This page shows how to estimate the errors in Monte Carlo simulations. General results from probability theory are illustrated with numerical examples.

To be specific, suppose that we need to estimate the fraction of power emitted by a light source that will be received by a detector. The answer of course depends on the water IOPs; the size, orientation, and location of the detector relative to the light source; the angular distribution of the emitted light; and perhaps on other things like boundary surfaces that can reflect or absorb light scattered onto the boundary. The numerical results to be shown below use the source and detector geometry shown in Fig. figure1. The source emits a collimated bundle of rays toward a detector that is $\tau = 5$ optical path lengths away and is 0.1 optical path lengths in diameter. The water IOPs have $a = 0.2 \, \mathrm{m}^{-1}$ and $b = 0.8 \, \mathrm{m}^{-1}$ so that $c = 1 \, \mathrm{m}^{-1}$ and one meter is one optical path length. The albedo of single scattering is $\omega_o = 0.8$. The scattering phase function is a Henyey-Greenstein phase function with a scattering-angle mean cosine of $g = 0.8$.
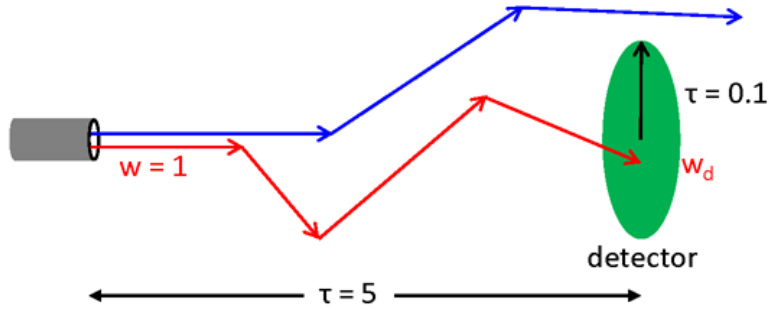

Figure 1: Source and detector geometry used for numerical simulations.

In the simulations, N rays will be emitted from the source, each with an initial weight of $w = 1$. Most of those rays will miss the detector, as illustrated by the blue arrows in Fig. figure1. However, some rays will hit the detector, as illustrated by the red arrows, at which time their current weight will be tallied to the accumulating total weight w_d received by the detector. After all rays have been traced, the the Monte Carlo estimate of the fraction of emitted power received by the detector is simply $f_d = w_d/N$.

If we do only one simulation tracing, say, $N = 10^4$ ray bundles, then the resulting estimate of f_d is all we have. In particular, we have no estimate of the statistical error in the estimated f_d.

Probability Theory

To develop a quantitative error estimate for the result of a Monte Carlo simulation, we begin with a review of some results from basic probability theory. Recalling the notation introduced on the Monte Carlo Introduction page, let $p_W(w)$ be the probability density function (pdf) for random variable W. Capital W represents the random variable, e.g. the ray bundle weight received by the detector, and lower case w represents a specific value of W, e.g., $w = 0.72$. In the present example, $p_W(w)$ is the pdf that a ray strikes the detector

with a weight $w, 0 \leq w \leq 1$. rays that miss the detector do not contribute to accumulating weight or power received by the detector and do not enter into the calculations below; they are simply wasted computer time. Note that we have no idea what mathematical form $p_W(w)$ has: it results from a complicated sequence of randomly determined ray path lengths and scattering angles.

For any continuous pdf $p_W(w)$ the *expected* or *mean value* of W is defined as

$$mean(W) \equiv \mu \equiv \mathcal{E}\{W\} = \int w \, p_W(w) \, dw \,, \tag{1}$$

where $\mathcal{E}$ denotes expected value and the integral is over all values for which W is defined. If the random variable is discrete, the integral is replaced by a sum over all allowed values of W. Similarly, the *variance* of W is defined as

$$var(W) \equiv \sigma^2 \equiv \mathcal{E}\{(W - \mu)^2\} = \int (w - \mu)^2 \, p_W(w) \, dw = \mathcal{E}\{W^2\} - [\mathcal{E}\{W\}]^2 \,.$$

Note that if c is a constant, then

$$\mathcal{E}\{cW\} = c\,\mathcal{E}\{W\} \quad \text{and} \quad var(cW) = c^2 \, var(W) \,.$$

Greek letters μ and σ^2 are used to denote the *true* or *population* mean and variance of a pdf.

Suppose that N_d ray bundles actually reach the detector. The total weight received by the detector is then given by the sum of these randomly determined weights:

$$S_{N_d} = \sum_{i=1}^{N_d} w_d(i) \,,$$

where $w_d(i)$ is the weight of ray bundle i when it reached the detector.

Each of the N ray bundles emitted by the source and traced to completion is independent of the others. In particular, a different sequence of random numbers is used to determine the path lengths and scattering angles for each emitted bundle. Moreover, the underlying pdfs for ray path length and scattering angles (i.e., the IOPs) are the same for each bundle. The random variables are then said to be *independent and identically distributed* (iid), and S_{N_d} is said to be a random sample of size N_d of random variable W.

The linearity of the expectation (i.e., the integral of a sum is the sum of the integrals) means that for iid random variables such as W,

$$\mathcal{E}\{S_{N_d}\} = N_d \, \mathcal{E}\{W\} = N_d \, \mu \quad \text{and} \quad var(S_{N_d}) = N_d \, var(W) = N_d \, \sigma^2 \,. \tag{2}$$

In the Monte Carlo simulation, the *sample mean*, i.e. the estimate of the average detected weight obtained by from the N_d detected ray bundles is

$$m_{N_d} \equiv \frac{1}{N_d} S_{N_d} = \frac{1}{N_d} \sum_{i=1}^{N_d} w_d(i) \,.$$

Equation (likesection2) now gives two extremely important results. First,

$$\mathcal{E}\{m_{N_d}\} = \frac{1}{N_d} \mathcal{E}\{S_{N_d}\} = \mu \,. \tag{3}$$

2

That is, the expectation of the sample mean m_{N_d} is equal to the true mean μ. The sample mean is then said to be an *unbiased estimator* of the true mean of the pdf. Second,

$$var(m_{N_d}) = var(\frac{1}{N_d}S_{N_d}) = \frac{1}{N_d^2}var(S_{N_d}) = \frac{\sigma^2}{N_d}. \tag{4}$$

Thus, the variance of the sample mean goes to zero as $N_d \to \infty$, that is, as more and more ray bundles are *detected.* In other words, the Monte Carlo estimate of the average power received by the detector is guaranteed to give a result that can be made arbitrarily close to the correct result if enough rays are detected. This result is known as *the law of large numbers.* Again, you can emit and trace all the rays you want, but if they don't hit the detector, they don't count.

It is often convenient to think in terms of the standard deviation, e.g., when plotting data and showing the spread of values. The standard deviation of the error in m_{N_d} is

$$s_{N_d} = \sqrt{var(m_{N_d})} = \frac{\sigma}{\sqrt{N_d}}. \tag{5}$$

The dependence of the standard deviation of the estimate on $1/\sqrt{N_d}$ *is a very general and important result.* However, this "approach to the correct value" is very slow. If we want to reduce the standard deviation of the error in the estimated average power received by the detector by a factor of 10, we must detect 100 times as many rays. That can be computationally very expensive.

It is to be emphasized that result (likesection4) that the variance of a sample mean equals the true variance divided by the sample size holds for any situation for which the individual samples are independent and identically distributed random variables.

Note, of course, that if we knew the pdf for the received power, $p_W(w)$, then we could simply evaluate Eq. (likesection1) to get the desired true mean μ, and no Monte Carlo simulation would be required.

Finally, it must be remembered that the discussion here assumes that "all else is the same" when considering the number of detected rays. For example, we emit and trace more rays without changing the physics of the simulation. The page on importance sampling presents ways to increase the number of detected rays and thereby reduce the variance, but with a change in the physics that sometimes may invalidate the simple $1/\sqrt{N_d}$ dependence.

Numerical Examples

Numerical simulations were performed for the geometry and conditions described for Fig. figure1. For these simulations, tracing type 1 of the previous page was used. That is, ray bundles were traced until they either hit the target (still with weight $w = 1$) or were absorbed. For a given number N of emitted ray bundles, various numbers N_{run} of independent runs were done. That is, N rays were traced and the number N_d of detected ray bundles and their weights were tallied for each run. The fraction of emitted power received by the detector was then computed by the total detected weight for the run divided by N. Then another run was made with everything the same except that a different sequence of random numbers

3

was used (i.e., each run was started with a different seed for the $U[0, 1]$ random number generator used to determine path lengths and scattering angles).

The first set of simulations used $N = 10,000$ emitted ray bundles for each run, with $N_{\text{run}} = 10, 100, 1000$, and $10,000$ runs being made. Figure figure2 shows the distributions of the sample means m_{N_d} and other information for these four sets of run numbers. The upper left panel of the figure is for only $N_{\text{run}} = 10$ runs, or trials, or simulations. In this panel, the histogram shows that one run, or 0.1 of the total number of runs, gave an estimated fraction between 0.0088 and 0.0090 of the emitted power; two runs, or 0.2 of the total, gave a fraction between 0.0104 and 0.0106, and so on. As the number of runs increases, the estimates of the fraction of power received range from slightly less than 0.008 to slightly more than 0.014, with most estimates centering somewhere near 0.0108.
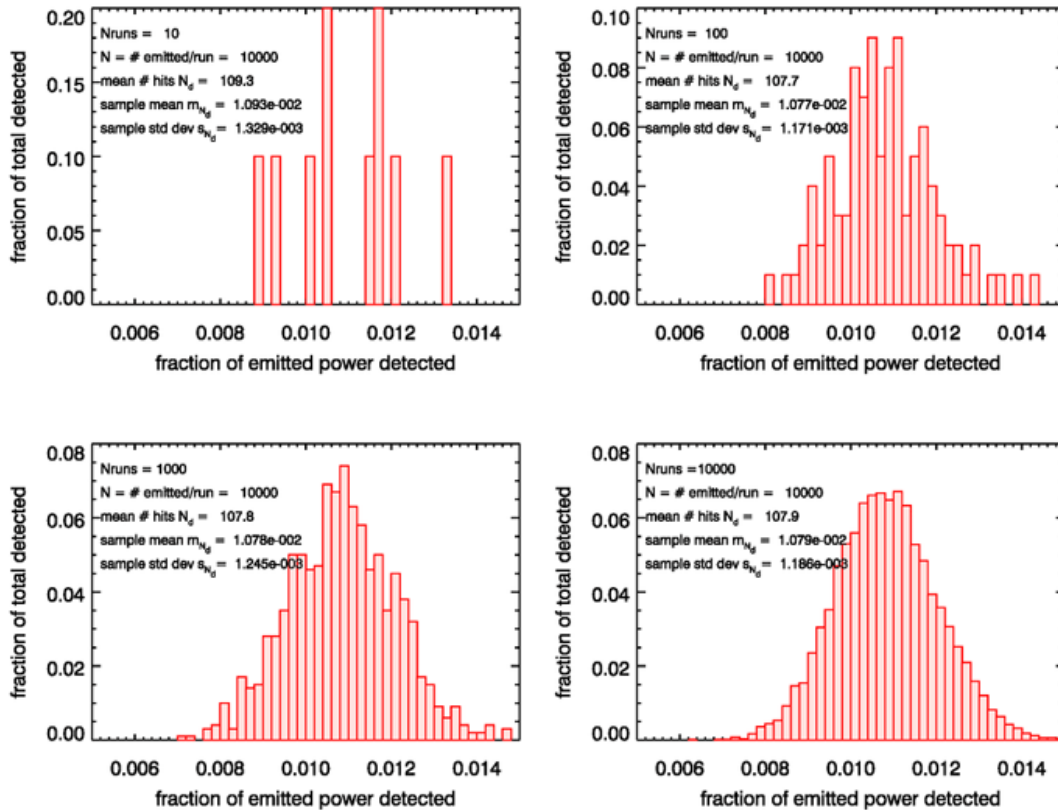

Figure 2: Estimates of the fraction of detected power for four sets of runs with $N = 10^4$ emitted rays in each run.

As the number of runs increases, something very remarkable happens: *the distribution of the fraction of the emitted power appears to be approaching a Gaussian or normal shape, even though the underlying pdf $p_W(w)$ is certainly not Gaussian.* This distribution can be thought of as the distribution of errors in the estimated mean of the distribution, or the distribution of $\mathcal{E}\{m_{N_d} - \mu\}$, where μ is the unknown true mean of the distribution $p_W(w)$. This approach to a Gaussian distribution is a consequence of the *central limit theorem. The central limit theorem states that the **sum** of a large number of independently distributed random variables*

with finite means and variances is approximately normally distributed regardless of what the distribution of the random variable itself may be. This is one of the most profound results in probability theory. Indeed, it explains why so many natural phenomena tend to have a Gaussian shape. Phenomena as disparate as average student exam scores, noise in electrical circuits, daily water usage in a city, or the fraction of people who develop cancer can all result from sums of many individual contributions. Such sums then tend toward a Gaussian distribution as the number of individual contributions increases. The theorem was first proved for a specific pdf in 1733, but it was not proven to hold for all pdfs (having finite means and variances) until the early 1900's. (By the way, in spite of what you see in the tabloid press, there is nothing in probability theory called "the law of averages." The central limit theorem is maybe the closest thing to the often involked buy mythical "law of averages.")

Figure figure3 shows the corresponding results for series of $N_{\text{run}} = 10, 100, 1000$, and 10,000 runs being made, but with each run now having $N = 100,000$ emitted rays. Now the spread in the estimated values is much less, from slightly less than 0.010 to about 0.012, again centering somewhere around 0.0108. Again, we see the approach to a Gaussian shape as more and more runs are made, but the Gaussian has a narrower width, i.e. less variance about the mean. The standard deviation of the sample estimates of the mean is usually called the "standard error of the mean."

The lower right panel of Fig. figure2 shows that the sample standard deviation for the case of 10,000 runs each with 10,000 emitted rays is $s_{N_d} = 1.186 \cdot 10^{-3}$ and the average number of detected rays is $N_d = 107.9$. The corresponding panel of Fig. figure3 shows $s_{N_d} = 3.718 \cdot 10^{-4}$ and $N_d = 1080.5$. The ratio of these sample standard deviations is

$$\frac{s_{N_d}(N_d = 107.9)}{s_{N_d}(N_d = 1080.5)} = \frac{1.186 \cdot 10^{-3}}{3.718 \cdot 10^{-4}} = 3.19\,.$$

The corresponding ratio of $\frac{\sigma}{\sqrt{N_d}}$ values is

$$\frac{\frac{\sigma}{\sqrt{N_d=107.9}}}{\frac{\sigma}{\sqrt{N_d=1080.5}}} = \sqrt{\frac{1080.5}{107.5}} = 3.16\,.$$

This nicely illustrates the dependence of the sample standard deviation, or the standard error of the mean, on the square root of the number of ray bundles detected, just as predicted by Eq. (likesection5).

Error Estimation

In the present example of the fraction of emitted power received by a detector, the central limit theorem guarantees that the errors in the fraction of received power computed by many Monte Carlo runs approaches a Gaussian. We can thus use all of the results for Gaussian, or normal, probability distributions to estimate the errors in the Monte Carlo results.

It is often desirable to know the probability that the computed sample mean m_{N_d} is within some prechosen amount, say 1 standard deviation, of the (unknown) true mean μ. Conversely, we may want to compute the error range so that the sample mean is within
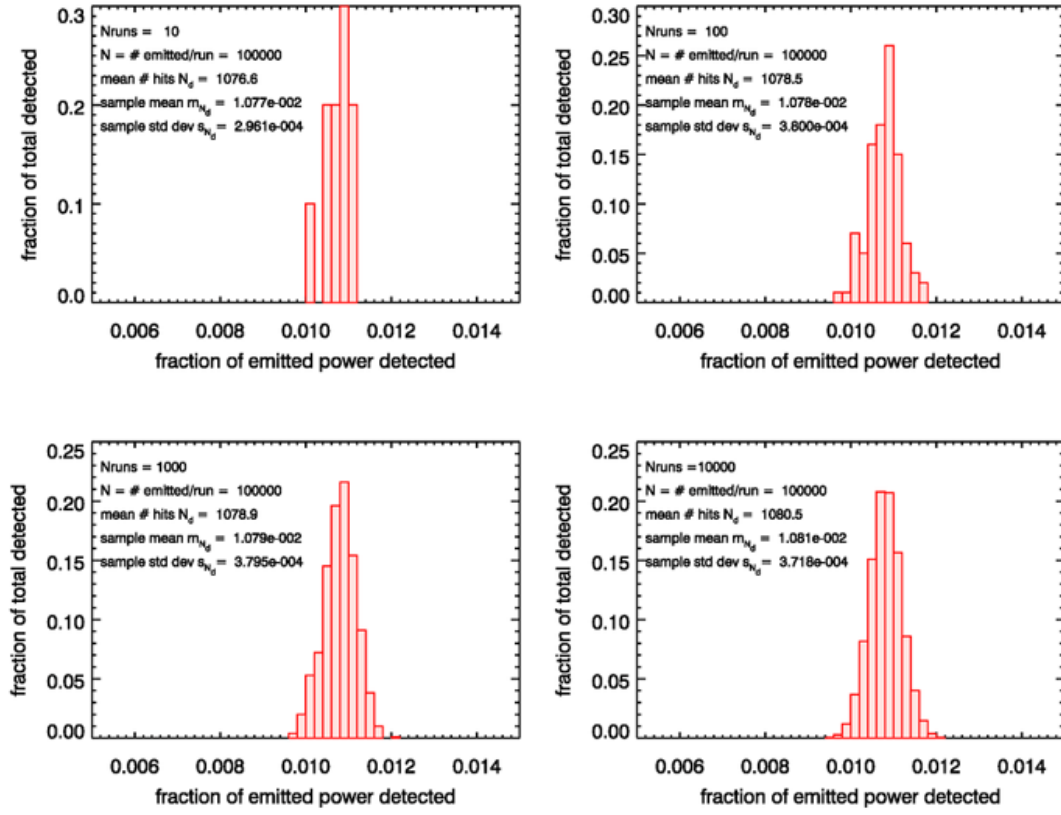
Figure 3: Estimates of the fraction of detected power for four sets of runs with $N = 10^5$ emitted rays in each run.

that error range of the true mean with some prechosen probability. Such questions can be answered starting with the statement

$$Prob\{\mu - \beta\, s_{N_d} \leq m_{N_d} \leq \mu + \beta\, s_{N_d}\} = 1 - \alpha\,.$$

This equation states that the probability is $1 - \alpha$ that the sample mean is within a range $\beta\, s_{N_d}$ of the true mean μ; β is a fraction of the sample standard deviation s_{N_d}. In other words, the probability is $1 - \alpha$ that $\mu = m_{N_d} \pm \beta\, s_{N_d}$. The central limit theorem guarantees that, if the sample size is large enough, the deviation of the sample mean from the true mean, $m_{N_d} - \mu$, is approximately normally distributed:

$$pdf(m_{N_d} - \mu) \approx \frac{1}{\sqrt{2\pi} s_{N_d}} \exp\left\{ -\frac{(m_{N_d} - \mu)^2}{2 s_{N_d}^2} \right\}\,. \tag{6}$$

Assuming that we have enough samples to get a good approximation to the normal distribution, the probability that m_{N_d} is greater than μ by an amount $\beta\, s_{N_d}$ is then

$$Prob\{m_{N_d} - \mu \geq \beta\, s_{N_d}\} = \frac{1}{\sqrt{2\pi} s_{N_d}} \int_{\beta s_{N_d}}^{\infty} \exp\left\{ -\frac{t^2}{2 s_{N_d}^2} \right\} dt\,.$$

Letting $y = t/s_{N_d}$ gives

$$Prob\{m_{N_d} - \mu \geq \beta\, s_{N_d}\} = \frac{1}{\sqrt{2\pi}} \int_{\beta}^{\infty} \exp\left\{ -\frac{y^2}{2} \right\} dy \equiv Q(\beta)\,. \tag{7}$$

The $Q(\beta)$ integral in the last equation cannot be performed analytically, but it is tabulated in probability texts, and software packages such as MATLAB and IDL have routines to compute it. In to also common to find tables and subroutines for

$$\Phi(\beta) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta} \exp\left\{ -\frac{y^2}{2} \right\} dy\,.$$

Note that $Q(\beta) + \Phi(\beta) = 1$.

Let us now apply Eq. (likesection7) to various examples. We first compute the probability that the sample mean is within one standard deviation of the true mean. Letting $\beta = 1$, we compute the probability $\alpha/2$ that $m_{N_d} - \mu$ lies in the "right-hand tail" of the normal distribution beyond $\beta = 1$. This is the area shaded in red in Fig. figure4.

This probability is

$$Prob\{m_{N_d} - \mu \geq s_{N_d}\} = Q(1.0) = \frac{\alpha}{2}\,.$$

$Q(1) \approx 0.1587$, so $\alpha = 2Q(\beta) = 0.3174$. The Gaussian distribution is symmetric about the mean, so $Prob\{m_{N_d} - \mu \leq -s_{N_d}\}$ that $m_{N_d} - \mu$ lies in the left-hand (green-shaded) tail of the distribution also equals 0.1587. Thus the probability that $m_{N_d} - \mu$ does *not* lie in either tail of the distribution, i.e. that $\mu \leq m_{N_d} \pm s_{N_d}$ is $1 - \alpha/2 - \alpha/2 = 0.6826$.

For the simulations of Fig. figure2, the lower right panel shows values of $m_{N_d} = 0.0179$ and $s_{N_d} = 1.186 \cdot 10^{-3}$. Thus there is a roughly 68% chance that the true fraction of received power is within the range $0.01079 \pm 1.186 \cdot 10^{-3}$. For the corresponding run of
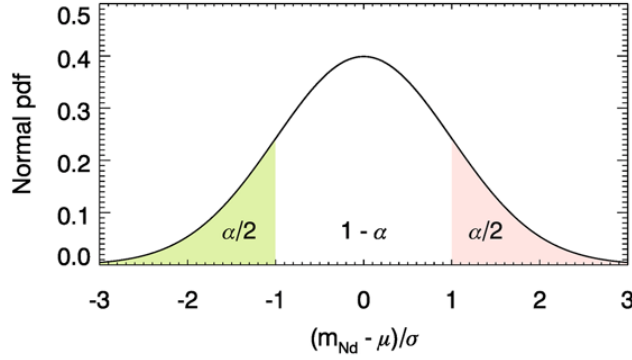
Figure 4: The Gaussian or normal distribution of Eq. likesection6

.

Fig. figure3, which traced ten times as many ray bundles, the corresponding numbers are $0.01081 \pm 3.718 \cdot 10^{-4}$. Thus, in the last set of simulations, we are 68% certain that the sample mean is within about $\pm 3.4\%$ of the true mean.

Suppose we need the probability that we are within, say, 5% of the correct value. For the lower right simulation of Fig. figure3, 0.05 of 0.01081 is about 0.00054. From $0.00054 = \beta 3.178 \cdot 10^{-4}$ we get $\beta = 1.454$. $Q(1.454) = 0.0729 = \alpha/2$, so the probability of being within 5% is $1 - \alpha = 0.854$. If being 85% confident that the Monte Carlo estimate of the mean is within 5% of the true value is adequate for your application, then you are done. If you need to be 95% confident that you are within 5% of correct, then you need to continue tracing rays until you get enough rays on the target to reduce the sample variance to a value small enough to achieve the desired 95% confidence.

As a final example, we might ask how big is the error so that we can say that we are within that range with 90% certainly. We now set $1 - \alpha = 0.9$, and solve

$$Q(\beta) = \frac{\alpha}{2} = 0.05$$

Again, the inverse of $Q(\beta)$ is also tabulated. This equation gives $\beta = 1.645$. From the last panel of Fig. figure3 we then get $\beta s_{N_d} = 1.645 \times 3.718 \cdot 10^{-4} = 6.12 \cdot 10^{-4}$, so that $\mu = 0.01081 \pm 6.12 \cdot 10^{-4}$ with 90% confidence.

As a final caveat to this section, keep in mind that the central limit theorem says that the error becomes Gaussian as the number of samples, N_d in the present examples, becomes very large. How large is large enough depends on the particular problem and the user's accuracy requirement. In the present examples, 10,000 runs each with 10,000 or more emitted rays, resulting in 100 or more detected rays for each run, gives distributions that visually appear close to Gaussian (the lower right panels of the preceding two figures). There are various ways to quantify how close a data distribution is to a Gaussian, but that is a topic for somewhere else. Just do a search on "normality tests."