CHAPTER 2: THE MONTE CARLO METHOD

The Monte Carlo method is a method of approximately solving mathematical and physical problems by the simulation of random quantities. The terminology "Monte Carlo" comes from the city of Monte Carlo in the principality of Monaco, famous for its gambling houses.

The computational algorithm is relatively simple in Monte Carlo calculations. The algorithm consists, in general, of a process for producing a random event. The process is repeated *N* times, each trial being independent of the others, and the results of all trials are averaged together to provide an *estimate* of the quantity of interest.
The process is similar to performing a scientific experiment and is sometimes called the method of stochastic, or statistical experiments or trials.

The error associated with the estimated quantity is, as a rule, inversely proportional to the square root of the number of trials. that is

$$\text{error} \propto (1/N)^{1/2}$$

It is clear that, to decrease the error by a factor of 10 (in order to obtain another significant digit in the result), it is necessary to increase N (and the computational effort) by a factor of 100. To attain high precision in Monte Carlo calculations is clearly impossible.

The Monte Carlo method is most effective in soiving problems in which the results need to be accurate to less than a few percent. It is important to point out here however that, unlike other deterministic methods, the Monte Carlo method provides an answer with an error associated with it, so that a confidence level in the result can be established.

The main advantage of the Monte Carlo method is its ability to handle complex geometry. Its main limitation is that it only provides solutions at specified locations, unlike deterministic methods which provide solutions at all points in the space considered.

Since the Monte Carlo method is a computational process in which random variables are used, we begin by explaining what it is meant by a random variable and reviewing some important statistical concepts.

2.1 Random Variables

In ordinary English usage, a random variable is the outcome of any process that proceeds without any discernible aim or direction.

Mathematically the word random variable means that we do not know the value of a particular quantity in any given case, but we know what values it can assume and we know the probabilities with which it assumes these values. Then a random variable, X, is defined discretely by the table

$$X = (x_1 \;\; x_2 \ldots \;\; x_n ; \;\; p_1 \;\; p_2 \ldots \;\; p_n)$$

where the x_is are possible values of X and the p_is are the corresponding probabilities. Then one writes

$$P(X=x_i)=p_i, \text{ or } p_X(x_i)=p_i.$$

For continuous random variables, a function $p(x)$ in some interval (a,b) is assigned and called the *probability density function* (pdf), or the density distribution, such that

$$P(a < X < b) = \int_a^b p(x')dx' \text{ such that; } p(x) \geq 0$$

The zero-th moment of this function is normalized such that

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

The first moment of the distribution provides the so-called expected value or mathematical expectation

$$E(X) = \int_a^b xp(x)dx$$

The second *central* moment defines the variance of the distribution

$$\sigma^2(X) = \int_a^b [x - E(X)]^2 p(x)dx$$

The cumulative density function, cdf is defined as

$$F(x_O)=P(X\leq x_O)=\int_a^{x_0} p(x)dx$$

Then,

$$P(a' < X < b') = \int_{a'}^{b'} p(x)dx = F(b')-F(a')$$

The cumulative density function represents an area under the pdf extending from a to x_o.

The cdf is particularly useful in Monte Carlo calculations as shown later.

2.2 Abstract Analysis

An event means the occurrence of a specified outcome of an experiment. Let Ω be the event that includes all possible events. Then any event Λ, Ω, that is Λ is contained in Ω.

The probability is a real-valued function of the events of an experiment satisfying

$$P(\varnothing)=0,\ \ P(\Omega)=1;\ \ 0 \le P(\Lambda) \le 1; \text{ for all } \Lambda \subset \Omega$$

$$P\left(\bigcup_{I=1}^{\infty} \Lambda_i\right)= \sum_{I=1}^{\infty} p(\Lambda_i); \text{ if } \Lambda_i \cap \Lambda_j = \varnothing,\ i \neq j$$

where $\cup$ reads *union* and signifies the fact that the event exists simultaneously in the spaces considered, while $\cap$ reads *intersection* and defines events common to the concerned spaces.

The above abstract notion of probability is more general than the frequency notion usually used in statistical analysis. In the frequency probability analysis, in an experiment repeated n times, with an event Λ occurring $n\Lambda$ times, one expects $(n\Lambda)/n$ to cluster about a unique number $P(\Lambda)$.

The abstract notion of probability requires however only that the function P assigns to every event Λ a number with the above probability.

The probability is a real-valued function in certain subsets of Ω, which we call the event of Ω. Certain real-valued functions of the *points* of Ω are called *random variables*.

Let a point of Ω denoted by ω and let ξ be a real-valued function on the points of Ω. Let

$$\Lambda(t)=\{ \omega|\xi(\omega) \leq t\}$$

This defines the set of all points ω such that $\xi(\omega) \leq t$. Then $\Lambda(t)$ is a subset of Ω which depends on the real number t.

If for every t, the set $\Lambda(t)$ is an event, then ξ is called a *random variable*. Then

$$P\{\Lambda(t)\} = P\{\omega|\xi(\omega) \leq t\} = P\{\xi \leq t\};\ \text{defined for every } t$$

The above is the formal definition of random variables.

The real-valued function of a real variable defined by

$$F(t) = P\{\xi \leq t\}$$

is called the distribution function or the cumulative density function.
It has the following characteristics

1. F is continuous on the right at every event t

2. F is a monotone non-decreasing function

3. $F(-\infty)=0$ and $F(\infty)=1$

4. $F(a)-F(b)=P(a < x \leq b)$, for $a<b$

5. If t_O is a point of discontinuity of F with a jump of height p, then $P\{\xi = t_O\}=p$ and there is a non-zero probability that the random variable takes on the value t_O .

6. If the derivative of F with respect to t exists at point t, then

$$lim_{\Delta \to 0} \quad P\{t-\Delta/2 < x \leq t + \Delta/2\} = (dF/dt)\Delta = f(t)$$

If the derivative, $f(t)$ exists, it is called the probability density function, or simply the density function.

If N independent trials of an experiment are performed, the probability space Ω consisting of all N-tuples $(\omega_1,.., \omega_N)$ of points of Ω is

$$P^N(\omega_1,.., \omega_N) = \prod_{i=1}^{N} P(\omega_i)$$

Define N random variables ξ_i on Ω^N by

$$\xi_i(\omega_1,.., \omega_N) = \xi_i(\omega_i);\ 1 \leq i \leq N$$

If ξ is a discrete random function of N variables, such that

$$\xi^{(N)} = \sum_{I=1}^{N} \xi_i$$

is also a random variable on Ω^N and represents the total number of occurrences of the event ω in N repetitions of the experiment.

2.2.1 Tchebycheff Theorem

This theorem states mathematically that for any random variable, ξ, of any distribution function with a mean, or expected value m, and a standard deviation σ

$$P\{|\xi - m| > k\,\sigma\} \leq 1/k^2, \quad k > 0$$

If we use k to define an "error" $\varepsilon = k\,\sigma$, then for the random variable $\xi^{(N)}$

$$P\{|\xi^{(N)} - p| > \varepsilon\} \leq 1/(4N\varepsilon^2)$$

This theorem reiterates the fact represented by equation (2.1) {error $\propto (1/N)^{1/2}$} that to reduce the error by a factor of two, the number of trials of the experiment must be quadrupled.

2.2.2 Central Limit Theorem

This is also called the law of large numbers and states essentially that $\xi^{(N)}$ (will be approximately normally distributed even if ξ is not.

The theorem states formally that if ξ_1, .. ξ_N is a sequence of *independent* and *identically distributed* random variables with a common mean *m* and variance σ^2, then

$$\bar{\xi} = (1/N) \sum_{I=1}^{N} \xi_I$$

is asymptotically normal (m, $\sigma/N^{1/2}$), that is

$$\lim_{N\to\infty} P\{(\bar{\xi} - m)/(\sigma/N^{1/2})\} \le x\} = \{1/(2\pi^{1/2})\} \int_{-\infty}^{x} exp(-t^2/2)\, dt$$

The theorem assumes that both *m* and σ exist, that is they are given by absolutely convergent integrals. Applying the Tchebycheff's theorem, then

$$P\{|\{(\bar{\xi} - m)/(\sigma/N^{1/2})\}| < \varepsilon'\} \text{ leads to } \{1/(2\pi^{1/2})\} \{ \int_{-\infty}^{\varepsilon'} exp(-t^2/2)\, dt - \int_{\varepsilon'}^{\infty} exp(-t^2/2) dt\}$$

In the above equation leads to implies that they are asymptotically equal.

$$P\{|\bar{\xi}-m|<\varepsilon\}=\{2/\pi^{1/2}\}\int_0^{\varepsilon/(\sigma/\sqrt{N})}\exp(-t^2/2)dt$$

The central-limit theorem is the backbone of the Monte Carlo method. The average value $\bar{\xi}$ is used as an *estimate* of the random variable ξ. This value approaches the true expected value, m, as the number of trials approach infinity.

The variability estimated by

$$s^2=(1/N)\sum_{I=1}^{N}\xi_I^2-\{(1/N)\sum_{I=1}^{N}\xi_i\}^2$$

is called the sample variance. It is not directly an estimate of the distribution variance. It can be stated however that

$$s^2=\{N/(N-1)\}E(\sigma^2)$$

The confidence interval in the estimated value of $\bar{\xi}$ can be defined by

$$[\bar{\xi}+\sigma_e\ ,\ \bar{\xi}-\sigma_e\ ],\ \text{where}\ \sigma_e=\sigma/N^{1/2}$$

Since σ is not known, the following estimate is used

$$\sigma_e^2=\sigma^2/N=\{1/(N-1)\}[(1/N)\sum_{I=1}^{N}\xi^2-\{(1/N)\sum_{I=1}^{N}\xi_i\}^2]$$

A useful quantity used in Monte Carlo computations is called the *fraction standard deviation*, fsd, defined as

$$\text{fsd}=\sigma_e/\bar{\xi}$$

An fsd of less than 0.05, or 5%, is usually required in Monte Carlo calculations.

2.3 Construction of Samples

The realization of a random variable, ξ_i from a pdf $f(x)$ is obtained by constructing a sequence of numbers $t_1, \ldots, t_n$, such that

$$P\{a < t \leq b\} = \int_a^b f(x)\, dx$$

and

$$P\{a < t_{i,1}, \ldots, t_{i,n} \leq b\} = P\{a < t_{i,1} \leq b\} \ldots P\{a < t_{i,n} \leq b\} = \left[\int_a^b f(x)dx\right]^n$$

The above equation implies that the random variables $\xi_{i,1}, .., \xi_{i,n}$, are mutually independent, if $t_{i,1}, ., t_{i,n}$, are all different.

The sequence of *random numbers*, $\rho_1, \ldots, \rho_n$, such that $0 \leq \rho_i \leq 1$, represents samples drawn independently from a *uniform* pdf in the interval [0,1]. that is

$$P\{a < \rho_i \leq b\} = \int_a^b 1\ dx = b-a$$

and

$$P\{a < \rho_{i1}, \ldots, \rho_{in} \leq b\} = (b-a)^n, \quad i_1, \ldots, i_n, \text{ are all different}$$

Now by setting a sampled random number ρ equal to the cdf $F(x)$, that is

$$F(x)=P\{\xi \leq x\}=\int_{a}^{x} f(t)dt = \rho$$

one can solve for x, and consequently obtain a value that is sampled from the distribution $f(x)$.

For a discrete distribution, one constructs the cdf

$$F(x)= \sum p_i \quad \text{for } x \leq x_i$$

where p_i is the probability of occurrence of x_i. The sampling of an x_j is then achieved as follows

$$\sum_{I=1}^{j-1} p_i < \rho \leq \sum_{I=1}^{j} p_i$$

where ρ is a random number uniformly distributed in the interval (0,1).

2.4 Random Number Generation

Digital random number generators are nowadays a standard feature in almost all computer systems. The generated numbers are called *pseudo random numbers* as they are not purely random. They must satisfy however two important criteria:

Equi-distribution: each number has the same probability of occurrence as any other number in the set.

Independence: the occurrence of any given number should not depend on the previous occurrence or any subsequent occurrence of any other number.

The modulus method is perhaps the most widely used method. Given any constant *a* the random numbers are generated as follows

$$\rho_i = a\ \rho_{i-1} \ (\text{mod } M)$$

where $M=2^k$, and k is the number of bits per word in the computer being used. Modulus is a number or quantity that produces the same remainder when divided into each of two quantities. $A = B$ (mod M) reads A is congruent to B module M and means A is the remainder of B/M[1].

[1] For example: 15 (mod 13)=2

2.5 Monte Carlo Simulation of Particle Transport

2.5.1 Essential Requirements

Source

The position, geometry, directional distribution and energy distribution of the source must be specified. In transient analysis, the change of the source with time must be also known. Fission sources and collision sources are determined by the cross section of the material and need not be specified as input parameters. The fission distribution with energy, $\chi(E)$, must however be specified, in order to determine the energy of the emerging neutron. A source particle is usually assigned a statistical weight, W, the significance of which is examined later.

Geometry

The Monte Carlo method can handle complex geometries. The geometry must however be specified in such a way that enables tracking of the particle throughout the system and relating the position of the particle within the system to the material, or more specifically material cross section.

The geometry can be specified via analytical geometry procedures, which define the surfaces of different geometrical objects. Alternatively, the geometry may be specified via a set of elementary bodies, combined together using logical operators to form a zone of a particular material. This is called the combinatorial geometry method and is utilized in the MORSE code.

Material Cross Sections

The cross sections for the different materials encountered must be supplied as a function of energy. The Legendre expansion coefficients for each material are also needed, if an anisotropic scattering is considered.

The cross sections are processed prior to the simulation to provide the probability table, which determines the distance to be travelled by the particle until the next collision, the outcome of the collision, and the outgoing energy and angle of a scattering event; in addition to the number of neutrons per fission for fissile materials.

Scoring

The scoring process is determined by a variety of estimators which evaluate the fluence, or fluence-like quantities, at a point or a region. Statistical estimates, including the average and the variance of the average are estimated at the end of the random walk process.

The *surface crossing estimator* evaluates the flux crossing a surface, by accumulating the weight of particles crossing the surfaces divided by the absolute value of the cosine of the angle between the normal to the surface and direction of the incident particle. Provisions are made to avoid small angles cosines.

The *track length estimator* evaluates the fluence by summing the track length of particles crossing a given zone, divided by the volume of the zone.
This is usually suitable for evaluating the fluence in void or air regions, and regions containing a low density material.

The *collision density estimator* adds up the weight of particles colliding within a zone, divided by the total cross section of the material and the volume of the zone. The estimator provides adequate estimates for the fluence in regions of high density materials, where a large number of collisions are anticipated.

In all the above estimators, the particle must visit the region, or surface of interest. In situations where the probability of the particles reaching the region of interest is low, indirect estimates, called statistical estimation are used. These estimators evaluate the probability of the next collision being at the detector site. This called the *next event estimator* and is particularly useful for point detectors, where there is only one possible position for the "next collision". Note that the particle being tracked does not alter its original position, only the probability of the next collision being at the detector site is evaluated and stored.

2.5.2 **Example**

In order to illustrate the above points, let us consider the relatively simple problem of evaluating the fluence through a shielding slab, with a neutron source on one side and a detector on the other side.

Source Parameters

If we assume a point, mono-energetic and isotropic source, then only the direction of the incident particle need to be sampled, in a steady state problem. Since we are not interested in neutrons directed away from the target, the angular probability becomes

$$p(\Omega)d\Omega = d\Omega/2\pi = (\sin\theta d\theta\, d\varphi)/2\pi = (d\cos\theta d\varphi)(2\pi)$$

where θ is the polar angle and ϕ is the azimuthal angle.

Equating the cumulative probability for cos θ to a random number ρ_1 sampled from a uniform distribution in the interval (0,1), then

$$\rho_1 = \int_{-1}^{\cos\theta} (1/2)\, d\cos\theta' = (\cos\theta + 1)/2$$

The inversion of the above leads to an equation for selecting θ

$$\theta = \cos^{-1}(2\rho_1 - 1)$$

Similarly, the angle φ is sampled from the relationship $\varphi = \pi\rho_2$ where ρ_2 is another random number.

Distance of Travel

Next, one needs to determine the distance the neutron will travel until it collides. The probability of a neutron experiencing its first interaction between the distances x and $x+dx$ is equal to $\Sigma_t \exp(-\Sigma x)$, where Σ_t is the total cross section of the material encountered.

As shown earlier, equation (2.44) $\{x=-(1/\Sigma) \ln \rho\}$ provides the method for sampling the distance, x, the neutron will travel until it collides.

3.2.3 Type of Interaction

The type of interaction which takes place at the position defined by the distance x is determined by the so-called *activation cross sections*, which are usually $\Sigma_{scatter}$, $\nu \Sigma_{fission}$, and $\Sigma_{capture}$. In certain circumstances, the cross section of some particular reactions, such as the (n,p) reaction, may be specified. The activation table of probabilities is converted into a cumulative probability table, enabling the selection of the proper interaction.

If the interaction is determined to be an absorption process, the random walk of the particle may be terminated. This is called *analog Monte Carlo* and is not often used as it may result in early termination of the random walk. Alternatively, a *non-analog* process is used in which the particle weight is reduced by the non-absorption probability, $(\Sigma_{total}-\Sigma_{capture})/ \Sigma_{total}$, and a particle scattering or fission is sampled.

This process allows the particle to fully complete its path within the system, until it escapes the system or is terminated by a weight cut-off, or an energy cut-off, or some other pre-specified process.

3.2.4 Energy of Outgoing Particle

In a non-fissile material, the only interaction possible in a non-analog Monte Carlo is particle scattering. One needs then to determine the energy and angle of the particle emerging from the collision.

Let us assume an elastic isotropic neutron scattering process.
Then, the energy of the outgoing particle can lie any where from the energy of the incident particle E_i to the minimum possible energy a E, where $\alpha=[(A-1)/(A+1)]^2$, with A being the mass number of the element considered. The probability of the particle reaching an energy E is given by

$$p(E)dE=dE/[E_i(1-\alpha)]$$

Equating the cumulative probability to some random number ρ_3, one obtains

$$E=\rho_3 E_i(1-\alpha) + \alpha E_i$$

Problem. Devise a method for determining the outgoing energy for isotropic scattering in a chemical compound such as water.

The outgoing energy is sampled from the above equation. Since isotropic scattering is assumed, the outgoing direction can be sampled using a procedure similar to that used for the source, except that the whole 4π of the azimuthal angle must be considered.

Once the direction and energy of the scattered particle are determined, the distance of flight until the next collision is evaluated, and so on. Note, however, in this one-dimensional problem it is sufficient to determine the x position of the collision site, as $x=x_i+d\cos\theta \sin\varphi$, where x_i is the initial position of the particle and d is the distance the particle travels between collisions.

Scoring

A simple scoring process is to employ the boundary crossing estimator at the boundary far away from the source. For a deep penetration problem, e.g. thick shield, the probability of the particle crossing the shielding slab is very low. Some so-called biasing or *importance sampling* techniques can be employed. These techniques involve *splitting* which increases the number of particles travelling towards the location of interest (forward in the problem considered), and R*ussian roulette* which kills most of the particles travelling in the "wrong" direction. The *exponential transformation* technique may also be employed. In this technique, the total cross section is artificially decreased, to enable the particle path length between collisions to stretch, and consequently be able to cross the slab. In all these biasing techniques, the particle weight is adjusted such that the resulting estimates are *unbiased*.

2.6 Work Problems

1. Prove that the $\sigma^2(X)=E[\{X-E(X)\}^2]$

2. Show that the following procedure represents sampling from the distribution $\Sigma \exp(-\Sigma x)$, where Σ is a constant

$$x=-(1/\Sigma) \ln \rho$$