Written by
Doulo SOW
-

M2 in Economics
IP-Paris
2022-2023
-

Theme

Approximation of integrals

Quadrature (Hermite) Methods vs Monte Carlo Simulation Method applied in Mixed Logit Models to estimate Maximum Likelihood

Course
CREST,[a]
Advanced Methods in Computational Economics
-

[a] *Center for Research in Economics and Statistics*

Professor
Simon Scheidegger,
HEC, University of Lausanne
-

1 Project Presentation and Theoretical Framework

1.1 Presentation

Maximum likelihood estimation (MLE) is one of the most common methods econometricians use to estimate the parameters of a given model. In short, let $x, \ldots, x_n$ be observations from n independent (not necessary) and identically distributed random variables drawn from a Probability Distribution f_{θ_0} which is know to be from a family of distribution, depending on a parameter θ, i.e $\mathcal{F} = \{f(\cdot;\theta) : \theta \in \Theta\}$. The goal of MLE is to maximize the likelihood function:

$$\mathcal{L}(\theta) = f(x, \ldots, x_n; \theta)$$

Although for some types of models this likelihood function can be easily computed, there are many other models whose likelihood functions involve multidimensional integrals with no analytically tractable solution. Therefore, a simple way to approximate this complicated likelihood function is needed to estimate these models. In this project, I use mixed logit models (inspired by Heiss et al., 2008) with some simplifications, to investigate (and try to understand) how to approximate these likelihood functions using the quadrature method and Monte Carlo simulation.

1.2 Mixed Logit Models

Assume a population of N individuals $i = 1, \ldots, N$. For each of these individuals, we observe T choice situations. For each $t = 1, \ldots, T$ choice situation, an individual i makes a choice j among J available choices. To make a choice when facing J possible choices, the individual is supposed to pick up one of them that gives the highest utility: Random Utility Models framework. The unobserved utility is supposed to be explained by a set of K-dimensional observed characteristics and additional unobserved random i.i.d variables: $u_{itj} = x'_{itj}\beta_i + \epsilon_{itj}$ where u_{itj} the utility provided by the choice j at the choice situation t to the individual i and β_i corresponds to the taste levels and varies across individuals. Furthermore, β_i are supposed to be distributed according to a parameter θ with p.d.f $f(\beta_i;\theta)$ with support $\Psi \subseteq \mathbf{R}^K$. The main goal is then to estimate the parameter of taste levels θ.

Let's define: $y_{itj} = \begin{cases} 1 \text{ if i choices j at the t choice situation} \\ 0 \text{ otherwise} \end{cases}$

For the sack of simplicity, I suppose:

- one situation of choice: T=1;
- two types of choices: J=2;
- $x'_{i2} - x'_{i1} = 1'_K = (1, \ldots, 1)$

In general, the probability that the underlying random variable $Y_i = [y_{itj}; t = 1, \ldots, T; j = 1, \ldots, J]$ equals the observed choices y_i conditional on the characteristics $x_i = [x_{itj}; t = 1, \ldots, T; j = 1, \ldots, J]$

and the taste level β_i is given by:

$$\mathbf{P}^*(\beta_i) = \mathbf{P}(Y_i = y_i | x_i, \beta_i) = \prod_{t=1}^{T} \frac{\prod_{j=1}^{J} e^{(x'_{itj}\beta_i y_{itj})}}{\sum_{j=1}^{J} e^{x'_{itj}\beta_i}} \tag{1}$$

With the simplification assumptions above and noticing that in such case $y_{i2} + y_{i1} = 1$, I show that this formula can be expressed as following (what, by the way, corresponds to the Logit Model with random parameters β_i):

$$\mathbf{P}^*(\beta_i) = \frac{e^{-1'_K \beta_i y_{i1}}}{1 + e^{-1'_K \beta_i}}$$

The likelihood contribution of individual i is therefore equal to the joint probability as a function of θ and is expressed as following:

$$\mathbf{P}_i(\theta) = \int_{\Psi} \mathbf{P}^*(\beta_i) f(\beta_i;\theta) d\beta_i = \int \cdots \int \mathbf{P}^*(\beta_i) f(\beta_i;\theta) d\beta_{i1} \dots d\beta_{iK}$$

2 Approximation

2.1 Univariate Approximation

Before going to an approximation of the multivariate form, I suppose (as done in the main paper) that I observed one explanatory feature x i.e K=1. Furthermore, assuming that the taste level, which is now a scalar, has a normal distribution with mean μ and variance σ^2 and by variable changing ($z = \frac{\beta_i - \mu}{\sigma}$), the likelihood contribution is given by:

$$\mathbf{P}_i(\theta) = \int_{\mathbf{R}} g(z)\phi(z) dz = \mathbf{E} g(Z), Z \sim \mathcal{N}(0,1)$$

, where $g(z) = \frac{e^{-(\mu+\sigma z) y_{i1}}}{1+e^{-(\mu+\sigma z)}}$[1] To get rid of the y_{i1}, one can suppose to focus on the probability that the individual i chooses a specific choice (e.g, choice 2 i.e $y_{i1} = 0$). This gives the same results (observation equivalence) due to the fact that I suppose two choices and 1 situation of choice. Therefore knowing the probability of chosing y_{i1} is equivalent to know that of $y_i = [y_{i1}, y_{i1}]$. Subsequently, I then consider $g(z) = \frac{1}{1+e^{-(\mu+\sigma z)}}$. To find analogy with the original paper of Heiss et al., 2008, this case corresponds to that of $T_1 = 0$ and $T_2 = 1$.
Now that I have defined the theoretical framework, I can either simulate the probability using the Monte Carlo method or approximate it using the standard Gaussian quadrature method.

2.1.1 Monte Carlo method

Using R simulation draws, I represent below how the Monte Carlo method approximates the integral. The exact value of the integral is given by the value of the area delimited by the blue

[1] Note that the same formula can be drawn by analogy from the original article where, contrary to my assumptions, the authors gave the formula as a function of the number of times a choice is chosen.

curve line and x-axis. The estimation made fron the Monte Carlo corresponds to the sum of R rectangles each of which has a width of $1/R$ and a height that corresponds to the function value at randomly chosen points. This approximation turns arround 0.70 (see notebook attached). Obviously and as expected the approximation becomes more and precise when I increase the number of draws. For instance, with order n=4, the estimation is about 0.72 quite close to the one obtained from the Monte Carlo Metrho (See notebook for details).

Figure 1: Monte Carlo Results with small numbers of simulation

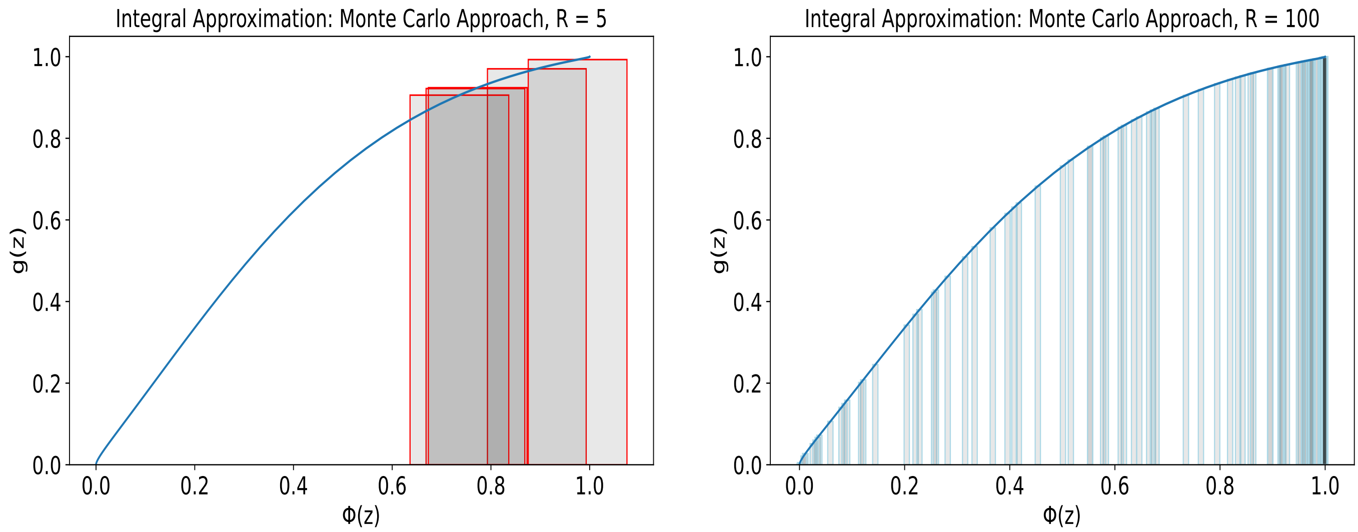
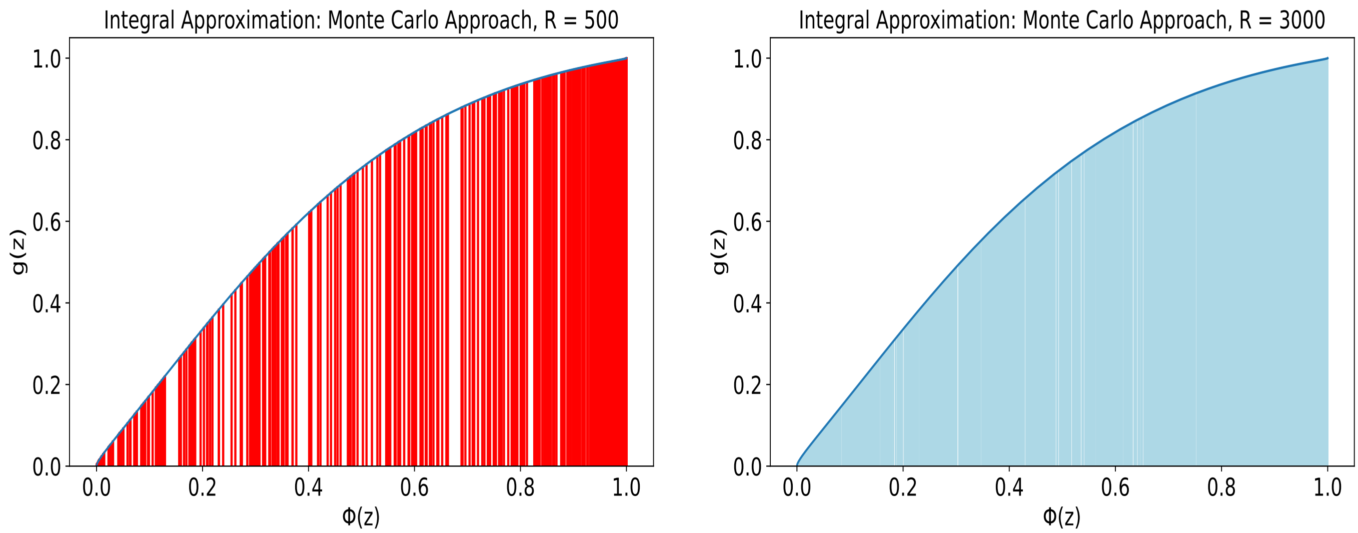


Figure 2: Monte Carlo Results with high numbers of simulation



2.1.2 Gauss-Hermite Quadrature

Quadrature exactly integrates a polynomial of a given degree that represents an approximation to the integrand.

$$\int_{-\infty}^{\infty} f(x)\exp(-x^2)dx = \sum_{i=1}^{n} \omega_i f(x_i) + \frac{n!\sqrt{\pi}}{2^n}\frac{f^{(2n)}(\xi)}{(2n)!}$$

I use Nodes and weights that come from Hermite polynomials (i.e:- Domain $[-\infty, \infty]$ - Weighting $\exp(-x^2)$) to approximate the integrand defined as $\int_{-\infty}^{\infty} g(\mu + \sqrt{2}\sigma z)e^{-(z)^2}dz$ where $g(z) = \pi^{\frac{-1}{2}}(1 + e^{-z})^{-1}$

That is:

$$\int_{-\infty}^{\infty} g(\mu + \sqrt{2}\sigma z)e^{-(z)^2}dz \approx \sum_{i=1}^{n} \omega_i g(\sqrt{2}\sigma x_i + \mu)$$

2.2 Multivariate Approximation

In this section, I consider the case where the number of features equals 2. Therefore, the taste level $\beta_i = (\beta_{i1}, \beta_{i2})'$ is bivariate normal distributed. I assume the two component are correlated and have the same marginal distribution. Then,

$$\beta_i \sim \mathcal{N}((\mu, \mu)', \Sigma = \begin{pmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{pmatrix}),$$

In this case, our integral of interest becomes:

$$\mathbf{P}_i(\theta) = \int \int \frac{1}{1 + e^{-(\beta_{i1}+\beta_{i2})}} f(\beta_i; \theta) d\beta_{i1} d\beta_{i2}$$

2.2.1 Monte Carlo method

Figure 3: Estimated Integrand with Monte Carlo

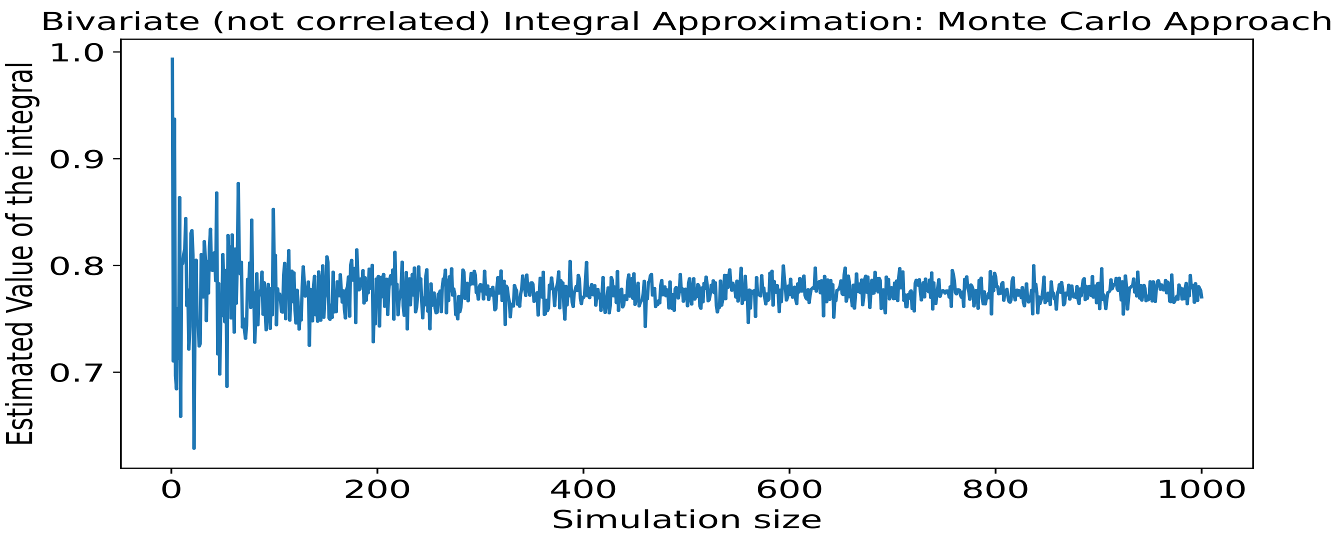
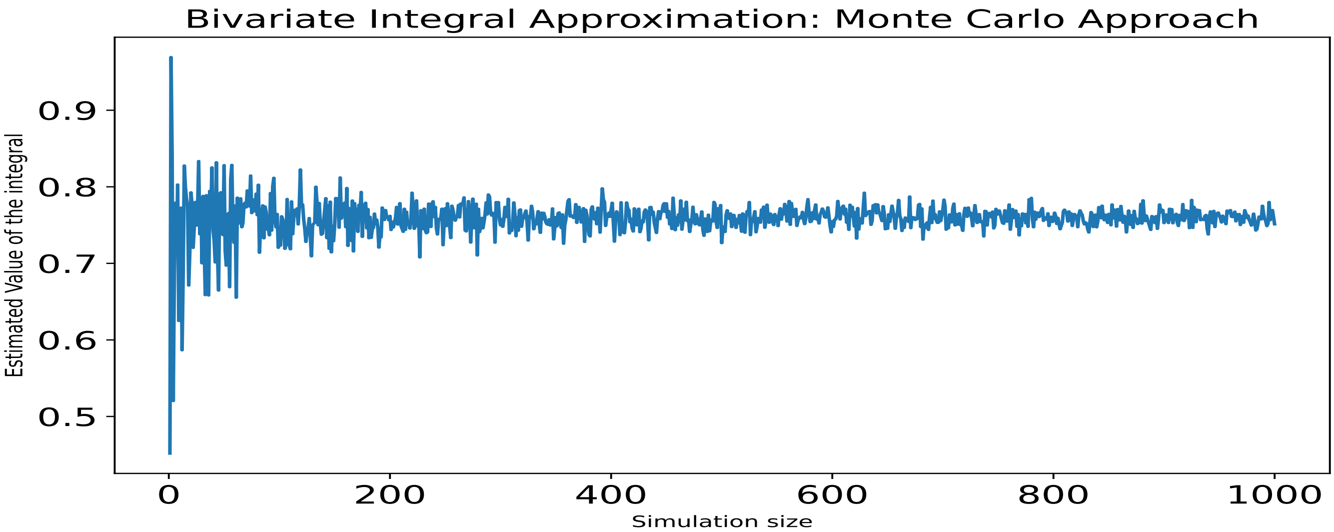
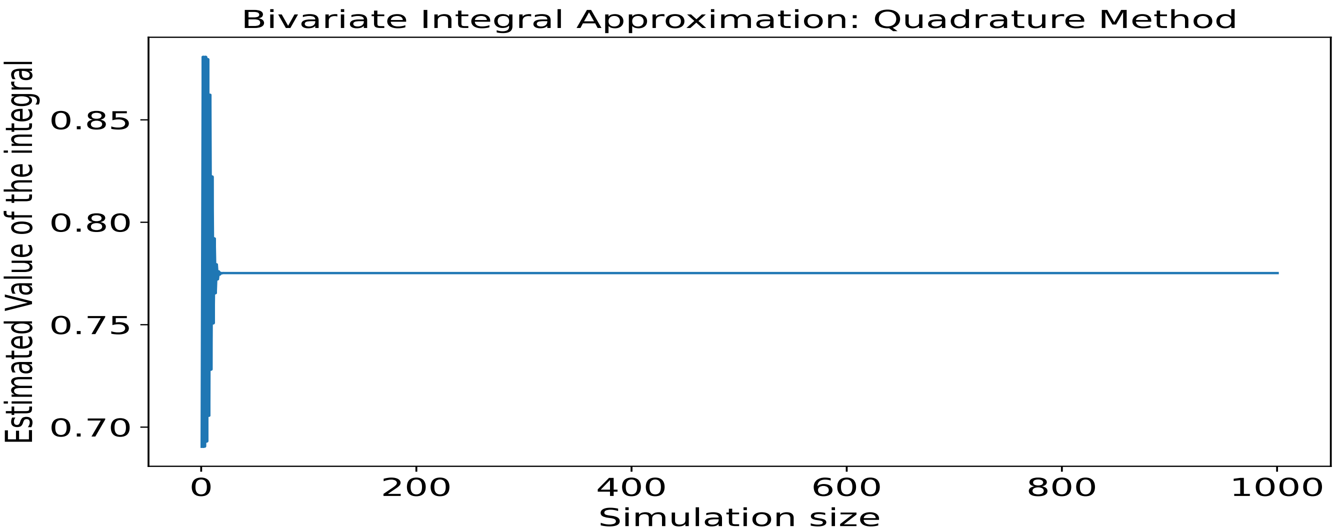


Figure 4: Estimated Integrand with Monte Carlo



2.2.2 Quadrature Method

Figure 5: Estimated Integrand with Quadrature Method



Appendix

Figure 6: Different types of quadrature

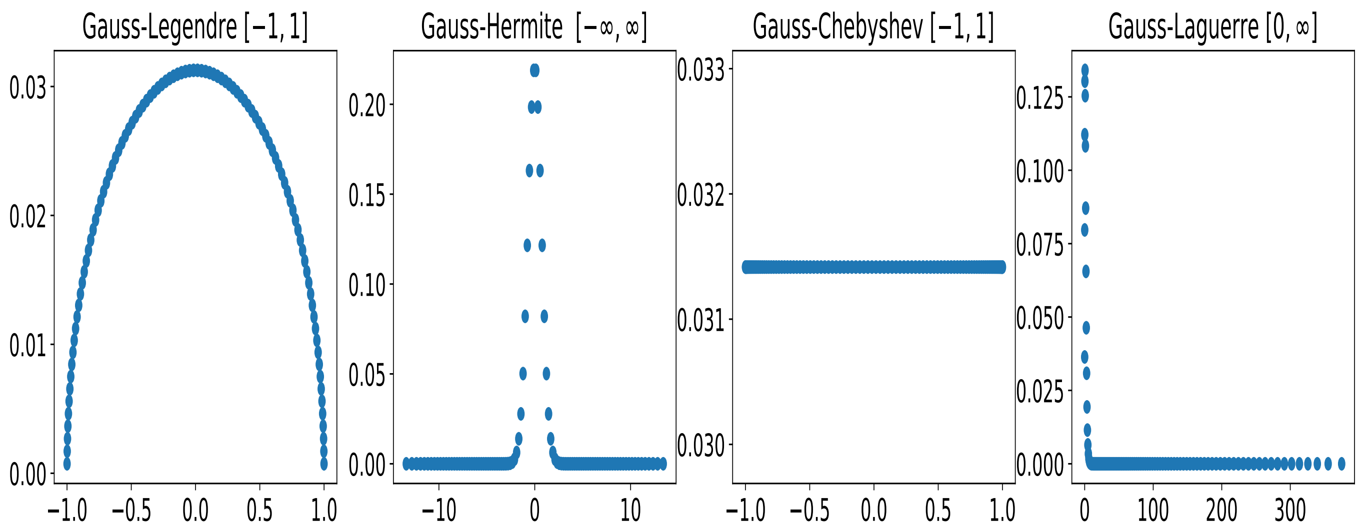


Figure 7: Bivariate Normal Distribution

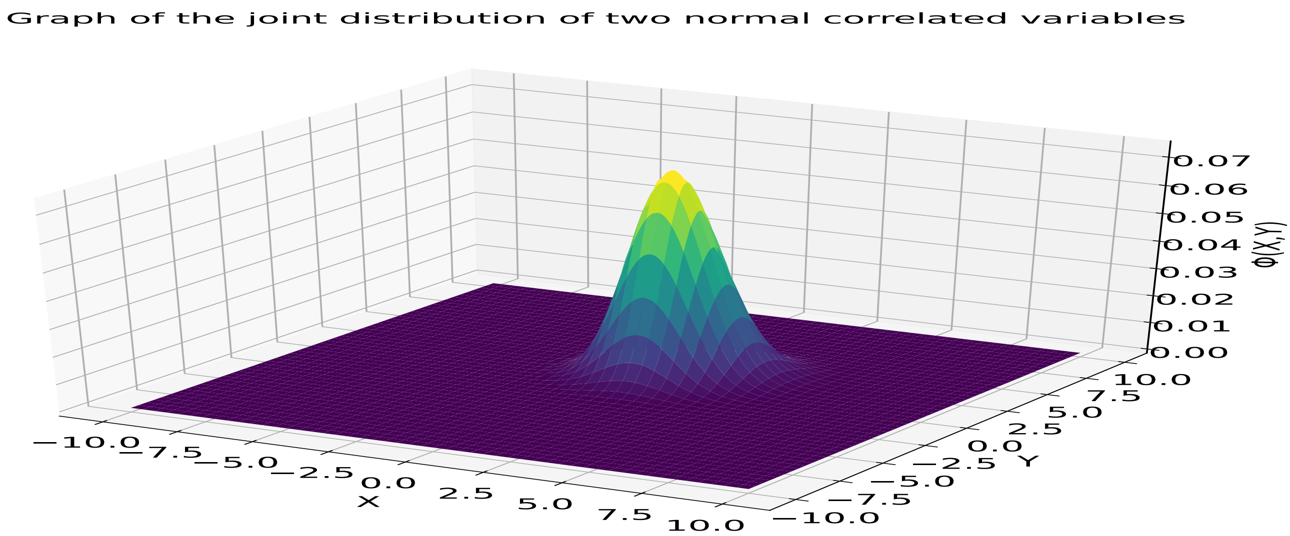


Figure 8: Linear Interpolation