Monte Carlo Methods

Peter Shirley

November 2, 1990

1 Introduction

In this appendix the basic Monte Carlo solution methods for definite integrals and sums are outlined. These techniques are then straightforwardly extended to certain integral and linear equations. All of the material of this appendix is also covered in several of the classic Monte Carlo texts. This appendix differs by being geared toward classes of problems that crop up in Computer Graphics. Readers interested in a broader treatment of Monte Carlo techniques should consult one of the classic Monte Carlo texts[2, 3, 1, 4].

2 Background

Before getting to the specifics of Monte Carlo techniques, we need several definitions, the most important of which are: *random variable*, *expected value*, and *variance*.

Loosely speaking, a *random variable* x is a scalar or vector quantity that 'randomly' takes on some value, and the behavior of x is entirely described by the distribution of values it takes. This distribution of values can be quantitatively described by the *probability density function*, f, associated with x (the relationship is denoted $x \sim f$). If x ranges over some region Ω, then the probability that x will take on a value in some subregion $\Omega_i \subset \Omega$ is given by the integral:

$$P(x \in \Omega_i) = \int_{x' \in \Omega_i} f(x') d\mu(x') \quad (f : \Omega \to \Re^1) \tag{1}$$

Here $P(event)$ is the probability that *event* is true, so the integral is the probability that x takes on a value in the region Ω_i. The measure μ is the measure on our probability space. In graphics Ω is typically an area ($d\mu = dA = dxdy$), or a set of directions (points on a unit sphere: $d\mu = d\omega = \sin\theta d\theta d\phi$). The density f has two characteristics:

$$f(x) \geq 0 \quad \text{(Probability is nonnegative)} \tag{2}$$

$$\int_{x' \in \Omega} f(x') d\mu(x') = 1 \quad (x \text{ has a value in } \Omega) \tag{3}$$

As an example, the canonical random variable ξ takes on values between zero and one with uniform probability. This implies that:

$$f(\xi) = \begin{cases} 1 & \text{if } 0 \leq \xi \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The probability that ξ takes on a value in a certain region is:

$$P(a < \xi < b) = \int_a^b 1 d\xi' = b - a$$

The average value a random variable will take on is called its *expected value*, $E(x)$:

$$E(x) = \int_{x' \in \Omega} x' f(x') d\mu(x')$$

The expected value has a surprising and useful property: *the expected value of the sum of two random variables is the sum of the expected values of those variables* ($E(x + y) = E(x) + E(y)$). This property holds whether or not the variables are independent! Since the sum of two random variables is itself a random variable, this principle generalizes. Since a function of x is itself a random variable, we can write down the expected value of a function $g(x)$:

$$E(g(x)) = \int_{x' \in \Omega} g(x') f(x') d\mu(x')$$

The *variance*, $var(x)$, of a random variable is the expected value of the square of the difference between x and $E(x)$:

$$var(x) = E([x - E(x)]^2) = E(x^2) - [E(x)]^2$$

The variance of a sum of random variables is the sum of the variances *if the variables are independent.* The square root of the variance is called the *standard deviation*, which gives some indication of absolute deviation from the expected value.

Many problems involve sums of independent random variables x_i, where the variables share a common density f. Such variables are said to be *independent identically distributed* random variables. When the sum is divided by the number of variables, we get an estimate of $E(x)$:

$$E(x) \approx \frac{1}{N} \sum_{i=1}^{N} x_i$$

This idea can be generalized to the *Law of Large Numbers*:

$$P\left[E(x) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i\right] = 1$$

This idea leads naturally to the idea of Monte Carlo estimation of integrals.

3 Monte Carlo Integration

From the last section we saw that for a function g and a random variable $x \sim f$, we can approximate the expected value of $g(x)$ by a sum:

$$E(g(x)) = \int_{x'\in\Omega} g(x')f(x')d\mu(x') \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i) \tag{4}$$

Because the expected value can be expressed as an integral, the integral is also approximated by the sum. The form of Equation 4 is a bit awkward; we would usually like to approximate an integral of a single function h rather than a product gf. We can get around this by substituting $h = gf$ as the integrand:

$$\int_{x'\in\Omega} h(x')d\mu(x') \approx \frac{1}{N}\sum_{i=1}^{N} \frac{h(x_i)}{f(x_i)} \tag{5}$$

For this formula to be valid, f must be positive where h is nonzero.

Variance can be used to measure the reliability of the estimate. Both estimates are *unbiased*, which means that the expected values are what we would expect. The simple term $h(x_i)/f(x_i)$ is called the *primary estimator*, and the average of many primary estimators is a *secondary estimator*. The secondary estimator is preferred because its variance is lower. The variance of the estimate is:

$$var\left(\frac{1}{N}\sum_{i=1}^{N} \frac{h(x_i)}{f(x_i)}\right) = \frac{var(\frac{h(x)}{f(x)})}{N} \tag{6}$$

So to get a good estimate, we want as many samples as possible, and we want the density h/f to have a low variance (similar shape). Choosing f intelligently is called importance sampling, because if f is large where h is large, there will be more samples in important regions. Equation 4 also shows the fundamental problem with Monte Carlo integration: *diminishing return*. Because the variance of the estimate is proportional to $1/N$, the standard deviation is proportional to $1/\sqrt{N}$. Since the error in the estimate behaves similarly to the standard deviation, we will need to quadruple N to halve the error.

Another way to reduce variance is to partition Ω, the domain of the integral, into several smaller domains Ω_i, and evaluate the integral as a sum of integrals

method	*sampling function*	*variance*	*samples needed for standard error of 0.008*
importance	$(6-x)/(16)$	$56.8N^{-1}$	887,500
importance	$1/4$	$21.3N^{-1}$	332,812
importance	$(x+2)/16$	$6.3N^{-1}$	98,437
importance	$x/8$	0	1
stratified	uniform	$21.3N^{-3}$	70

Table 1: Variance for Monte Carlo Estimate of $\int_0^4 x\,dx$

over the Ω_i. This is called stratified sampling. Normally only one sample is taken in each Ω_i (with density f_i), and in this case the variance of the estimate is:

$$var\left(\sum_{i=1}^{N}\frac{h(x_i)}{f_i(x_i)}\right)=\sum_{i=1}^{N}var\left(\frac{h(x_i)}{f_i(x_i)}\right) \tag{7}$$

As an example of the Monte Carlo solution of an integral I set $h(x)$ to be x over the interval (0, 4):

$$I=\int_0^4 x\,dx=8 \tag{8}$$

The great impact of the shape of the function f on the variance of the N sample estimate is shown in Table 1. Note that the variance is lessened when the shape of f is similar to the shape of h. The variance drops to zero if $p=Ch$, but h is not usually known or we would not have to resort to Monte Carlo. One important principle illustrated in Table 1 is that stratified sampling is often *far* superior to importance sampling. Although the variance for this stratification on I is inversely proportional to the cube of the number of samples, there is no general result for the behavior of variance under stratification. There are some functions where stratification does no good. An example is a white noise function, where the variance is constant for all regions. A poorly chosen stratification can even increase the variance for some functions.

4 Monte Carlo Solution to an Integral Equation

The 'rendering equation' is a *Fredholm Equation of the Second Kind.* Such equations have the form:

$$a(x)=b(x)+\int_{x'\in\Omega}k(x,x')a(x')d\mu(x')$$

Where b and k are known. To apply the equipment of the last section we can repeatedly substitute a into the integral:

$$\begin{aligned} a(x) &= b(x) + \\ &\int_{x' \in \Omega} k(x,x')b(x')d\mu(x') + \\ &\int_{(x',x'') \in \Omega^2} k(x,x')k(x',x'')b(x'')d\mu(x')d\mu(x'') + \\ &\int_{(x',x'',x''') \in \Omega^3} k(x,x')k(x',x'')k(x'',x''')b(x''')d\mu(x')d\mu(x'')d\mu(x''') + \end{aligned} \quad (9)$$

A primary estimator for the first integral in the series is:

$$\frac{k(x,x')b(x')}{f_1(x')} \quad : x' \sim f_1 \quad (10)$$

A primary estimator for the second integral is:

$$\frac{k(x,x')k(x',x'')b(x'')}{f_2(x',x'')} \quad : (x',x'') \sim f_2 \quad (11)$$

And the third integral:

$$\frac{k(x,x')k(x',x'')k(x'',x''')b(x''')}{f_3(x',x'',x''')} \quad : (x',x'',x''') \sim f_3 \quad (12)$$

We could simple estimate each integral separately and add these estimates to form a estimate for the truncated series. What is usually done, however, is to reuse the share sample points between integrals. To do this, we choose a chain of samples $(x^1, x^2, x^3, \cdots, x^n)$. The estimator for the first n integrals in Equation 9 is:

$$\begin{aligned} a(x) &= b(x) + \\ &\frac{k(x,x^1)b(x^1)}{f_1(x^1)} + \\ &\frac{k(x,x^1)k(x^1x^2)b(x^2)}{f_2(x^1,x^2)} + \\ &\frac{k(x,x^1)k(x^1,x^2)k(x^2,x^3)b(x^3)}{f_3(x^1,x^2,x^3)} + \\ &\vdots \\ &\frac{k(x,x^1)\cdots k(x^{n-1},x^n)b(x^n)}{f_n(x^1,\cdots,x^n)} \end{aligned} \quad (13)$$

We should probably have some misgivings about reusing the sample points for each integral, but this will not bias our sample because, as stated earlier, the expected value of a sum is the sum of expected values, even if the variable being summed are *not* independent.

5 Monte Carlo Estimates of Sums and Linear Equations

The techniques used to estimate integrals can also be used to estimate sums. One way to see this is to consider a sum to be a special case of an integral (and discrete probability a special case of continuous probability). A primary estimator for a sum is:

$$\sum_{i=1}^{N} h_i \approx \frac{h_i}{f(i)}$$

Here $f(i)$ is the probability of choosing h_i. As with integrals, a lower variance secondary estimator can be developed by averaging multiple instances of the primary estimator.

Just as Monte Carlo integration extends to integral equations, Monte Carlo summation extends to linear equations. Consider the linear system:

$$x = b + Ax$$

where x is an unknown column vector of length N, b is a known column vector of length N, and A is a known $N \times N$ matrix. As with integrals, we first expand the equation into a series:

$$x = b + Ab + A^2b + A^3b + \cdots$$

Any particular element of x is given by:

$$\begin{aligned} x_i \quad = \quad & b_i + \\ & \sum_{j=1}^{N} A_{ij} b_j + \\ & \sum_{j=1,k=1}^{N} A_{ij} A_{jk} b_k + \\ & \sum_{j=1,k=1,l=1}^{N} A_{ij} A_{jk} A_{kl} b_l + \cdots \end{aligned} \tag{14}$$

As with the integral equations, we can generate a series $(j, k, l, m, \cdots)$, and generate a primary estimator for x_i:

$$\begin{aligned} x_i \quad \approx \quad & b_i + \\ & \frac{A_{ij} b_j}{f_1(j)} + \\ & \frac{A_{ij} A_{jk} b_k}{f_2(j,k)} + \\ & \frac{A_{ij} A_{jk} A_{kl} b_l}{f_3(j,k,l)} + \cdots \end{aligned} \tag{15}$$

6 Generating Random Numbers With Non-Uniform Densities

For Monte Carlo methods we must often generate random points according to some probability density function, or random rays according to a directional probability density. In this section a method for one and two dimensional random variables is described. The discussion closely follows that of Shreider[3].

If the density is a one dimensional $f(x)$ defined over the interval $x \in [a, b]$, then we can generate random numbers α_i that have density f from a set of uniform random numbers ξ_i, where $\xi_i \in [0, 1]$. To do this we need the probability distribution function $F(x)$:

$$F(x) = \int_a^x f(x')d\mu(x') \tag{16}$$

To get α_i we simply transform ξ_i:

$$\alpha_i = F^{-1}(\xi_i) \tag{17}$$

where F^{-1} is the inverse of F. If F is not analytically invertable then numerical methods will suffice because an inverse exists for all valid probability distribution functions.

If we have a two dimensional density (x, y) defined on $[a, b : c, d]$ then we need the two dimensional distribution function:

$$F(x, y) = \int_c^y \int_a^x f(x', y')d\mu(x', y') \tag{18}$$

We first choose an x_i using the marginal distribution $F(x, d)$, and then choose y_i according to $F(x_i, y)/F(x_i, d)$. If $f(x, y)$ is separable (expressable as $g(x)h(y)$), then the one dimensional techniques can be used on each dimension.

To choose reflected ray directions for zonal calculations or distributed ray tracing, we can think of the problem as choosing points on the unit sphere or hemisphere (since each ray direction ψ can be expressed as a point on the sphere). For example, suppose that we want to choose rays according to the density:

$$p(\theta, \phi) = \frac{n+1}{2\pi} \cos^n \theta \tag{19}$$

Where n is a Phong-like exponent, θ is the angle from the surface normal and $\theta \in [0, \pi/2]$ (is on the upper hemisphere) and ϕ is the azimuthal angle ($\phi \in [0, 2\pi]$). The distribution function is:

$$P(\theta, \phi) = \int_0^\phi \int_0^\theta p(\theta', \phi') \sin\theta' d\theta' d\phi' \tag{20}$$

The $\cos\theta'$ term arises because on the sphere $d\omega = \cos\theta d\theta d\phi$. When the marginal densities are found, p (as expected) is separable and we find that a (r_1, r_2) pair of uniform random numbers can be transformed to a direction by:

$$(\theta, \phi) = (\arccos((1 - r_1)^{\frac{1}{n+1}}), 2\pi r_2) \tag{21}$$

One nice thing about this method is that a set of jittered points on the unit square can be easily transformed to a set of jittered points on the hemisphere with a distribution of Equation 19. If n is set to 1 then we have a diffuse distribution needed for a Monte Carlo zonal method.

Other example results are: to choose points uniformly from a disk of radius R, apply the transformation $\theta = 2\pi r_1$, $r = R\sqrt{r_2}$. To choose random points on a triangle defined by vertices p_0, p_1, and p_2, apply the transformation: $a = 1 - \sqrt{1 - r_1}$, $b = (1 - a)r_2$, and the random point p will is: $p = p_0 + a(p_1 - p_0) + b(p_2 - p_0)$.

References

[1] John H. Halton. A retrospective and prospective of the monte carlo method. *SIAM Review*, 12(1):1–63, January 1970.

[2] J. M. Hammerley and D. C. Handscomb. *Monte Carlo Methods.* Wiley, New York, N.Y., 1964.

[3] Y. A. Screider. *The Monte Carlo Method.* Pergamon Press, New York, N.Y., 1966.

[4] Sidney J. Yakowitz. *Computational Probability and Simulation.* Addison-Wesley, New York, N.Y., 1977.