

Numerical Integration Using Monte Carlo and Quasi-Monte Carlo Methods

- There are many problems in economics and econometrics for which the evaluation of high dimensional integrals is essential. If we have to compute an integral in more than two dimensions, Monte Carlo based methods are the only ones really viable.
- Monte Carlo methods were developed in the 1940s by von Neumann, Ulam, Metropolis, and others, primarily motivated by problems in nuclear physics such as neutron diffusion in connection with the development of the atom bomb. Since this research was top secret, it was given a code name appropriate for a process dealing with random events: Monte Carlo, referring to the world-famous gambling casino. The name persisted, and was for a time used to refer to any simulation effort.
- Von Neumann is credited with one of the earliest arithmetic algorithms to produce random numbers, using a middle-square method. In this algorithm random numbers are generated by extracting the middle digits from the square of a previously-generated number. This was widely used by those pioneers of MC methods, but it had the undesirable property of degenerating when the middle digits are close to zero, and the method is slow.
- The idea of Monte Carlo integration is simple, the function is sampled at n points distributed randomly in the domain of integration, and then the mean of these function values is multiplied by the area (or volume, etc.) of the domain to obtain an estimate for the integral. The error in this estimate goes to zero as \sqrt{n} , which means that, for example, to gain a decimal place of accuracy the number of sample points must be increased by a factor of 100.
- Monte Carlo integration is not competitive in integrals in one or two dimensions, but its strength is that its convergence rate is independent of the number of dimensions. As we will see below the efficiency of this method can be enhanced by various methods for biasing the sampling, either to achieve more uniform coverage of the sampled volume, or to concentrate sampling in regions where the integrand is largest in magnitude (importance sampling) or in variability (stratified sampling).

Monte Carlo Methods

- Monte Carlo integration methods are motivated by the Strong Law of Large Numbers. The Law basically states that if x_1, x_2, \dots, x_n are independent realizations of a random variable \bar{X} and f is a continuous function then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(x_i) = E f(\bar{X}) \quad (1)$$

with probability one. Furthermore,

$$\text{var} \left(N^{-1} \sum_{i=1}^N X_i \right) = \frac{\sigma_x^2}{N}, \quad (2)$$

where $\sigma_x^2 = \text{var}(X_1)$. If σ_x^2 is not known a priori we can use the unbiased estimator

$$\hat{\sigma}_x^2 = (N-1)^{-1} \sum_{i=1}^N (X_i - \bar{X})^2, \quad (3)$$

where \bar{X} is the sample mean.

The Monte Carlo integration scheme is thus a simple one: To compute an approximation to the expectation of $f(X)$, one draws a random sample x_1, x_2, \dots, x_n from the distribution of X and sets

$$E[f(X)] \approx 1/n \sum_{i=1}^n f(x_i). \quad (4)$$

If for example one wants to compute $I_f = \int_0^1 f(x)dx$, and $X \sim U[0, 1]$, then

$$E[f(X)] = \int_0^1 f(x)dx. \quad (5)$$

It is then easy to see that we just have to take N draws from $U[0, 1]$ to compute an estimate of the integral.

Monte Carlo procedures differs substantially from other quadrature based methods in that the approximation is itself a random variable, \hat{I}_f , with variance $\sigma_{\hat{I}_f}^2 = N^{-1}\sigma_f^2$. Since the variance of the function is unknown, we have to take an estimate $\hat{\sigma}_f^2$,

$$\hat{\sigma}_f^2 = (N - 1)^{-1} \sum_{i=1}^N (f(x_i) - \hat{I}_f)^2. \quad (6)$$

- Then, the standard error of \hat{I}_f is σ_f/\sqrt{N} . Since the Monte Carlo quadrature procedures yields a random variable, any estimate of \hat{I}_f should be accompanied with an estimate of $\sigma_{\hat{I}_f}^2$. Otherwise is like presenting a parameter estimate without reporting a confidence interval.
- Also, approximations generated by Monte Carlo integration will vary from one integration to the next, unless initiated at the same point, making the use of Monte Carlo integration in conjunction or within other iterative schemes, such as dynamic programming or Maximum Likelihood estimation, problematic.
- We have seen that using the probability integral transform we can basically draw from any distribution by using draws from a uniform.
- An important problem that arises with Monte Carlo integration is that it is almost impossible to generate a truly random sample for any distribution. For this reason, numerical random number generators are more accurately said to generate sequences of “pseudo-random” rather than random numbers.
- A good random number generator should have as many of the following properties as possible:
 - Random pattern: it should pass statistical tests of randomness, for example serial correlation tests.
 - Long period: it should go as long as possible before repeating.
 - Efficiency: it should execute rapidly and require little storage, since many simulations require millions of random numbers.

- Repeatability: it should produce the same sequence if started with the same initial conditions.
- Portability: it should run on different kinds of computers and be capable of producing the same sequence on each.
- Monte Carlo integration is easy to implement and may be preferred over Gaussian quadrature if the routine for computing Gaussian mass points and probabilities is not readily available or if the integration is over many dimensions.
- The crude Monte Carlo method is often used in conjunction with variance reduction techniques because its variance is too large. We will not get into details here but you can read Judd (1998, Ch. 8) for details. Methods often used are

- **Stratified Sampling:** The variance of f over an interval of $[0, 1]$ is often less than over the whole interval. Suppose we divide $[0, 1]$ into $[0, \alpha]$, and $[\alpha, 1]$. Then, if we have N points sampled over $[0, \alpha]$, and N over $[\alpha, 1]$, we form the estimate:

$$\hat{I}_f = \frac{\alpha}{N} \sum_i f(x_{1i}) + \left(\frac{1-\alpha}{N} \right) \sum_i f(x_{2i}), \quad (7)$$

where $x_{1i} \in [0, \alpha]$ and $x_{2i} \in [\alpha, 1], i = 1, \dots, N$. Notice that the main idea from stratified sampling is to keep the draws from clumping in one region. This feature is typical of acceleration schemes.

- **Antithetic Variates:** The idea is that if f is monotonically increasing, then $f(x)$ and $f(1-x)$ are negatively correlated. We can apply this idea to reduce the variance of \hat{I}_f which we can write as:

$$\hat{I}_f^a = \frac{1}{2N} \sum_{i=1}^N (f(x_i) + f(1-x_i)). \quad (8)$$

The antithetic estimate is an unbiased estimate of I_f but will have smaller variance than the crude estimate when the antithetic summands are negatively correlated.

- **Control Variates:** Suppose we know a function, ϕ , that is similar to f but easily integrated. Then $\int f = \int \phi + \int (f - \phi)$, reduces the problem to a Monte Carlo integration of $\int (f - \phi)$ plus the known integral $\int \phi$. The variance of this method will be smaller than the crude method if the covariance between f and ϕ is large.
- **Importance Sampling:** The crude method of Monte Carlo integration samples the draws unrelated to the integrand $f(x)$. This can be wasteful, since the value of f in some parts of $[0, 1]$ is much more important than in other parts. Importance Sampling tries to sample $f(x)$ where its value is most important in determining the integral $\int_0^1 f(x) dx$. If $p(x) > 0$ and in the $[0, 1]$ integrates to 1, then $p(x)$ is a density and

$$I_f = \int_0^1 f(x) dx = \int_0^1 \frac{f(x)}{p(x)} p(x) dx. \quad (9)$$

Which under the right conditions has a lower variance than the crude estimate. The cost of this method is that it may take some effort to construct a non-uniform random variable that has a density similar to $f(x)$.

Quasi-Monte Carlo Methods

- These methods do not rely on probability theory but on number theory to derive improved methods for calculating high dimensional integrals.
- In this case the sequences of numbers generated are not random, but deterministic. These sequences attempt to fill the space in a regular manner. These sequences exhibit improved convergence properties.
- A well established literature (Niederreiter 1992, and Traub and Wozniakowski 1992) validate these techniques, which have been shown to clearly outperform traditional (pseudo) Monte Carlo Methods. This is especially true when dimensions and sample sizes increase.
- The objective is to construct sequences that do well at integration. We use Monte Carlo methods because if we were able to draw random sequences we could invoke the LLN to obtain unbiased estimates of integrals (a sequence of random i.i.d. draws is almost surely equidistributed). But we can circumvent the randomness property by appropriately choosing the sequence, which are required to converge to the integral, but also all subsequences have to have this property. This is a very strong requirement.
- The most used equidistributed sequences are the Niederreiter, Weyl, and Haber. These are low discrepancy sequences, and that is why this method is sometimes called low-discrepancy integration.