubature method, the PMC method, and the QMC method. Section 4 describes the design for the numerical experiments. Section 5 identifies the performance criteria used in comparing alternative methods and presents the experimental results. The final section closes the paper.

2. Mixed Multinomial Logit Model

The mixed Multinomial logit (MMNL) model is a generalization of the well-known multinomial logit (MNL) model. It involves the integration of the multinomial logit formula over the distribution of unobserved random parameters. It takes the structure shown below:

$$P_{qi}(\theta) = \int_{-\infty}^{+\infty} L_{qi}(\beta) f(\beta \mid \theta)\, d(\beta), \quad L_{qi}(\beta) = \frac{e^{\beta' x_{qi}}}{\sum_j e^{\beta' x_{qj}}}, \tag{1}$$

where P_{qi} is the probability that individual q chooses alternative i, x_{qi} is a vector of observed variables specific to individual q and alternative i, β represents parameters which are random realizations from a density function *f(.)*, and θ is a vector of underlying moment parameters characterizing *f(.)*.

The MMNL model structure of equation (1) can be motivated from two very different (but formally equivalent) perspectives. Specifically, a MMNL structure may be generated from an intrinsic motivation to allow flexible substitution patterns across alternatives (error-components structure) or from a need to accommodate unobserved heterogeneity across individuals in their sensitivity to observed exogenous variables (random-coefficients structure).

The error-components structure partitions the overall random term associated with each alternative's utility into two components: one component which allows the unobserved error terms to be non-identical and non-independent across alternatives, and the other which is specified to be independent and identically (type I extreme-value) distributed across alternatives. Specifically, consider the following utility function for individual q and alternative i:

$$\begin{aligned} U_{qi} &= \gamma' y_{qi} + \zeta_{qi} \\ &= \gamma' y_{qi} + \mu' z_{qi} + \epsilon_{qi} \end{aligned} \tag{2}$$

where $\gamma' y_{qi}$ and ζ_{qi} are the systematic and random components of utility, and ζ_i is further partitioned into two components, $\mu' z_{qi}$ and ϵ_{qi} . z_{qi} is a vector of observed data associated with alternative i, some of whose elements might also appear in the vector y_{qi}. μ is a random vector with zero mean. The component $\mu' z_{qi}$ induces heteroscedasticity and correlation across unobserved utility components of the alternatives. The emphasis in the error-components structure is to allow a flexible substitution pattern among alternatives in a parsimonious fashion. This is achieved by the "clever" specification of the variable vector z_{qi} combined with (usually) the specification of independent normally distributed random elements in the vector μ. Defining $\beta = (\gamma', \mu')'$ and $x_{qi} = (y_{qi}', z_{qi}')'$, we obtain the MMNL model structure for the choice probability of alternative i for individual q. Applications of the error-components formulation include Brownstone and Train (1999) and Bhat (1998a, 1999b).

The random-coefficients structure allows heterogeneity in the sensitivity of individuals to exogenous attributes. The utility that an individual q associates with alternative i is written as:

$$U_{qi} = \beta_q' x_{qi} + \epsilon_{qi} \tag{3}$$

where x_{qi} is a vector of exogenous attributes, β_q is a vector of coefficients that varies across individuals with density $f(\beta)$, and ϵ_{qi} is assumed to be an independently and identically distributed (across alternatives) type I extreme value error term. With this specification, the unconditional choice probability of alternative *i* for individual *q* is given by the mixed logit formula of equation (1). While several density functions may be used for *f(.)*, the most commonly used is the normal distribution. A log-normal distribution may also be used if, from a theoretical perspective, an element of beta has to take the same sign for every individual (such as a negative coefficient for the travel time parameter in a travel mode choice model). Applications of the random-coefficients formulation include Revelt and Train (1998), Train (1998), Bhat (1998b, 1999), and Ben-Akiva and Bolduc (1996).

The reader will note that the error-components specification in equation (2) and the random-coefficients specification in equation (3) are structurally equivalent. Specifically, if β_q is distributed with a mean of γ and deviation μ, then equation (3) is identical to equation (2) with $x_{qi} = y_{qi} = z_{qi}$. However, this apparent restriction for equality of equations (2) and (3) is purely notational. Elements of x_{qi} that do not appear in z_{qi} can be viewed as variables whose coefficients are deterministic in the population, while elements of x_{qi} that do not enter in y_{qi} may be viewed as variables whose coefficients are randomly distributed in the population with mean zero.

In the numerical experiments in this paper, we use a random-coefficients interpretation of the mixed-logit structure. However, the results from the experiments should be generalizable to any model structure with a mixed-logit form.

3. Estimation Methodology

This section discusses the details of the estimation procedure for the random-coefficients mixed-logit model using each of the three methods: the cubature method, the PMC method, and the QMC method.

Consider equation (3) and separate out the effect of variables with fixed coefficients (including the alternative specific constant) from the effect of variables with random coefficients:

$$U_{qi} = \alpha_{qi} + \sum_{k=1}^{K} \beta_{qk} x_{qik} + \epsilon_{qi}, \tag{4}$$

where α_{qi} is the effect of variables with fixed coefficients. Let $\beta_{qk} \sim N(\mu_k, \sigma_k)$, so that $\beta_{qk} = \mu_k + \sigma_k s_{qk}$ $(q = 1,2,\ldots,Q;\ k = 1,2,\ldots,K)$. [3] s_{qk} $(q = 1,2,\ldots,Q;\ k = 1,2,\ldots,K)$ is a standard normal variate. Further, let $V_{qi} = \alpha_{qi} + \sum_k \mu_k x_{qik}$.

The log-likelihood function for the random-coefficients logit model may be written as:

$$\mathcal{L} = \sum_q \sum_i y_{qi} \log P_{qi} = \sum_q \sum_i y_{qi} \log \left\{ \int_{s_{q1}=-\infty}^{s_{q1}=+\infty} \int_{s_{q2}=-\infty}^{s_{q2}=+\infty} \cdots \int_{s_{qK}=-\infty}^{s_{qK}=+\infty} \frac{e^{V_{qi} + \sum_k \sigma_k s_{qk} x_{qik}}}{\sum_j e^{V_{qj} + \sum_k \sigma_k s_{qk} x_{qjk}}} \, d\Phi(s_{q1})\, d\Phi(s_{q2}) \cdots d\Phi(s_{qK}) \right\}, \tag{5}$$

where $\Phi(.)$ represents the standard normal cumulative distribution function and

$$y_{qi} = \begin{cases} 1 \ \text{if the } q\text{th individual chooses alternative } i \\ 0 \ \text{otherwise}, \end{cases} \qquad (q = 1,2,\ldots Q,\ i = 1,2,\ldots,I) \tag{6}$$

---

[3]In this notation, we are implicitly assuming that the $\beta_{qk}'s$ are independent of one another. Even if they are not, a simple Choleski decomposition can be undertaken so that the resulting integration involves independent normal variates (see Revelt and Train, 1998).

The cubature method, the PMC method, and the QMC method represent three different ways of evaluating the multi-dimensional integral involved in the log-likelihood function.

3.1. Polynomial-Based Cubature Method

To apply the cubature method, define $\varpi_k = s_{qk}/\sqrt{2}$ for all q. Then, the log-likelihood function in equation (5) takes the following form:

$$\mathcal{L} = \sum_q \sum_i y_{qi} \log \left\{ \left(\frac{1}{\sqrt{\pi}} \right)^K \int_{\varpi_1=-\infty}^{\varpi_1=+\infty} \int_{\varpi_2=-\infty}^{\varpi_2=+\infty} \cdots \int_{\varpi_K=-\infty}^{\varpi_K=+\infty} \frac{e^{V_{qi}+\sqrt{2}\sum_k \sigma_k \varpi_k x_{qik}}}{\sum_j e^{V_{qj}+\sqrt{2}\sum_k \sigma_k \varpi_k x_{qjk}}} e^{-\varpi_1^2} e^{-\varpi_2^2} \ldots e^{-\varpi_K^2} d\varpi_1 \, d\varpi_2 \cdots d\varpi_K \right\}. \quad (7)$$

The above integration is now in an appropriate form for application of a multi-dimensional product formula of the Gauss-Hermite quadrature (see Stroud, 1971).

3.2. Pseudo-Random Monte Carlo (PMC) Method

This technique approximates the choice probabilities by computing the integrand in equation (5) at randomly chosen values for each s_{qk}. Since the s_{qk} terms are independent across individuals and variables, and are distributed standard normal, we generate a matrix s of standard normal random numbers with $Q*K$ elements (one element for each variable and individual combination) and compute the corresponding individual choice probabilities for a given value of the parameter vector to be estimated. We then repeat this process R times for the given value of the parameter vector and approximate the integrand by averaging over the computed choice probabilities in the different draws. This results in an unbiased estimator of the actual individual choice probabilities. It's variance decreases as R increases. It also has the appealing properties of being smooth (*i.e.*, twice differentiable) and being strictly positive for any realization of the finite

R draws. The parameter vector is estimated as the vector value that maximizes the simulated log-likelihood function. Under rather weak regularity conditions, the PMC estimator is consistent, asymptotically efficient, and asymptotically normal (see Hajivassiliou and Ruud, 1994 and Lee, 1992). However, the estimator will generally be a biased simulation of the maximum likelihood (ML) estimator because of the logarithmic transformation of the choice probabilities in the log-likelihood function. The bias decreases with the variance of the probability simulator; that is, it decreases as the number of repetitions increase.

An important point in implementation of the PMC method is that the same matrix s of standard normal random numbers has to be used across iterations of the likelihood function maximization for each of the R repetitions. This can be achieved by re-setting the seed to the same pre-determined value at the beginning of each likelihood (and gradient) function evaluation. This seed is automatically updated for each repetition within a single likelihood (and gradient) function evaluation, and is updated to the same value for any repetition across function evaluations by assigning the same initial seed value at each function evaluation.

3.3. Quasi-Random Monte Carlo (QMC) Method

The quasi-random Halton sequence is designed to span the domain of the S-dimensional unit cube uniformly and efficiently (the interval of each dimension of the unit cube is between 0 and 1). In one dimension, the Halton sequence is generated by choosing a prime number r (r > = 2) and expanding the sequence of integers 0,1,2,..g,…G in terms of the base r:

$$g = \sum_{l=0}^{L} b_l r^l, \text{ where } 0 \le b_l \le r-1 \text{ and } r^L \le g < r^{L+1}. \quad (8)$$

Thus, g ($g = 1,2,\ldots G$) can be represented by the r-*adic* integer string $b_l \ldots b_1 b_0$. The Halton sequence in the prime base r is obtained by taking the radical inverse of g (g = 1,2,…G) to the base r by reflecting through the radical point:

$$\varphi_r(g) = 0.b_0 b_1 \ldots b_L (base\ r) = \sum_{l=0}^{L} b_l r^{-l-1} \tag{9}$$

The sequence above is very uniformly distributed in the interval (0,1) for each prime r. The Halton sequence in K dimensions is obtained by pairing K one-dimensional sequences based on K pairwise relatively prime integers, $r_1, r_2, \ldots, r_S$ (usually the first K primes):

$$\psi_g = (\varphi_{r_1}(g), \varphi_{r_2}(g), \ldots, \varphi_{r_S}(g)) \tag{10}$$

The Halton sequence is generated number-theoretically rather than randomly and so successive points at any stage “know” how to fill in the gaps left by earlier points, leading to a uniform distribution within the domain of integration. This is illustrated in Figure 1, where we have plotted 1000 points from a two-dimensional uniform pseudo-random sequence and correspondingly 1000 points generated by a two-dimensional Halton sequence. The “clumping” of points in the pseudo-random sequence is clearly noticeable, while the Halton sequence has a much more even spread.

The simulation technique to evaluate the integral in the log-likelihood function of equation (5) involves generating the K-dimensional Halton sequence for a specified number of "draws" R for each individual. To avoid correlation in simulation errors across individuals, we make separate independent draws of R Halton numbers in K dimensions for each individual. We achieve this by generating a Halton "matrix" Y of size G x K, where G = R*Q + 10 (Q is the total number of individuals in the sample). We then discard the first ten terms in each dimension because the integrand may be sensitive to the starting point of the Halton sequence (see Morokoff and Calflisch, 1995; Bratley *et al.*, 1992). This leaves a (R*Q) x K Halton matrix which is partitioned into Q sub-matrices of size R x K, each sub-matrix representing the R Halton draws in K dimensions for each individual (thus, the first R rows of the Halton matrix Y are assigned to the first individual, the second R rows to the second individual, and so on). It is important to note that storing a matrix of size (R*Q) x K in memory may not require substantial memory space since typically the number of repetitions R in the QMC estimation method is not high. Alternatively, one could store the matrix on disk and extract the appropriate Q*K matrix corresponding to each draw as needed during the likelihood function iterations.

The Halton sequence is uniformly distributed over the multi-dimensional cube. To obtain the corresponding multivariate normal points over the multi-dimensional domain of the real line, we take the inverse standard normal distribution transformation of Y. By the integral transform result, $X = \Phi^{-1}(Y)$ provides the Halton points for the multi-variate normal distribution (see Fang and Wang, 1994; Chapter 4). The integrand in equation (5) is computed at the resulting points in the columns of the matrix X for each of the R draws for each individual and then the simulated likelihood function is developed in the usual manner as the average of the values of the integrand across the R draws.

4. Experimental Design

The data used in the numerical experiments are obtained from an intercity mode choice survey conduced by VIA Rail (the Canadian national rail carrier) in 1989 to develop travel demand models to forecast future intercity travel in the Toronto-Montreal corridor. We extracted the level-of-service information for each of three modes (car, air and train) and for the trip of each of 2000 weekday-business travelers from the data. The reason for using actual level-of-service data in the experiments is to ensure that the simulation data resemble field data. However, the choice process is generated by simulating the normally distributed random coefficients on the level-of-service parameters. Thus, the experimental data may be characterized as "pseudo-simulated" data, with independent variables obtained from an actual field survey, but the choice being generated using an experimental design.

Five level-of-service variables are included in the x_{qi} vector of equation (3): line-haul cost, access cost, in-vehicle time, access time, and terminal time. To examine the performance of the three alternative numerical estimation methods (*i.e.*, the cubature method, the PMC method and the QMC method) *vis-a-vis* the dimensionality of the integration, we started the experiments by allowing the parameter on line-haul cost to be randomly (normal) distributed and maintaining fixed coefficients on other variables (this leads to one-dimensional integration). Subsequently, we generated new choice data by also allowing the coefficient on access cost to be randomly (normal) distributed (leading to a two-dimensional integration). This process of adding random coefficients was continued till all the level-of-service variables are random (leading to five-dimensional integration). Thus, we generated five data sets, each data set representing a specific number of underlying random coefficients in the choice process.

The mean (and standard deviations) of the coefficients adopted in the data generating process are as follows: line-haul cost (-4.5 and 3), access cost (-6.0 and 4.5), in-vehicle time (-1 and 0.75), access time (-3 and 2.25), and terminal time (-2 and 1.5). The alternative specific constants were assumed to be 3.5 (for the air mode) and -0.3 (for the train mode), with the car mode being the base (the alternative specific constant values and the mean values of the level-of-service coefficients used in the experiments were informed by a multinomial logit estimation on the original data; the standard deviations used were obtained as roughly three-fourths of the mean values). The reader will note that the standard deviation of a level-of-service coefficient applies only if that coefficient is allowed to be random; if it is fixed in a particular experiment, the corresponding standard deviation is zero.

The utility for each mode is computed based on equation (3) after generating values for β_q for each individual from the normal distribution and for ϵ_{qi} from the standard type I extreme value distribution. Then, the brand with the highest computed utility value is identified as the chosen brand in the experiments.

5. Computational Results and Comparative Performance

5.1. Background

The three numerical estimation methods were implemented using the GAUSS matrix programming language. The log-likelihood function and the gradients of the function with respect to relevant parameters were coded. Estimations were carried out using an Intel pentium II 300 Mhz. processor with 128 MB of RAM. The parameters used to generate the data were used as the common starting point for all estimations. This allows a comparison of time to convergence across estimations.

For the estimations involving one and two-dimensional integration, we used the results from the 20 point cubature method as the benchmark for comparing the performance of alternative methods. For estimations in three to five dimensions, we obtained benchmark results using the PMC method with 20,000 draws. While there may be simulation error even when using 20,000 draws, it is expected that this error would be very small.

For one-dimensional integration, cubature estimation using 20 points can be achieved very quickly and so we did not estimate models with lower number of cubature points. For two-dimensional integration, we estimated models with 2, 4, and 10-point cubature. For three-dimensional integration, we obtained results using 2- and 4- point cubature. For higher dimensions, we estimated models with only 2-point cubature (more than 2 points of cubature dramatically increases computational time).

PMC estimation was based on four different numbers of draws: 250, 500, 1000, and 2000. We used 250 draws as the minimum, though it is generally preferable to use higher number of pseudo-random draws to reduce simulation variance. We did not go beyond 2000 draws because computation time starts to increase quite substantially beyond this many number of draws.

We estimated models with 25, 50, and 75 Halton points for one and two-dimensional integration. For higher dimensions, we also estimated models with 100 and 125 Halton points.

5.2. Performance Criteria and Evaluation Results

To evaluate the performance of different methods, we used measures that indicate the ability of each method to recover the "true" parameters and the choice probabilities as obtained in the benchmark estimation. The evaluation of the proximity of estimated and true parameters was based on the root mean square error as well as the mean absolute error ratio across parameters. The evaluation of the ability to reproduce true choice probabilities was similarly based on the root mean square error and mean absolute error ratio of computed choice probabilities (*vis-a-vis* true choice probabilities) across modes and across individuals. Thus, four measures of performance are used. These measures, and the time required for convergence of the maximum likelihood iterations, were compared across methods

Tables 1 through 5 present the error measures for different dimensions of integration. In all the tables, the error measures tend to become smaller as the number of cubature points or the number of draws increase. For the PMC and QMC methods, this is not always the case. For example, the parameter-based error measures increase between 500 and 1000 draws for the PMC method in three dimensions (Table 3). Similarly, the parameter-based error measures increase between 75 and 100 draws for the QMC method in three dimensions. These minor aberrations are not surprising because of the way that draws are assigned to individuals. For example, in the case of 75 Halton draws, the procedure involves drawing 75*N Halton points and assigning each set of 75 points to each individual. Thus, the first 75 points get assigned to the first individual, the second 75 points to the second individual, and so on. For 100 Halton draws, 100*N Halton points are obtained and each set of 100 is assigned to an individual as before. Thus, in the sampling scheme, moving from 75 to 100 draws does not imply maintaining the same 75 draws for each individual and supplementing this with an additional 25 draws. It involves a fresh set of points for each individual, and therefore, there is always a small chance that fewer draws might provide better results than a larger fresh set of draws. The important point to note, however, is that these are minor aberrations to the more stable general trend of reduced errors with higher draws.

The results for one-dimensional integration (Table 1) indicate that with as few as 50 Halton draws, the error measures from the QMC method are smaller than from 1000 draws of the PMC method; those from 75 Halton draws are much smaller than from 2000 pseudo-random draws. Besides, the times to convergence for the QMC method are considerably lesser than for the PMC method. As indicated earlier, we are not presenting cubature results in one dimension because it is easy to estimate one-dimensional integrals with a high number of cubature points. The benchmark used for one-dimensional integration is 20 points, and this estimation takes just 7.5 minutes. Obviously, the cubature method is the preferred method for 1 dimensional integration, but its effectiveness drops significantly beyond one-dimensional integration (as we will see later). The QMC method, on the other hand, does not take significantly longer than the cubature method and is also much more effective in higher dimensions.

The maximum simulated likelihood estimation results for the mixed logit model with two-dimensional integration are provided in Table 2. Polynomial-cubature methods with 2 and 4 points converge quickly, but the associated error measures are rather large. The cubature technique is quite accurate with 10 points, though it requires about 2.5 hours for convergence. However, 10 point cubature is superior to the 2000 draw PMC method both in terms of accuracy as well as convergence time. Among all the three methods, the Halton method comes out as being a clear winner; with as few as 75 draws, it is able to recover the true parameters and choice probabilities more accurately and in substantially less time than 10 point cubature or a 2000-draw PMC method.

In the case of three-dimensional integration, cubature-based estimation of the mixed logit with more than 4 points becomes very expensive. Between the PMC and QMC methods, the QMC method with as few as 75 draws is better than the PMC method with 1000 draws. Also, the QMC method with 125 draws is far superior to the PMC method with 2000 draws, and takes less than one-tenth the time required by the 2000-draw PMC method.

Similar results may be observed for the estimation of the mixed logit model in four and five-dimensions. A particularly important point in 4- and 5-dimensions is that the cubature method beyond 2 points becomes very time-intensive, and estimation with just 2 points leads to very substantial errors.

Overall, the QMC method dominates the other two methods in performance. Also, it takes substantially lesser time for convergence. This is a valuable finding. It suggests that the estimation of the mixed logit model using the Halton method is very efficient and practical.

6. Summary and Conclusions

This paper proposes the use of a quasi-random Monte Carlo (QMC) method (specifically, the Halton method) for the estimation of the flexible mixed logit model for discrete choice analysis. The QMC method uses "cleverly" crafted non-random and uniformly distributed sequences in the domain of integration. The basic idea of the method is that it is really inconsequential whether the discrete points are truly random; of primary importance is the even distribution (or maximal spread) of the points in the integration space.

The results from our simulation experiments indicate that the QMC method out-performs the polynomial-cubature and pseudo-Monte Carlo (PMC) methods for mixed logit model estimation. In less than or equal to three dimensions, simulation estimation with as few as 75 Halton draws provides considerably better accuracy than with 2000 pseudo-random draws. In higher dimensions (4-5), 100 Halton draws provide about the same level of accuracy as 2000 pseudo-random draws and 125 Halton draws provides much better accuracy at about one-tenth the time required for convergence using 2000 pseudo-random draws.

The experimental result have the potential to dramatically influence the use of the mixed logit model in practice. Given the flexibility of the mixed logit model to accommodate very general patterns of competition among alternatives and/or random coefficients, the use of the QMC simulation method of estimation should facilitate the application of behaviorally rich structures for discrete choice modeling.

The author is currently pursuing the comparison of potentially more efficient quasi-random sequences than the Halton sequence used in the current paper.

Acknowledgements

This research was supported by National Science Foundation grants DMS 9208758 and DMS 9313013 to the National Institute of Statistical Sciences (NISS). The author would also like to thank Frank Koppelman and three referees for valuable comments on an earlier version of the paper.

References

Ben-Akiva, M. and D. Bolduc (1996) Multinomial probit with a logit kernel and a general parametric specification of the covariance structure, working paper, Department of Civil and Environmental engineering, Massachusetts Institute of Technology, Cambridge, MA and Département d'économique, Université Laval, Sainte-Foy, Qc, Canada.

Bhat, C.R. (1998a) Accommodating flexible substitution patterns in multidimensional choice modeling: formulation and application to travel mode and departure time choice, *Transportation Research*, 32B, 425-440.

Bhat, C.R. (1998b) Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling, *Transportation Research*, 32A, 495-507.

Bhat, C.R. (1999a) A multi-level cross-classified model for discrete response variables, accepted for publication in *Transportation Research*.

Bhat, C.R. (1999b) Incorporaing observed and unobserved heterogeneity in urban work travel mode choice modeling, forthcoming, *Transportation Science*.

Bratley, P. and B.L. Fox (1988) Implementing Sobol's quasi-random sequence generator, *ACM Transactions on Mathematical Software*, 14, 88-100.

Bratley, P., B.L. Fox and H. Niederreiter (1992) Implementation and tests of low-discrepancy sequences, *ACM Transactions on Modeling and Computer Simulation*, 2, 195-213.

Bratten, E. and G. Weller (1979) An improved low-discrepancy sequence for multidimensional quasi-Monte Carlo integration, *Journal of Computational Physics*, 33, 249-258.

Brownstone, D. and K. Train (1999) Forecasting new product penetration with flexible substitution patterns, *Journal of Econometrics*, Vol. 89, pp. 109-129.

Cools, R. (1992) A survey of methods for constructing cubature formulae, in *Numerical Integration: Recent Development, Software and Application*, T.O. Espelid and A.C. Genz (Editors), Kluwer Academic Publishers, Dordrecht, 1-24.

Cools, R. and P. Rabinowitz (1993) Monomial cubature rules since “Stroud”: a compilation, *Journal of Computational and Applied Mathematics*, 43, 309-326.

Fang, K.-T and Y. Wang (1994) *Number-Theoretic Methods in Statistics*, Chapman and Hall, London.

Halton, J.H. (1970) A retrospective and prospective survey of the Monte Carlo method, *SIAM Review*, 12, 1-63.

Hajivassiliou, V.A., D. McFadden and P. Ruud (1996) Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results, *Journal of Econometrics*, 72, 85-134.

Hajivassiliou, V.A. and P.A. Ruud (1994) Classical estimation methods for LDV models using simulation, in Engle, R.F. and D.L. McFadden (eds.) *Handbook of Econometrics*, 4, 2383-2441.

Kocis, L. and W.J. Whiten (1997) Computational investigation of low-discrepancy sequences, *ACM Transactions on Mathematical Software*, 23, 266-294.

Krommer, A.R. and C.W. Ueberhuber (1994) *Numerical integration on advanced computer systems*, Springer-Verlag, Berlin, Germany.

L-F. Lee (1992) On the efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models", *Econometric Theory*, 8, 518-552.

Metropolis, N.C. and S.M. Ulam (1949) The Monte Carlo method, Journal of the American Statistical Association, 44, 335-341.

Morokoff, W.J. and R.E. Caflisch (1995) Quasi-Monte Carlo integration, *Journal of Computational Physics*, 122, 218-230.

Mullen, G.L., A. Mahalanabis, and H. Niederreiter (1995) Tables of (T,M,S)-net and (T,S)-sequence parameters, in *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, editors: H. Niederreiter and J-S. Shiue, 58-86, Springer, New York.

Niederreiter, H. (1992) *Random number generation and quasi-Monte Carlo methods*, 63, CBMS-NSF Regional Conference Series in Applied Math., SIAM, Philadelphia, PA.

Niederreiter, H. (1995) New developments in uniform psuedo-random number and vector generation, in *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, editors: H. Niederreiter and J-S. Shiue, 94-112, Springer, New York.

Owen, A.B. (1995) Randomly permuted (t,m,s)-nets and (t,s)-sequences, in *Monte Carlo Methods in Scientific Computing*, editors: H. Niederreiter and J-S. Shiue, 299-317, Springer, New York.

Owen, A.B. (1997) Scrambled net variance for integrals of smooth functions, *The Annals of Statistics*, 25, 1541-562.

Owen, A.B. (1998) Latin supercube sampling for very high dimensional simulations, *ACM Transactions on Modeling and Computer Simulation*, 8, 71-102.

Papageorgiou, A. and J.F. Traub (1997) Faster evaluation of multidimensional integrals, *Risk*, 9, 63-65.

Paskov, S.H. (1997) New methodology for valuing derivatives, in *Mathematics of Derivative Securities*, editors: S. Pliska and M. Dempster, Isaac Newton Institute, Cambridge University Press, Cambridge.

Press, W.H., S.A. Teukolsky and M. Nerlove (1992) Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, Massachusetts

Revelt, D. and K. Train (1998) Mixed logit with repeated choices: households' choices of appliance efficiency level, *forthcoming*, *Review of Economics and Statistics*.

Sloan, I.H. and S. Joe (1994) Lattice methods for multiple integration, Clarendon Press, Oxford University Press

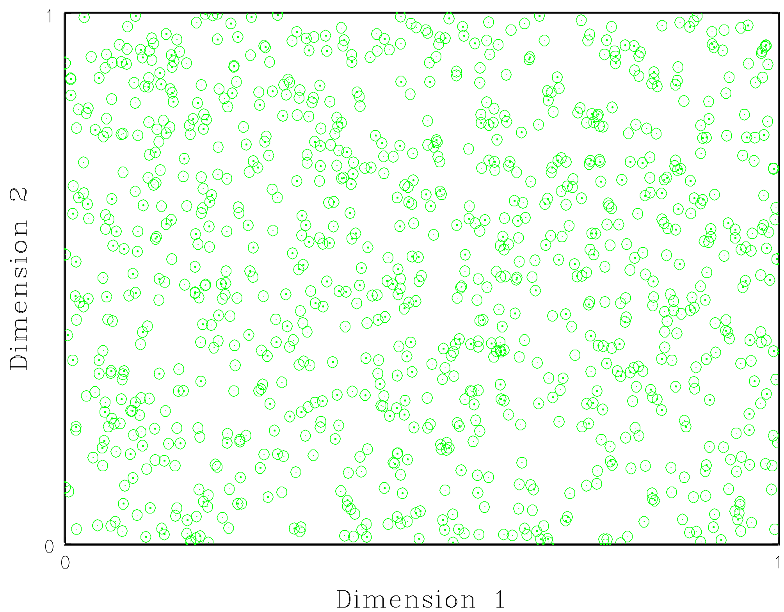
Sloan, J.H. and H. Wozniakowski (1998) When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?, *Journal of Complexity*, 14, 1-33.

Spanier, J. and E.H. Maize (1994) Quasi-random methods for estimating integrals using relatively small samples, *SIAM Review*, 36, 18-44.

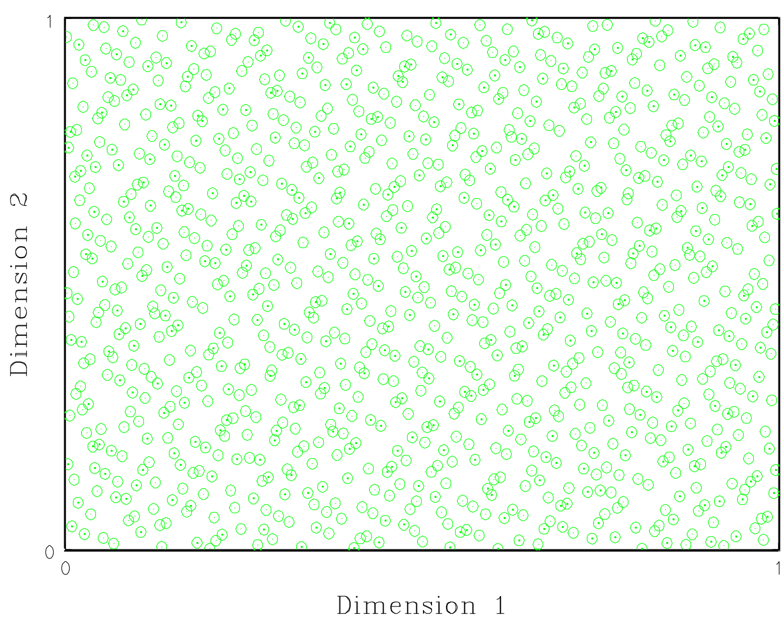
Stroud, A.H. (1971) Approximate Calculation of Multiple Integrals, Prentice-Hall, Englewood Cliffs, New Jersey.

Train, K. (1998) Recreation demand models with taste differences over people", *Land Economics*, 74, 230-239 (1998).

Wozniakowski, H. (1991) Average case complexity of multivariate integration, *Bulletin of the American Mathematical Society*, 24, 185-194.



Pseudo-random sequence



Halton sequence

Fig.1. 1000 Draws in Two Dimensions from the Uniform Distributiion.

Table 1. Maximum Simulated Likelihood Estimation Results for One Dimensional Integration

Evaluation Basis	Performance Measure	Pseudo-Random Monte Carlo (PMC) Estimation				Quasi-Random Monte Carlo (QMC) Estimation		
		Number of draws				Number of draws		
		250	500	1000	2000	25	50	75
Parameters	MAPE[1]	2.435	1.255	0.577	0.548	1.429	0.382	0.143
	RMSE[2]	0.089	0.047	0.024	0.021	0.050	0.007	0.004
Choice probabilities	MAPE	3.681	2.524	1.782	1.302	3.680	1.722	0.877
	RMSE	0.010	0.007	0.005	0.004	0.010	0.004	0.002
Time to convergence (mins)		40.8	71.8	143.0	330.3	6.1	9.4	14.5

[1]Mean Absolute Percentage Error (MAPE), [2]Root Mean Square Error

Table 2. Maximum Simulated Likelihood Estimation Results for Two Dimensional Integration

Evaluation Basis	Performance Measure	Polynomial-Based Cubature Estimation			Pseudo-Random Monte Carlo (PMC) Estimation				Quasi-Random Monte Carlo (QMC) Estimation		
		Number of cubature points			Number of draws				Number of draws		
		2	4	10	250	500	1000	2000	25	50	75
Parameters	MAPE	17.640	6.923	0.867	6.857	1.237	1.632	1.114	2.891	1.315	0.695
	RMSE	0.485	0.194	0.026	0.233	0.049	0.041	0.042	0.144	0.037	0.016
Choice probabilities	MAPE	6.162	2.356	1.815	4.852	3.218	2.331	1.784	4.851	2.517	1.652
	RMSE	0.015	0.009	0.004	0.014	0.010	0.007	0.005	0.014	0.007	0.004
Time to convergence (mins)		1.6	5.9	146.3	42.8	103.4	200.5	345.3	7.0	13.6	16.8

Table 3. Maximum Simulated Likelihood Estimation Results for Three Dimensional Integration

Evaluation Basis	Performance Measure	Polynomial-Based Cubature		Pseudo-Random Monte Carlo (PMC) Estimation				Quasi-Random Monte Carlo (QMC) Estimation				
		Number of cubature points		Number of draws				Number of draws				
		2	4	250	500	1000	2000	25	50	75	100	125
Parameters	MAPE	5.186	4.185	10.604	3.253	5.021	2.651	6.292	5.117	2.542	3.736	0.585
	RMSE	0.167	0.123	0.322	0.106	0.163	0.063	0.148	0.100	0.088	0.098	0.013
Choice probabilities	MAPE	3.532	1.815	5.403	3.649	2.626	1.787	5.846	3.565	2.599	1.768	1.411
	RMSE	0.012	0.006	0.015	0.011	0.008	0.005	0.017	0.010	0.008	0.005	0.004
Time to convergence (mins)		3.0	65.0	48.4	99.5	193.0	409.0	7.4	14.2	21.3	32.7	35.7

Table 4. Maximum Simulated Likelihood Estimation Results for Four Dimensional Integration

Evaluation Basis	Performance Measure	Polynomial-Based Cubature Estimation	Pseudo-Random Monte Carlo (PMC) Estimation				Quasi-Random Monte Carlo (QMC) Estimation				
		Number of cubature points	Number of draws				Number of draws				
		2	250	500	1000	2000	25	50	75	100	125
Parameters	MAPE	17.563	18.970	10.036	11.975	9.044	21.189	29.171	18.248	10.796	4.32
	RMSE	0.736	0.517	0.225	0.253	0.213	0.478	0.758	0.426	0.223	0.167
Choice probabilities	MAPE	5.113	5.493	4.133	2.857	2.129	7.499	4.373	3.655	2.586	2.071
	RMSE	0.015	0.017	0.012	0.009	0.006	0.022	0.012	0.010	0.007	0.006
Time to convergence (mins)		8.5	60.6	113.7	236.3	455.0	10.4	23.3	30.3	35.8	44.9

Table 5. Maximum Simulated Likelihood Estimation Results for Five Dimensional Integration

Evaluation basis	Performance measure	Polynomial-Based Cubature Estimation	Pseudo-Random Monte Carlo (PMC) Estimation				Quasi-Random Monte Carlo (QMC) Estimation				
		Number of cubature points	Number of draws				Number of draws				
		2	250	500	1000	2000	25	50	75	100	125
Parameters	MAPE	41.77	37.121	13.435	18.354	14.17	56.685	52.433	29.000	12.804	6.657
	RMSE	0.742	0.687	0.360	0.306	0.221	0.945	0.791	0.424	0.240	0.134
Choice probabilities	MAPE	4.862	6.665	4.771	3.332	2.418	8.812	5.073	3.905	3.090	2.283
	RMSE	0.014	0.020	0.014	0.010	0.007	0.025	0.014	0.010	0.008	0.007
Time to convergence (mins)		26.9	73.7	121.7	283.7	501.0	14.5	24.0	36.1	47.9	59.3

List of Tables

List of Figures