Economics Bulletin

Volume 39, Issue 4

Further evidence on sparse grids-based numerical integration in the mixed logit model

Zsolt Sándor
Sapientia Hungarian University of Transylvania

Abstract

We study the performance of Gauss quadrature methods based on sparse grids for approximating integrals involved in mixed logit models. In Monte Carlo experiments we consider data generating processes in which consumer heterogeneity has low variance and data generating processes in which it has high variance. In the former case we find that, in line with previous literature, sparse grids produce very accurate estimates even when the number of points used for approximating integrals is small. However, in the latter case sparse grids yield biased estimates and are outperformed by quasi-Monte Carlo methods.

Citation: Zsolt Sándor, (2019) "Further evidence on sparse grids-based numerical integration in the mixed logit model", *Economics Bulletin*, Volume 39, Issue 4, pages 2726-2731
Contact: Zsolt Sándor - sandorzsolt@cs.sapientia.ro.
Submitted: July 25, 2019. **Published:** November 29, 2019.

1 Introduction

The mixed logit model is widely used because it models flexibly unobserved consumer heterogeneity, which can be recovered by maximum likelihood (henceforth ML) estimation. Its panel version accommodates repeated choice information in a random coefficient logit framework. We refer to Train (2009) and the references therein for a discussion of the main contributions and applications.

A well-known difficulty with this model is that only an approximation of the ML estimator can be computed because the log-likelihood involves integrals that cannot be computed analytically. To evaluate integrals of this type, typically Gauss quadrature or Monte Carlo integration or a more efficient version of these is used. So-called quasi-Monte Carlo (henceforth QMC) integration have been shown to outperform the Monte Carlo method (e.g., Bhat 2001; Sándor and Train 2004). In an influential work Heiss and Winschel (2008) propose Gauss quadratures based on sparse grids and find that they outperform Monte Carlo and a simple type of QMC (i.e., Latin hypercube).

Heiss and Winschel (2008) evaluate the performance of sparse grid integration (henceforth SGI) in Monte Carlo experiments. We complement their experimental setting by considering higher variances of the random coefficients. High variances may occur in practice in markets where consumer heterogeneity is pronounced. In order to have a more complete picture of SGI, it is important to evaluate its performance in such situations. As a main contribution we find that in such situations QMC methods outperform SGI and we also provide a possible explanation for this phenomenon.

2 Model and estimation

Suppose that each consumer $i = 1, ..., n$ chooses one alternative from each of S different choice sets. Each choice set has J alternatives. Let $y_{js}^i \in \{0, 1\}$ denote the choice of consumer i in choice set s, that is, $y_{js}^i = 1$ if and only if alternative j is chosen. The utility of i corresponding to alternative j in s is assumed to be

$$u_{ijs} = x_{js} \left(\boldsymbol{\beta} + V_i \boldsymbol{\sigma}\right) + \varepsilon_{ijs},$$

where x_{js} is a $1 \times K$-vector of attributes of alternative j in choice set s, $\boldsymbol{\beta} + V_i\boldsymbol{\sigma}$ is the random coefficient with unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ that are $K \times 1$-vectors and V_i being a $K \times K$ diagonal matrix having standard normal random variables v_i on its main diagonal, and ε_{ijs} is a type I extreme value distributed error term.

The likelihood corresponding to consumer i is

$$l_i \left(y^i | \theta\right) = \int_{\mathbb{R}^K} q_i \left(\theta | v\right) d\Phi \left(v\right) \quad \text{with} \quad \theta = \left(\boldsymbol{\beta}', \boldsymbol{\sigma}'\right)', \tag{1}$$

$$q_i \left(\theta | v\right) = \prod_{s=1}^{S} \prod_{j=1}^{J} p_{js}^{y_{js}^i} \left(\theta | v\right) \quad \text{and}$$

$$p_{js} \left(\theta | v_i\right) = \frac{\exp\left(x_{js}\left(\boldsymbol{\beta} + V_i\boldsymbol{\sigma}\right)\right)}{\sum_{h=1}^{J} \exp\left(x_{hs}\left(\boldsymbol{\beta} + V_i\boldsymbol{\sigma}\right)\right)},$$

where Φ is the cumulative distribution function of the K-dimensional standard normal distribution. This yields the log-likelihood for the n consumers

$$L(y|\theta) = \sum_{i=1}^{n} \ln\left(l_i\left(y^i|\theta\right)\right), \tag{2}$$

the maximum of which with respect to θ is the ML estimator. As already mentioned, integrals of type (1) involved in the log-likelihood cannot be computed analytically.

3 Computation of the likelihood

Gauss quadrature and Monte Carlo integration are two well-established methods for approximating integrals of type (1). Here we briefly present these and their versions used in the paper.

3.1 Gauss quadrature

Gauss quadrature approximates the integral (1) with the (deterministic) formula

$$\sum_{r=1}^{R} w_r q_i(\theta|z_r),$$

where z_r are abscissas, which are predetermined points in $\mathbb{R}^K$, and $w_r \in \mathbb{R}$ are their weights, $r = 1, ..., R$. Sparse grids are sets of abscissas that are especially suited for multi-dimensional integrals (for details we refer to Heiss and Winschel 2008).

A drawback of SGI is that some weights can be negative, which may cause the integral approximation to be negative (Heiss and Winschel 2008). When in the mixed logit the variance of the random coefficients is relatively large, this phenomenon is more likely to occur. In order to deal with this issue we replace the logarithm in the log-likelihood (2) by the function

$$\ln_*(x) = \begin{cases} \ln(x), & x > d, \\ \ln(d) - 1.5 + \frac{2x}{d} - \frac{x^2}{2d^2}, & x \leq d, \end{cases} \quad d > 0, \tag{3}$$

which is strictly increasing and is twice continuously differentiable on $\mathbb{R}$. We take $d = \exp(-20)$ in our computations, so on $(0, \infty)$ the $\ln_*(\cdot)$ function differs from the logarithm only for very small values.

3.2 Monte Carlo integration

Monte Carlo integration uses the stochastic approximation

$$\frac{1}{R}\sum_{r=1}^{R} q_i(\theta|v_r),$$

where $v_r,\ r = 1, ..., R$ is a sample generated from the underlying distribution, which is the K-dimensional standard normal in our case. This approximation is an unbiased estimator of the integral (1). Several methods have been proposed to reduce the variance of this estimator. One such method is QMC integration.

QMC integration uses a deterministic sample constructed in the unit hypercube, which is transformed by the inverse cumulative distribution function. One class of such quasi-random samples, called digital nets, has the property that the points of its low-dimensional projections fill the corresponding unit hypercubes evenly. The simplest example of a digital net is a Latin hypercube, which is a $(0, 1, s)$-net, that is, due to the regular grid structure of a one-dimensional projection of a Latin hypercube its points cover the $[0, 1]$ interval evenly. A so-called $(0, 2, s)$-net has the property that its one- and two-dimensional projections cover the $[0, 1]$ and $[0, 1] \times [0, 1]$ intervals evenly, respectively.

4 Results

We compare the performance of SGI and QMC in a Monte Carlo simulation experiment. The data generating process and estimation are similar to those in Heiss and Winschel (2008). The characteristics vector x_{js} is generated as a $1 \times K$ vector of uniforms on $[0, 1]$, the number of choice situations is $S = 5$, the number of alternatives within a choice situation is $J = 5$, and the number of consumers is $n = 1000$. We consider two values for the dimension of the characteristics vector, namely, $K = 5$ and 10. The parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$, similar to Heiss and Winschel, have equal components, that is, $\boldsymbol{\beta} = (\beta, ..., \beta)'$, $\boldsymbol{\sigma} = (\sigma, ..., \sigma)'$, where $\beta = 1$ in each case while σ takes two values, namely, 0.5 and 3. In the estimation, similar to Heiss and Winschel, only the scalars β and σ are estimated.

For SGI we use abscissas and weights corresponding to Kronrod-Patterson rules.[2] For QMC integration we use Latin hypercube samples and $(0, 2, s)$-nets as well as $(0, 3, s)$-nets constructed in the way described in Sándor and András (2004). The bias and root mean squared error (henceforth RMSE) reported are based on 100 replications over the product characteristics.

Table 1 presents results for $\sigma = 0.5$ for SGI, Latin hypercubes and digital nets (denoted LH and Net in the tables, respectively). In the 5-dimensional case SGI with only $R = 11$ points produce estimates that have very low bias and low RMSE. The Latin hypercube of size 11 produces both higher biases and RMSE's.[3] The performance of SGI with more points ($R = 51$ and 401) is not better than that with $R = 11$ and is rather comparable to LH of size $R = 51$ and 401 and Net of size $R = 49$ (a $(0, 2, s)$-net) and 343 (a $(0, 3, s)$-net). The relative performance of the methods shows a similar pattern in the 10-dimensional case: the SGI with $R = 21$ points matches the performance of the methods with more points (i.e., LH of size $R = 201$ and 1201 and Net of size $R = 169$

[2] Downloaded from the web page `http://www.sparse-grids.de/` created as a companion to the paper by Heiss and Winschel (2008).

[3] We note here that there is no $(0, 2, s)$-net of size 11 or comparable. In Section 5 we argue that this is not an important drawback (see also Chiou and Walker, 2007).

(a $(0,2,s)$-net) and 1331 (a $(0,3,s)$-net)). Again, LH of size $R = 21$ produces estimates with higher biases and RMSE's. We note that these results are very much in line with those obtained by Heiss and Winschel (2008, Table 5) for SGI and LH.

Table 1. Bias and RMSE when the standard deviation is 0.5

	SGI		LH		Net	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
			dimension 5			
	$R = 11$		$R = 11$			
β	-0.002	0.030	-0.010	0.032	–	–
σ	-0.013	0.079	-0.085	0.118	–	–
	$R = 51$		$R = 51$		$R = 49$	
β	-0.002	0.030	-0.001	0.033	-0.004	0.029
σ	-0.005	0.080	-0.026	0.081	-0.042	0.092
	$R = 401$		$R = 401$		$R = 343$	
β	-0.004	0.030	-0.003	0.028	0.001	0.030
σ	-0.013	0.094	-0.020	0.088	-0.005	0.081
			dimension 10			
	$R = 21$		$R = 21$			
β	-0.002	0.025	-0.020	0.032	–	–
σ	-0.023	0.064	-0.089	0.115	–	–
	$R = 201$		$R = 201$		$R = 169$	
β	0.004	0.026	0.001	0.024	-0.006	0.024
σ	0.006	0.068	-0.012	0.068	-0.023	0.068
	$R = 1201$		$R = 1201$		$R = 1331$	
β	-0.002	0.025	-0.000	0.029	0.002	0.027
σ	-0.006	0.068	-0.010	0.073	-0.005	0.079

The situation is quite different when $\sigma = 3$, as shown in Table 2. Both bias and RMSE are higher in general. It is rather striking that SGI are outperformed both in terms of bias and RMSE by the other two methods. Although the performance of SGI improves as the number of points reaches $R = 401$ (5 dimensions) and $R = 1201$ (10 dimensions), the estimates are still rather biased. For such sample sizes or comparable (specifically, $R = 343$ and 1331 for Net) LH and Net produce estimates with small biases and RMSE's.

The explanation for the relative poor performance of SGI is that for larger σ there are more consumers for which integrals of type (1) involved in the likelihood are approximated to be negative, so the likelihood will be computed with larger error, which is eventually reflected in the ML estimator. In order to illustrate this, in a small Monte Carlo study we estimated the likelihood components (1) and their logarithms at the true parameter values by using SGI and Net and compared their performance based on their relative magnitude with respect to the 'true' values of (1) computed by Monte Carlo based on a sample of size $100,000$. Then we computed RMSE's based on 100 replications. The RMSE's corresponding to the likelihood components are 0.000002 (SGI) and 0.000043 (Net) when $\sigma = 0.5$, and 0.00085 (SGI) and 0.00111 (Net) when $\sigma = 3$, that is, in both cases SGI outperforms Net. The RMSE's corresponding to the logarithms of the likelihood components are 0.0015 (SGI) and 0.0216 (Net) when $\sigma = 0.5$, and 2.889 (SGI) and 0.427 (Net) when $\sigma = 3$, so in the latter case SGI is outperformed by Net. In the case $\sigma = 0.5$ these findings are in line with the results obtained (in Table 1) for the parameter estimates. When $\sigma = 3$ the results obtained for the logarithms are in line with those obtained for the parameter estimates (see Table 2), but for the likelihood components SGI outperforms Net. We argue that this is due to the fact that in 5 out of the 100 replications the estimates of (1) based on SGI are negative (as opposed to the case $\sigma = 0.5$ when all SGI estimates are positive). The phenomenon that some estimates of the likelihood components (1) are negative does not mean that they are imprecise (at least compared to Net), nevertheless, when computing the log-likelihood one needs to replace the logarithm by a function of type (3), and this introduces error in the log-likelihood. This error is then transmitted to the parameter estimates, as seen in Table 2. Recall that the estimators computed with LH or Net are not affected by this problem because the integrals are always estimated to be positive.

Table 2. Bias and RMSE when the standard deviation is 3

	SGI		LH		Net	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
			dimension 5			
	$R = 11$		$R = 11$			
β	-0.326	0.334	-0.374	0.377	–	–
σ	-1.467	1.469	-1.046	1.047	–	–
	$R = 51$		$R = 51$		$R = 49$	
β	-0.169	0.182	-0.079	0.094	-0.210	0.217
σ	-0.981	0.985	-0.462	0.467	-0.289	0.298
	$R = 401$		$R = 401$		$R = 343$	
β	-0.099	0.121	-0.046	0.076	-0.020	0.063
σ	-0.456	0.465	-0.050	0.100	-0.049	0.106
			dimension 10			
	$R = 21$		$R = 21$			
β	-0.454	0.458	-0.430	0.431	–	–
σ	-1.907	1.908	-1.644	1.645	–	–
	$R = 201$		$R = 201$		$R = 169$	
β	-0.353	0.356	-0.211	0.217	-0.273	0.276
σ	-1.758	1.759	-0.817	0.819	-0.876	0.878
	$R = 1201$		$R = 1201$		$R = 1331$	
β	-0.373	0.376	-0.080	0.098	-0.065	0.085
σ	-1.556	1.558	-0.292	0.309	-0.234	0.262

5 Conclusions

In this paper we conduct Monte Carlo experiments to investigate Gauss quadrature methods based on sparse grids for approximating integrals involved in panel mixed logit models. In doing so we complement the literature by considering data generating processes in which consumer heterogeneity has high variance. We find that in this case QMC methods (Latin hypercubes and digital nets) yield estimates of the parameters that are closer to the truth. This has important consequences for the estimation of panel mixed logit models in practice because if the underlying consumer heterogeneity turns out to be pronounced, estimation based on a small number of points for evaluating integrals may yield misleading results. As a safeguard against such situations we recommend using a large digital net (preferably a $(0, 3, s)$-net) for approximating integrals involved in panel mixed logit likelihoods.

These conclusions are in line with the idea that empirical identification of the parameters may depend on the number of points used for integration. Several papers in the literature starting with Chiou and Walker (2007) recommend using a large number of points for the evaluation of integrals in order to detect identification problems that can yield misleading estimates (see also, e.g., Brunner et al. 2017).

Our results also confirm previous findings by Heiss and Winschel (2008) that sparse grids provide accurate approximations to integrals when consumer heterogeneity has low variance. It is important to note that sparse grids can potentially be useful even when consumer heterogeneity has large variance within estimation procedures where the approximated integrals are allowed to take negative values (e.g., method of moments).

References

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

Brunner, D., F. Heiss, A. Romahn, and C. Weiser (2017) "Reliable estimation of random coefficient logit demand models" Heinrich Heine Universität Düsseldorf discussion paper number 267.

Chiou, L. and J.L. Walker (2007) "Masking identifi cation of discrete choice models under simulation methods" *Journal of Econometrics* **141**, 683–703.

Heiss, F. and V. Winschel (2008) "Likelihood approximation by numerical integration on sparse grids" *Journal of Econometrics* **144**, 62-80.

Sándor, Z. and P. András (2004) "Alternative sampling methods for estimating multivariate normal probabilities" *Journal of Econometrics* **120**, 207–234.

Sándor, Z. and K.E. Train (2004) "Quasi-random simulation of discrete choice models" *Transportation Research* B **38**, 313–327.

Train, K.E. (2009) *Discrete Choice Methods with Simulation*, Cambridge University Press: New York.