A

Monte Carlo Integration

THE techniques developed in this dissertation are all Monte Carlo methods. Monte Carlo methods are numerical techniques which rely on random sampling to *approximate* their results. Monte Carlo integration applies this process to the numerical estimation of integrals. In this appendix we review the fundamental concepts of Monte Carlo integration upon which our methods are based. From this discussion we will see why Monte Carlo methods are a particularly attractive choice for the multidimensional integration problems common in computer graphics. Good references for Monte Carlo integration in the context of computer graphics include Pharr and Humphreys [2004], Dutré et al. [2006], and Veach [1997].

The term "Monte Carlo methods" originated at the Los Alamos National Laboratory in the late 1940s during the development of the atomic bomb [Metropolis and Ulam, 1949]. Not surprisingly, the development of these methods also corresponds with the invention of the first electronic computers, which greatly accelerated the computation of repetitive numerical tasks. Metropolis [1987] provides a detailed account of the origins of the Monte Carlo method.

Las Vegas algorithms are another class of method which rely on randomization to compute their results. However, in contrast to Las Vegas algorithms, which always produce the exact or correct solution, the accuracy of Monte Carlo methods can only be analyzed from a statistical viewpoint. Because of this, we first review some basic principles from probability theory before formally describing Monte Carlo integration.

A.1 Probability Background

In order to define Monte Carlo integration, we start by reviewing some basic ideas from probability.

A.1.1 Random Variables

A *random variable* X is a function that maps outcomes of a random process to numbers. A random variable can be either discrete (e.g., the roll of a six-sided die where a fixed set of outcomes is possible, $X = \{1,2,3,4,5,6\}$), or continuous (e.g., a person's height, which can take on real values $\mathbb{R}$). In computer graphics we more commonly deal with continuous random variables which take on values over ranges of continuous domains (e.g., the real numbers $\mathbb{R}$ or the sphere of directions Ω).

A.1.2 Cumulative Distributions and Density Functions

The *cumulative distribution function*, or CDF, of a random variable X is the probability that a value chosen from the variable's distribution is less than or equal to some thresold x:

$$cdf(x) = Pr\{X \leq x\}. \tag{A.1}$$

The corresponding *probability density function*, or PDF, is the derivative of the CDF:

$$pdf(x) = \frac{d}{dx} cdf(x). \tag{A.2}$$

CDFs are always monotonically increasing, which means that the PDF is always non-negative. An important relationship arises from the above two equations, which allows us to compute the probability that a random variable lies within an interval:

$$Pr\{a \leq X \leq b\} = \int_a^b pdf(x)\, dx. \tag{A.3}$$

From this expression it is clear that the PDF must always integrate to one over the full extent of its domain.

A.1.3 Expected Values and Variance

The *expected value* or *expectation* of a random variable $Y = f(X)$ over a domain $\mu(x)$ is defined as

$$E[Y] = \int_{\mu(x)} f(x)\, pdf(x)\, d\mu(x), \tag{A.4}$$

while its *variance* is

$$\sigma^2[Y] = E\left[(Y - E[Y])^2\right], \tag{A.5}$$

where σ, the *standard deviation*, is the square root of the variance. From these definitions it is easy to show that for any constant a,

$$E[a\,Y] = a\,E[Y], \tag{A.6}$$

$$\sigma^2[a\,Y] = a^2\,\sigma^2[Y]. \tag{A.7}$$

Furthermore, the expected value of a sum of random variables Y_i is the sum of their expected values:

$$E\left[\sum_i Y_i\right] = \sum_i E[Y_i]. \tag{A.8}$$

From these properties it is possible to derive a simpler expression for the variance:

$$\sigma^2[Y] = E\left[Y^2\right] - E[Y]^2. \tag{A.9}$$

Additionally, if the random variables are *uncorrelated*, a summation property also holds for the variance[1]:

$$\sigma^2\left[\sum_i Y_i\right] = \sum_i \sigma^2[Y_i]. \tag{A.10}$$

A.2 The Monte Carlo Estimator

The Basic Estimator. Monte Carlo integration uses random sampling of a function to numerically compute an estimate of its integral. Suppose that we want to integrate the one-dimensional function $f(x)$ from a to b:

$$F = \int_a^b f(x)\,dx. \tag{A.11}$$

We can approximate this integral by averaging samples of the function f at uniform random points within the interval. Given a set of N uniform random variables $X_i \in [a,b)$ with a corresponding PDF of $1/(b-a)$, the Monte Carlo estimator for computing F is

$$\left\langle F^N \right\rangle = (b-a)\frac{1}{N-1}\sum_{i=0}^{N} f(X_i), \tag{A.12}$$

The random variable $X_i \in [a,b)$ can be constructed by warping a canonical random number uniformly distributed between zero and one, $\xi_i \in [0,1)$:

$$X_i = a + \xi_i(b-a). \tag{A.13}$$

Using this construction, we can expand the estimator to:

$$\left\langle F^N \right\rangle = (b-a)\frac{1}{N}\sum_{i=0}^{N-1} f\left(a+\xi_i(b-a)\right). \tag{A.14}$$

[1] This property is often made with the stronger condition that the variables are *independent*; however, it suffices for them to be uncorrelated.

Since $\langle F^N \rangle$ is a function of X_i, it is itself a random variable. We use this notation to clarify that $\langle F^N \rangle$ is an approximation of F using N samples.

Intuitively, the Monte Carlo estimator in Equation A.12 computes the mean value of the function $f(x)$ over the interval a to b, and then multiplies this mean by the length of the interval $(b-a)$. By moving $(b-a)$ into the summation, the estimator can be thought of as choosing a height at a random evaluation of the function and averaging a set of rectangular areas computed by multiplying this height by the interval length $(b-a)$. These two interpretations are illustrated in Figure A.1.

A.2.1 Expected Value and Convergence

It is easy to show that the expected value of $\langle F^N \rangle$ is in fact F:

$$\begin{aligned}
E\left[\langle F^N \rangle\right] &= E\left[(b-a)\frac{1}{N}\sum_{i=0}^{N-1} f(X_i)\right], \\
&= (b-a)\frac{1}{N}\sum_{i=0}^{N-1} E\left[f(X_i)\right], && \text{from Equations A.8 and A.6} \\
&= (b-a)\frac{1}{N}\sum_{i=0}^{N-1} \int_a^b f(x)\,pdf(x)\,dx, && \text{from Equation A.4} \\
&= \frac{1}{N}\sum_{i=0}^{N-1} \int_a^b f(x)\,dx, && \text{since } pdf(x) = 1/(b-a) \\
&= \int_a^b f(x)\,dx, \\
&= F. && \text{(A.15)}
\end{aligned}$$

Furthermore, as we increase the number of samples N, the estimator $\langle F^N \rangle$ becomes a closer and closer approximation of F. Due to the *Strong Law of Large Numbers*, in the limit we can guarantee that we have the exact solution:

$$Pr\left\{\lim_{N\to\infty} \langle F^N \rangle = F\right\} = 1. \tag{A.16}$$

In practice we are interested in knowing just how quickly this estimate converges to a

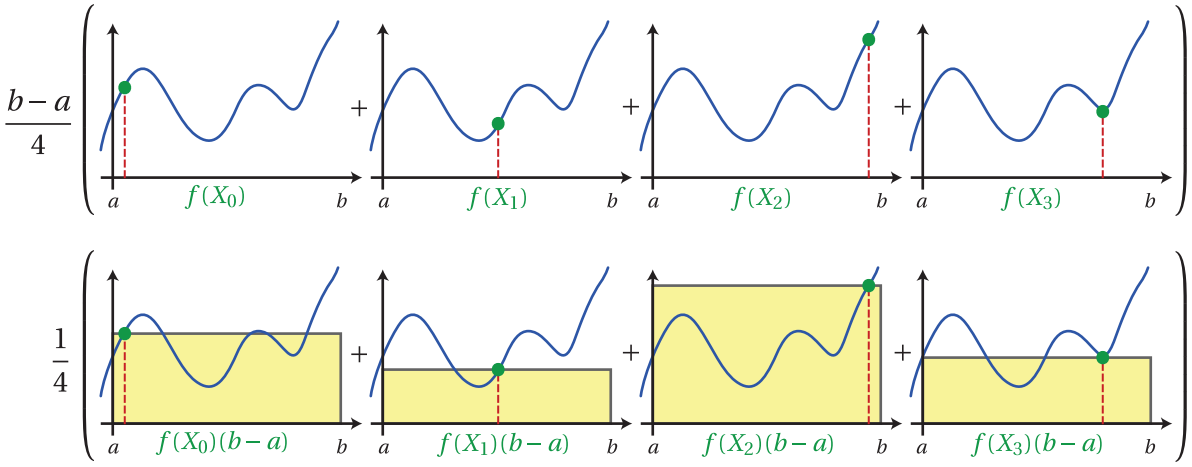

Figure A.1: An illustration of the two interpretations of the basic Monte Carlo estimator in Equation A.12 using four samples: computing the mean value, or height, of the function and multiplying by the interval length (top), or computing the average of several rectangular areas (bottom).

sufficiently accurate solution. This can be analyzed by determining the *convergence rate* of the estimators variance. In the next section we will show that the standard deviation is proportional to:

$$\sigma\left[\langle F^N \rangle\right] \propto \frac{1}{\sqrt{N}}. \tag{A.17}$$

Unfortunately, this means that we must quadruple the number of samples in order to reduce the error by half!

Standard integration techniques exist which converge much faster in one dimension; however, these techniques suffer from the *curse of dimensionality*, where the convergence rate becomes *exponentially* worse with increased dimensions. The basic Monte Carlo estimator above can easily be extended to multiple dimensions, and, in contrast to deterministic quadrature techniques, the convergence rate for Monte Carlo is independent of the number of dimensions in the integral. This makes Monte Carlo integration the only practical technique for many high dimensional integration problems, such as those encountered when computing global illumination.

A.2.2 Multidimensional Integration

Monte Carlo integration can be generalized to use random variables drawn from arbitrary PDFs and to compute multidimensional integrals, such as

$$F = \int_{\mu(x)} f(x)\, d\mu(x), \tag{A.18}$$

with the following modification to Equation A.12:

$$\left\langle F^N \right\rangle = \frac{1}{N} \sum_{i=0}^{N-1} \frac{f(X_i)}{pdf(X_i)}. \tag{A.19}$$

It is similarly easy to show that this generalized estimator also has the correct expected value:

$$\begin{aligned}
E\left[\left\langle F^N \right\rangle\right] &= E\left[\frac{1}{N} \sum_{i=0}^{N-1} \frac{f(X_i)}{pdf(X_i)}\right], \\
&= \frac{1}{N} \sum_{i=0}^{N-1} E\left[\frac{f(X_i)}{pdf(X_i)}\right], \\
&= \frac{1}{N} \sum_{i=0}^{N-1} \int_\Omega \frac{f(x)}{pdf(x)} pdf(x)\, dx, \\
&= \frac{1}{N} \sum_{i=0}^{N-1} \int_\Omega f(x)\, dx, \\
&= \int_\Omega f(x)\, dx, \\
&= F.
\end{aligned} \tag{A.20}$$

In addition to the convergence rate, a secondary benefit of Monte Carlo integration over traditional numerical integration techniques is the ease of extending it to multiple dimensions. Deterministic quadrature techniques require using N^d samples for a d-dimensional integral. In contrast, Monte Carlo techniques provide the freedom of choosing any arbitrary number of samples.

As mentioned previously, the Monte Carlo estimator has a constant $O(\sqrt{N})$ convergence rate in any dimension. However, in many situations it is possible to do much better. The efficiency of Monte Carlo integration can be significantly improved using a variety of techniques, which we

discuss in the next section.

A.3 Variance Reduction

Sources of Variance. In order to improve the quality of Monte Carlo integration we need to reduce variance. Since the samples in Monte Carlo integration are independent, using Equation A.10, the variance of $\langle F^N \rangle$ can be simplified to:

$$\begin{aligned}
\sigma^2 \left[\langle F^N \rangle \right] &= \sigma^2 \left[\frac{1}{N} \sum_{i=0}^{N-1} \frac{f(X_i)}{pdf(X_i)} \right] \\
&= \frac{1}{N^2} \sum_{i=0}^{N-1} \sigma^2 \left[\frac{f(X_i)}{pdf(X_i)} \right] \\
&= \frac{1}{N^2} \sum_{i=0}^{N-1} \sigma^2 \left[Y_i \right] \\
&= \frac{1}{N} \sigma^2 \left[Y \right], \qquad \text{(A.21)}
\end{aligned}$$

and hence,

$$\sigma \left[\langle F^N \rangle \right] = \frac{1}{\sqrt{N}} \sigma \left[Y \right], \qquad \text{(A.22)}$$

where $Y_i = f(X_i)/pdf(X_i)$ and Y represents the evaluation of any specific Y_i, e.g., $Y = Y_1$.

This derivation proves our earlier statement that the standard deviation converges with $O(\sqrt{N})$. Moreover, this expression shows that by reducing the variance of each Y_i we can reduce the overall variance of $\langle F^N \rangle$.

Variance-reduction techniques try to make each Y_i as constant as possible in order to reduce the overall error of the estimator. A significant amount of research has been put into this area, leading to a number of complementary techniques. Most of these methods rely on exploiting some previous knowledge of the function being integrated.

A.3.1 Importance Sampling

Importance sampling reduces variance by observing that we have the freedom to choose the PDF used during integration. By choosing samples from a distribution $pdf(x)$, which has a *similar shape* as the function $f(x)$ being integrated, variance is reduced. Intuitively, importance sampling attempts to place more samples where the contribution of the integrand is high, or "important." If we can properly guess the important regions during integration, the variance of the standard Monte Carlo estimator can be significantly reduced.

The Perfect Estimator

To demonstrate the effect of importance sampling, consider a PDF which is exactly proportional to the function being integrated, $pdf(x) = c\,f(x)$ for some normalization constant c. Since c is a constant, if we apply this PDF to the Monte Carlo estimator in Equation A.19, each sample X_i would have the same value,

$$Y_i = \frac{f(X_i)}{pdf(X_i)} = \frac{\cancel{f(X_i)}}{c\cancel{f(X_i)}} = \frac{1}{c}. \tag{A.23}$$

By using this PDF we have succeeded at reducing the variance of each Y_i. In fact, since each Y_i returns the same value, the overall variance is zero!

Since the PDF must integrate to one, it is easy to derive the value of c:

$$c = \frac{1}{\int f(x)\,dx}. \tag{A.24}$$

This unfortunately shows us that determining the normalization constant c in the PDF involves solving the integral we are interested in estimating in the first place. This best case is therefore not a realistic situation. In practice, all validly constructed PDFs will still produce a convergence rate of $\sigma \propto \sqrt{N}$; however, by choosing a PDF that is similar to $f(x)$, we can make the variance arbitrarily low.

The only restriction on the PDF, in order to maintain the correct expected value from Equation A.20, is that it must be non-zero everywhere where the function $f(x)$ is non-zero.

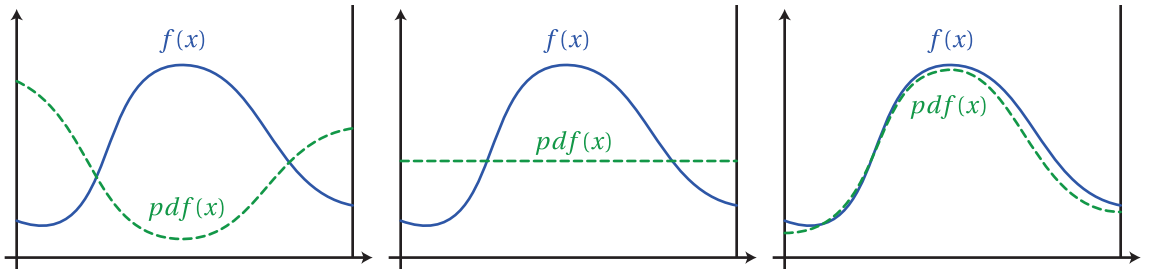

Figure A.2: Comparison of three probability density functions. The PDF on the right provides variance reduction over the uniform PDF in the center. However, using the PDF on the left would significantly increase variance over simple uniform sampling.

Importance sampling should not be used carelessly however. Though we can decrease the variance by using a "good" importance function, we can also make the variance arbitrarily large by choosing an importance function that increases the variance of Y. Figure A.2 shows a uniform PDF, as well as examples of PDFs which reduce and increase the overall variance.

Importance Sampling Complex Functions

Most of the time, the integrand $f(x)$ is very complicated, and we cannot guess its full behavior ahead of time. However, we may know something about its general structure. For instance, the integrand function $f(x)$ may in fact be the combination of more than one function, e.g., $f(x) = g(x)\,h(x)$. In these situations, it may not be possible to create a PDF exactly proportional to $f(x)$, but, if we know one of the functions in advance, we may be able to construct a PDF proportional to a portion of $f(x)$, e.g., $pdf_g(x) \propto g(x)$. In this situation, the Monte Carlo estimator simplifies to:

$$\begin{aligned}\langle F^N \rangle &= \frac{1}{N}\sum_{i=0}^{N-1}\frac{f(X_i)}{pdf_g(X_i)},\\ &= \frac{1}{N}\sum_{i=0}^{N-1}\frac{g(X_i)\,h(X_i)}{c\,g(X_i)},\\ &= \frac{1}{cN}\sum_{i=0}^{N-1} h(X_i).\end{aligned} \tag{A.25}$$

Since $h(x)$ will in general be smoother than the full $g(x)\,h(x)$, the variance of the Monte Carlo estimator is reduced.