Monte Carlo Integration. . . in a Nutshell

1 Preamble

In engineering analysis, we must often evaluate integrals defined over a complex domain or in a high-dimensional space. For instance, we might wish to calculate the volume of a complex three-dimensional part for an aircraft. We might even wish to evaluate a performance metric for an aircraft expressed as an integral over a very high-dimensional design space. Unfortunately, the deterministic integration techniques considered in the nutshell *Integration* are unsuited for these tasks: either the domain of integration is too complex to discretize, or the function to integrate is too irregular, or the convergence is too slow due to the high dimension of the space — the curse of dimensionality. In this nutshell, we consider an alternative integration technique based on random variates and statistical estimation: Monte Carlo integration.

We introduce in this nutshell the Monte Carlo integration framework. As a first application we consider the calculation of the area of a complex shape in two dimensions: we provide a statistical estimator for the area, as well as associated *a priori* and *a posteriori* error bounds in the form of confidence intervals; we also consider extension to higher space dimensions. Finally, we consider two different Monte Carlo approaches to integration: the "hit or miss" approach, and the sample mean method; for simplicity, we consider univariate functions.

Prerequisites: probability theory; random variables; statistical estimation.

2 Motivation: Example

Let us say that we are given a (pseudo-)random variate generator for the standard uniform distribution. How can we take advantage of this sample realization to estimate the value of the mathematical constant π? We first rephrase the question: how can we estimate the area of the unit disk — π — shown in Figure 1? How do we exploit our random variates? What can we say about the accuracy of our estimate? How does the cost of the estimate scale with the desired accuracy? The material in this nutshell will help you answer these questions.

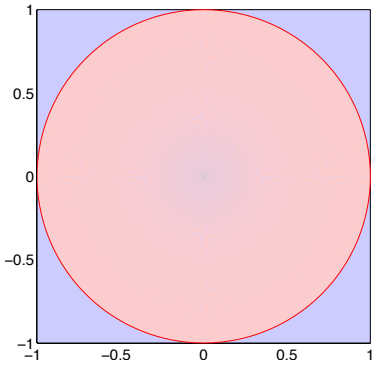


Figure 1: The estimation of π posed as the estimation of the area of the unit disk.

3 Area Estimation

We generalize our π question: we wish to estimate the area, A_D, of an arbitrary-shaped two–dimensional domain, $D \subset \mathbb{R}^2$, as illustrated in Figure 2. We now describe the associated Monte Carlo area estimation procedure in three steps.

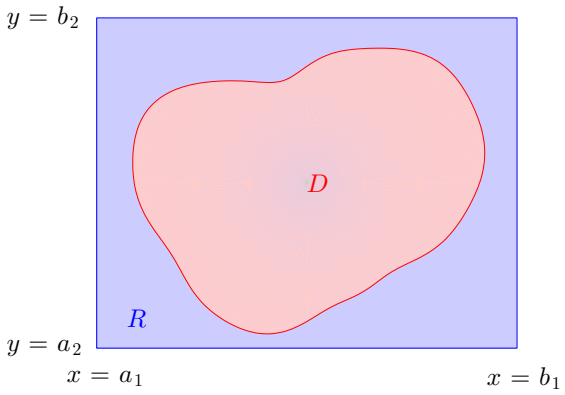


Figure 2: Illustration of a generic area estimation problem.

As the first step of Monte Carlo area estimation, we formulate the (deterministic) area estimation problem as the expectation of a random variable. Towards this end, we first introduce a rectangular region $R \equiv [a_1, b_1] \times [a_2, b_2]$ in which D resides, as shown in Figure 2; we denote the area of R by A_R, and we note that $A_R = (b_1 - a_1)(b_2 - a_2)$. We may think of R as a background, or bounding, rectangle. For instance, in our π-estimation example illustrated in Figure 1, D is the unit disk centered at the origin, and R may be chosen as the square $[-1, 1]^2 \equiv [-1, 1] \times [-1, 1]$ of area $A_R = 4$. (Any rectangle which includes D suffices, however we shall latter see that smaller is better.)

We now suppose that we randomly choose a point in the rectangle R. We describe the coordinate of the point as (X, Y) for X (respectively Y) a univariate continuous uniform random variable over the interval (a_1, b_1) (respectively (a_2, b_2)). We now observe — from the properties of the continuous uniform distribution — that the probability that our point (X, Y) resides in the region D is the

ratio of the area of the two regions:

$$P((X,Y) \in D) = \frac{A_D}{A_R} \, .$$

We next introduce a Bernoulli random variable,

$$B = \begin{cases} 0, & (X,Y) \notin D \\ 1, & (X,Y) \in D \end{cases} \quad ;$$

note that B is a function of the random variables (X,Y). We now note that the expectation of the random variable B is

$$E(B) = P((X,Y) \in D) = \frac{A_D}{A_R} \equiv \theta \, ,$$

where θ is the usual Bernoulli parameter (associated with B). We have thus recast the area-estimation problem as a mean-estimation problem — more precisely, estimation of the Bernoulli parameter θ. We have studied the latter extensively in the nutshell *Random Variables*.

CYAWTP 1. Let D be the region which results from the intersection of two unit disks, one of which is centered at $(0,0)$, and the other of which is centered at $(1,1)$. Sketch the region D. Introduce and sketch the rectangle R of smallest area and aligned with the coordinate axes which includes D. Evaluate A_R.

In the second step of Monte Carlo area estimation, we design a procedure to generate random variates of the Bernoulli random variable B. Here we employ a two-stage strategy. We first generate a (pseudo-)random variate of the pair of (independent) uniform random variables

$$(X,Y) \to (x,y). \tag{1}$$

We next determine if (x,y) is in D, and generate a (pseudo-)random variate of B, b, according to

$$b = \begin{cases} 0, & (x,y) \notin D \\ 1, & (x,y) \in D \end{cases} \quad . \tag{2}$$

We note that there are two key ingredients to generate a random variate of B. The first ingredient is a *(pseudo-)random variate generator* for uniform random variables; as discussed in the nutshell *Random Variables*, a (pseudo-)random variate generator for the standard uniform distribution is readily available, and simple transformations allow us to subsequently consider any interval. The second ingredient is a computational implementation of the rule (2) which, for any $(x,y) \in R$, determines whether (x,y) is in D; the rule often takes the form of a simple algebraic relation that reflects the geometry of the region D. Note that our rule allows us to evaluate b even though we do not know — and indeed, wish to estimate — the Bernoulli parameter, θ. As an example, we again consider our π-estimation problem of Figure 1: we may characterize D as $D \equiv \{(x,y) \mid x^2+y^2 \leq 1^2\}$; the value of the Bernoulli random variate is then determined from

$$b = \begin{cases} 0, & x^2+y^2 > 1 \\ 1, & x^2+y^2 \leq 1 \end{cases} \quad . \tag{3}$$

CYAWTP 2. Consider the region D introduced in **CYATWP 1**. Identify the appropriate parameters (respective lower and upper bounds) of the uniform random variables X and Y. Identify the algebraic rule, similar to (3) for the π-estimation problem, which determines the value of the "not–in" *vs* "in" Bernoulli random variate.

In the third step of Monte Carlo area estimation, we apply standard statistical techniques to estimate the mean of B, θ, from which we can then deduce the desired area as $A_R = A_D \cdot \theta$. We follow the procedure introduced in the nutshell *Random Variables* for estimation of a Bernoulli parameter. We first generate, by repeated application of the two-stage procedure described above, n random variates of B, $(B \to b_i)_{i=1,\ldots,n}$. We then construct our mean estimate from the sample mean,

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} b_i \ .$$

Finally, we appeal to the relationship $A_D = A_R \cdot \theta$ to provide the estimate of the area A_D,

$$(\hat{A}_D)_n = A_R \cdot \hat{\theta}_n.$$

CYAWTP 3. Consider the region D introduced in **CYATWP 1**. We invoke the procedure described in **CYATWP 2** to obtain $n = 20$ Bernoulli random variates,

$$0\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1.$$

Evaluate the mean estimate $\hat{\theta}_n$ and the area estimate $(\hat{A}_D)_n$.

We now proceed with an *a priori* error analysis of the Monte Carlo estimate. Towards this end, we recall that our mean estimate $\hat{\theta}_n$ is a realization of the sample-mean estimator $\hat{\Theta}_n$, $\hat{\Theta}_n = \frac{1}{n}\sum_{i=1}^{n} B_i$. We recall that $\hat{\Theta}_n$ is of mean θ and variance $\theta(1-\theta)/n$. Furthermore, for sufficiently large n,

$$P\left(|\theta - \hat{\Theta}_n| \le z_\gamma \sqrt{\frac{\theta(1-\theta)}{n}}\right) \approx \gamma \ . \tag{4}$$

Hence, with confidence level γ,

$$|\theta - \hat{\theta}_n| \lesssim (z_\gamma \sqrt{\theta(1-\theta)})\, n^{-1/2} \tag{5}$$

Here "$a_n \lesssim b_n$" is a shorthand for "$a_n < b_n$ as $n \to \infty$." We now mutiply (5) through by A_R and recall that $A_D = A_R \cdot \theta$ to obtain our *absolute* (γ-confidence) statistical *a priori* error bound,

$$|A_D - (\hat{A}_D)_n| \lesssim \left(z_\gamma A_R \sqrt{\frac{A_D}{A_R}\left(1 - \frac{A_D}{A_R}\right)}\right) n^{-1/2} \ , \tag{6}$$

We recognize $C \equiv z_\gamma A_R \sqrt{\frac{A_D}{A_R}(1 - \frac{A_D}{A_R})}$ as the constant of absolute error convergence, and $p = 1/2$ as the convergence rate. We may also divide the expression (6) through by A_D to obtain a *relative* (γ-confidence) statistical *a priori* error bound,

$$\frac{|A_D - (\hat{A}_D)_n|}{A_D} \lesssim \left(z_\gamma \sqrt{\frac{1 - A_D/A_R}{A_D/A_R}}\right) n^{-1/2} \ , \tag{7}$$

where we shall refer to $C_{\rm rel} \equiv \sqrt{\frac{1-A_D/A_R}{A_D/A_R}}$ as the constant of relative error convergence.

We make a few observations:

1. The statistical error bound $|A_D - (\hat{A}_D)_n|$ tends to 0 as $n \to \infty$: our Monte Carlo estimate is *convergent*.

2. The convergence rate, or order, of the Monte Carlo method is $p = 1/2$, which is, unfortunately, rather slow.

3. The constant of the relative error convergence is $C_{\rm rel} = \sqrt{\frac{1-A_D/A_R}{A_D/A_R}}$; we expect the quality of Monte Carlo estimate, in the sense of relative error, to degrade when A_D/A_R $(= \theta)$ is small.

4. The constant of the relative error convergence depends only on the ratio of the two areas A_D/A_R $(= \theta)$ and is independent of the particular shape of A_D: the (bound for the) error in the Monte Carlo estimate is independent of the geometric complexity of the region D.

5. The convergence rate, $p = 1/2$, is independent of the dimension of the problem, as we discuss further in Section 4.

We will focus exclusively in what follows on the more demanding relative error.

We now take the (base-10) logarithm of the *a priori* error bound (7) to obtain the logarithmic convergence asymptote

$$\log_{10}\left(\frac{|A_D - (\hat{A}_D)_n|}{A_D}\right) \lesssim \log_{10}(C_{\rm rel}) - \frac{1}{2}\log_{10}(n)\ ;$$

we anticipate our logarithmic convergence asymptote to have an intercept of order $\log_{10}(C_{\rm rel})$ and a slope of $-p = -1/2$. We present in Figure 3 a typical logarithmic convergence curve and logarithmic convergence asymptote associated with our π-estimation problem. Here "typical" is in reference to the random nature of our estimate: each realization shall be different. On a related note, because the error associated with the Monte Carlo estimate is noisy (within a single realization), here we redefine the logarithmic convergence asymptote as the straight line that bounds the error from above.

CYAWTP 4. Consider a region described in polar coordinates as

$$D_\beta \equiv \left\{(r,\theta) \mid r \le \frac{2}{3} + \frac{1}{3}\cos(4\beta\theta),\ 0 \le \theta \le \frac{\pi}{2}\right\},$$

where $r \equiv \sqrt{x^2+y^2}$, $\tan(\theta) = y/x$, and β is a positive integer parameter. Sketch, in two separate plots, the regions D_{β_1} and D_{β_2} associated with $\beta_1 = 1$ and $\beta_2 = 4$, respectively. Now suppose we invoke our Monte Carlo procedure to estimate the two areas $A_{D_{\beta_1}}$ and $A_{D_{\beta_2}}$; for each case, we choose for our background (bounding) rectangle $R \equiv [0,1]^2$. Sketch the logarithmic convergence asymptote for the two area estimates and compare the respective intercepts and slopes. (Note that we can readily show $A_{D_{\beta_1}} = A_{D_{\beta_2}} = \pi/8$.)

Numerical Experiment 5. Invoke the Monte Carlo Area GUI and empirically confirm your predictions of **CYATWP 4**.

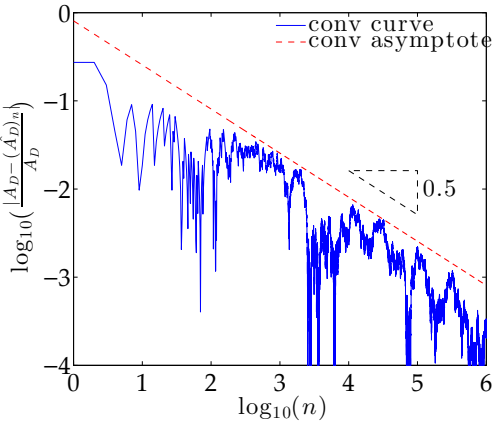


Figure 3: Convergence of the Monte Carlo area estimate for a realization of the π-estimation problem.

CYAWTP 6. Suppose we wish to estimate (i) the area of the square $D_1 \equiv [0.2, 0.7]^2$, and (ii) the area of the square $D_2 \equiv [0.2, 0.3]^2$, in both cases for a background rectangle $R \equiv [0, 1]^2$. Sketch the *relative* logarithmic convergence asymptote for these two area estimates and compare the respective intercepts and slopes.

Numerical Experiment 7. Invoke the Monte Carlo Area GUI and empirically confirm the results of **CYATWP 6**.

We now proceed with an *a posteriori* error analysis of our Monte Carlo estimate. An *a priori* error estimate is typically formulated in terms of unknown quantities — in our case, the mean θ — and serves to provide general insight into the requirements for, and general behavior of, convergence. In contrast, an *a posteriori* error bound is formulated in terms of known or readily calculable quantities and serves to certify a particular instantiation — a particular estimate for a particular problem. To construct an *a posteriori* error bound, we first recall that $\theta \in [\mathrm{CI}_\theta]_n(\hat{\Theta}_n, \gamma)$ with probability γ for

$$[\mathrm{CI}_\theta]_n(\hat{\Theta}_n, \gamma) \equiv \left[\frac{\hat{\Theta}_n + \frac{z_\gamma^2}{2n} - z_\gamma \sqrt{\frac{\hat{\Theta}_n(1-\hat{\Theta}_n)}{n} + \frac{z_\gamma^2}{4n^2}}}{1 + \frac{z_\gamma^2}{n}}, \frac{\hat{\Theta}_n + \frac{z_\gamma^2}{2n} + z_\gamma \sqrt{\frac{\hat{\Theta}_n(1-\hat{\Theta}_n)}{n} + \frac{z_\gamma^2}{4n^2}}}{1 + \frac{z_\gamma^2}{n}} \right];$$

here $z_\gamma = 1.96$ for the (typical) choice $\gamma = 95\%$. We then consider a realization of $[\mathrm{CI}_\theta]_n(\hat{\Theta}_n, \gamma)$, $[\mathrm{ci}_\theta]_n(\hat{\theta}_n, \gamma)$, such that $\theta \in [\mathrm{ci}_\theta]_n(\hat{\theta}_n, \gamma)$ with confidence level γ:

$$[\mathrm{ci}_\theta]_n(\hat{\theta}_n, \gamma) \equiv \left[\frac{\hat{\theta}_n + \frac{z_\gamma^2}{2n} - z_\gamma \sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n} + \frac{z_\gamma^2}{4n^2}}}{1 + \frac{z_\gamma^2}{n}}, \frac{\hat{\theta}_n + \frac{z_\gamma^2}{2n} + z_\gamma \sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n} + \frac{z_\gamma^2}{4n^2}}}{1 + \frac{z_\gamma^2}{n}} \right].$$

We may finally appeal to the relationship $A_D = A_R \cdot \theta$ to obtain a confidence interval for A_D,

$$[\mathrm{ci}_{A_D}]_n = A_R \cdot [\mathrm{ci}_\theta]_n \ ,$$

which we may readily evaluate in terms of our (known) estimate $\hat{\theta}_n$ and other available quantities.

CYAWTP 8. Consider the region D introduced in **CYATWP 1**. We invoke the procedure describe in **CYATWP 2** to obtain $n = 20$ Bernoulli random variates,

$$0\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1.$$

Evaluate the 95% confidence interval for θ, $[\mathrm{ci}_\theta]_n$, and the 95% confidence interval for, A_D, $[\mathrm{ci}_{A_D}]_n$. Does the latter confidence interval include the true area, $A_D \approx 0.5708$?

4 Volume Estimation: Higher Dimensions

We briefly comment on the calculation of volumes in higher dimensions. The three-step Monte Carlo area estimation procedure described in Section 3 readily extends to higher dimensions with only minor modification.

1. We introduce a d-dimensional background rectangular region $R \equiv [a_1, b_1] \times \cdots \times [a_d, b_d]$ of volume $A_R^{(d)}$ which encloses D. We next introduce a d-dimensional random vector $(X^{(1)}, \ldots, X^{(d)})$ for $X^{(j)}$ a univariate continuous uniform random variable over the interval (a_j, b_j), $j = 1, \ldots, d$. We then define, as in the two-dimensional case, a Bernoulli random variable B which takes on the value of 0 for $(X^{(1)}, \ldots, X^{(d)}) \notin D$ and 1 for $(X^{(1)}, \ldots, X^{(d)}) \in D$.

2. We invoke the d-dimensional generalization of (1) to generate random variates,

$$(X^{(1)}, \ldots, X^{(d)}) \rightarrow (x^{(1)}, \ldots, x^{(d)}) \ .$$

   We then generate random variates for B based on a d-dimensional generalization of the rule (2): $b = 0$ if $(x^{(1)}, \ldots, x^{(d)}) \notin D$, and $b = 1$ if $(x^{(1)}, \ldots, x^{(d)}) \in D$.

3. We invoke the statistical estimation procedure exactly as in the two-dimensional case: the algorithm "sees" only the Bernoulli random variates b_i, $i = 1, \ldots, n$.

We emphasize that because the statistical estimation procedure is identical to the two-dimensional case, the relative statistical *a priori* error bound is, as before,

$$\frac{|A_D^{(d)} - (\hat{A}_D^{(d)})_n|}{A_D^{(d)}} \leq \left(z_\gamma \sqrt{\frac{1 - A_D^{(d)}/A_R^{(d)}}{A_D^{(d)}/A_R^{(d)}}} \right) n^{-1/2} \ .$$

Note that, as expected from the two-dimensional case, the constant in the relative error estimate increases as the ratio of the two volumes, $\theta \equiv A_D^{(d)}/A_R^{(d)}$, decreases; note furthermore that $A_D^{(d)}/A_R^{(d)}$ may decrease with spatial dimension, d. However, and most importantly, the convergence rate, $p = 1/2$, is independent of the spatial dimension, d. In this sense, the Monte Carlo method, unlike classical deterministic methods, breaks the *curse of dimensionality*: the convergence rate does note degrade with increasing spatial dimension, d.

As an example, we consider estimation of the volume of a d-dimensional unit ball. We choose $R \equiv [-1, 1]^d$. We present in Figure 4 typical relative error convergence results for $d = 2, 3$, and 4; as predicted by theory, the convergence rate is independent of the dimension d. (The constant of the relative error convergence somewhat increases with the dimension because the relative volume of the d-ball with respect to the d-cube, $\theta \equiv A_D^{(d)}/A_R^{(d)}$, decreases with d.)

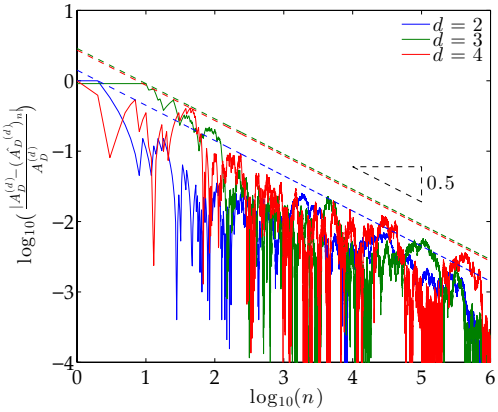


Figure 4: Relative error convergence for estimation of the volume of the d-dimensional unit sphere for $d = 2, 3$, and 4.

5 Monte Carlo Integration: "Hit or Miss" Method

We now apply the Monte Carlo area estimation technique to integrate a univariate function g over the interval (a, b),

$$I = \int_{x=a}^{b} g(x)dx \ ;$$

for simplicity we assume $g(x) > 0$ for all $a \leq x \leq b$.

We first recast the integral as an area estimation problem. Towards this end, we introduce the rectangular region

$$R \equiv [a, b] \times [0, c]$$

for $c \geq \max_{x \in (a,b)} g(x)$; we also introduce the target region

$$D \equiv \{(x, y) \in R \mid y \leq g(x)\}.$$

We now note that

$$I = \int_{x=a}^{b} g(x)dx = \int_{x=a}^{b} \int_{y=0}^{g(x)} dxdy = A_D.$$

We have thus transformed the function integration problem into an area estimation problem, as depicted in Figure 5. This Monte Carlo integration method as known as the "hit or miss" approach, because the approximation is based on the hit-or-miss estimate of the area.

We outline the rest of the estimation procedure. We introduce X (respectively Y) a univariate continuous uniform random variable over the interval (a, b) (respectively $(0, c)$). We then generate n associated random variates

$$((X, Y) \rightarrow (x_i, y_i))_{i=1,\ldots,n} \ .$$

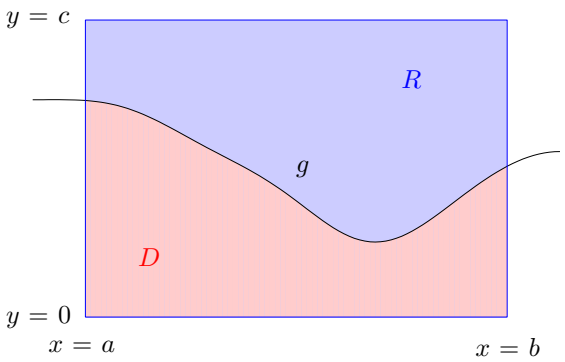


Figure 5: Illustration of univariate function integration recast as a Monte Carlo area estimation problem.

We then introduce random variates of the Bernoulli variable B, b_i, according to

$$b_i = \begin{cases} 0, & y_i > g(x_i) \\ 1, & y_i \leq g(x_i) \end{cases} \quad , \quad i = 1, \ldots, n \, . \tag{8}$$

Note this relation (8) derives directly from our general "not-in" *vs* "in" Bernoulli assignment rule, (2). Finally, we construct the integral estimate as

$$\hat{I}_n = A_R \cdot \hat{\theta}_n = c(b-a) \cdot \frac{1}{n} \sum_{i=1}^{n} b_i .$$

CYAWTP 9. Consider the integral $I = \int_0^1 g(x)dx$ for $g(x) \equiv \sin^2(\pi x)$. Instantiate the Bernoulli assignment rule (8) for estimation of this particular integral by the hit or miss approach.

CYAWTP 10. Suppose the function g that we wish to integrate takes on negative values in the interval (a, b), but is bounded from below:

$$g(x) \geq -m \, , \quad a \leq x \leq b \, . \tag{9}$$

Reformulate the hit or miss method such that the method may now treat this more general case. Note you may assume that you know a value of m which satisfies the necessary inequality, (9).

We proceed with an *a priori* error analysis. As before, we obtain a relative (γ-confidence) statistical *a priori* error bound,

$$\frac{|I - \hat{I}_n|}{I} \leq \left(z_\gamma \sqrt{\frac{1 - I/R}{I/R}} \right) n^{-1/2} .$$

We make a few observations, and in particular compare the result with the deterministic integration techniques introduced in a previous nutshell.

1. The constant of relative error convergence depends only on the ratio of the integral to the area of the backgound (bounding) box, I/R ($\equiv \theta$). Clearly we should choose c as small as possible.

2. The error bound does not depend on the regularity of the function: the function need not be smooth for the relationship to hold.

3. The error bound does not depend on the magnitude of the derivatives of g.

4. The method provides a reasonable estimate for small n: there is no resolution requirement.

5. The convergence is however slow: $p = 1/2$.

(Note that we might debate the resolution issue: in cases in which we must choose R (to contain g) such that θ is perforce small, we will not observe convergence until we throw sufficient darts to find $g(x)$. In such situations, if possible, we might break up the interval into different segments with adapted background rectangles over each segment; this is an example of "importance sampling.")

We finally proceed with an *a posteriori* error analysis. As before, the computable estimate of the confidence interval for θ is

$$[\mathrm{ci}_\theta]_n(\hat{\theta}_n, \gamma) = \left[\frac{\hat{\theta}_n + \frac{z_\gamma^2}{2n} - z_\gamma \sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n} + \frac{z_\gamma^2}{4n^2}}}{1 + \frac{z_\gamma^2}{n}}, \; \frac{\hat{\theta}_n + \frac{z_\gamma^2}{2n} + z_\gamma \sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n} + \frac{z_\gamma^2}{4n^2}}}{1 + \frac{z_\gamma^2}{n}} \right].$$

It follows that the confidence interval for $I = c(b-a) \cdot \theta$ is

$$[\mathrm{ci}_I]_n = c(b-a) \cdot [\mathrm{ci}_\theta]_n.$$

As always, these statements must be intepreted as "with confidence level γ."

6 Monte Carlo Integration: Sample-Mean Method

We introduce briefly another method to estimate the integral of a function: the sample-mean method. For simplicity, we develop the method for the integration of univariate functions. We recall that, given a function g and a random variable X defined over (a, b) with a probability density function f_X, the expectation of the random variable $Y = g(X)$ is

$$E(Y) = E(g(X)) = \int_a^b g(x) f_X(x) dx;$$

the rule is sometimes known as *law of the unconscious statistician*. Recall that the rule allows us to compute $E(Y)$ without explicit construction of, or reference to, f_Y. We now introduce a continuous uniform random variable X over (a, b); we also introduce a random variable $Y = g(X)$ for g the function we wish to integrate. We then note that the integral of g over (a, b) can be expressed as

$$\begin{aligned} I &= \int_a^b g(x)dx = (b-a) \int_a^b g(x) \frac{1}{b-a} dx = (b-a) \int_a^b g(x) f_X^{\mathrm{unif}}(x) dx \\ &= (b-a)E(g(X)) = (b-a)E(Y). \end{aligned}$$

In other words, the integral is $(b-a)$ times the mean of $E(Y)$. We have recast the integration problem as a mean-estimation problem.

We may readily apply standard statistical techniques to estimate the mean. Specifically, we first generate n random variates of X,

$$(X \rightarrow x_i)_{i=1,\ldots,n} \ .$$

We next generate n random variates of Y,

$$(Y \rightarrow y_i \equiv g(x_i))_{i=1\ldots,n} \ .$$

Note that we do not draw directly from f_Y: we draw x from f_X — uniform — and then evaluate $y = g(x)$. Finally, we estimate the integral as

$$\hat{I}_n = (b-a)\bar{y}_n = (b-a)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = (b-a)\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right),$$

which is our sample mean estimate of the integral. Note that $\bar{y}_n \equiv \frac{1}{n}\sum_{i=1}^{n} y_i$ is a realization of the sample mean of Y, $\bar{Y}_n$, which in turn is our estimator for $E(Y)$.

We briefly discuss the characteristics of the estimate. We first note that $E((b-a)\bar{Y}_n) = I$ and $E(\ ((b-a)\bar{Y}_n - I)^2\ ) = (b-a)^2\sigma_Y^2/n$; it follows that $(b-a)\bar{y}_n$ converges to I as $n \rightarrow \infty$, and in particular the error converges as $1/\sqrt{n}$ as in the Monte Carlo area estimate. Second, we note that $\sigma_Y^2 = \frac{1}{b-a}\int_a^b (g(x) - \bar{g})^2 dx$, where $\bar{g} \equiv \frac{1}{b-a}\int_a^b g(x)dx$; the error depends not on the smoothness of the function g, but rather on the "distribution" of the function.

CYAWTP 11. Consider the integral $I \equiv \int_{x=0}^{1/2} 3x dx$. We wish to estimate the integral using the sample mean method for $n = 5$. Our (pseudo-)random variates associated with X, a univariate continuous uniform random variable over the interval $(0, 1/2)$, are

$$0.3780 \quad 0.2353 \quad 0.1438 \quad 0.4396 \quad 0.0735.$$

Evaluate the sample-mean estimate of the integral, $\hat{I}_n$.

We remark that the sample-mean method for integration readily extends to higher spatial dimensions: we generate uniform random variates $x_i, 1 \leq i \leq n$, over the d-dimensional domain of interest, D; we next apply our transformation $g(x_i)$ to form the random variates $y_i, 1 \leq i \leq n$; we then evaluate the sample mean, $\bar{y}_n$, and scale by the volume of D to obtain our estimate for the integral.

CYAWTP 12. Consider the case in which we wish to perform an integral over a domain D which is, say, a disk in two spatial dimensions. How might we generate uniform random variates $x_i, 1 \leq i \leq n$, over the disk (as the first step in sample-mean estimation of an integral over D).

7 Perspectives

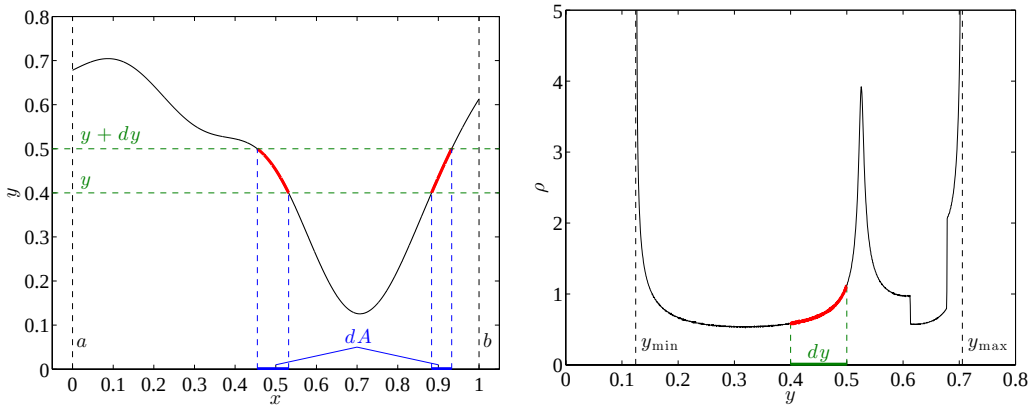
We have provided here only a first look at the topic of Monte Carlo integration. A more in-depth study may be found in *Math, Numerics, and Programming (for Mechanical Engineers)*, M Yano, JD Penn, G Konidaris, and AT Patera, available on MIT OpenCourseWare, which adopts similar notation to these nutshells and hence can serve as a companion reference. For a much more comprehensive view from both the computational and theoretical perspectives, we recommend *Simulation and the Monte Carlo Method*, RY Rubinstein, John Wiley & Sons, 2000. We note that the latter provides a more general perspective on pseudo-random variates and Monte Carlo methods; the application of the Monte Carlo method to deterministic problems (recast as random experiments) — as considered in this nutshell — is just one facet of a much broader framework.

Appendix A Monte Carlo Integration from the Perspective of Lebesgue Integration

We provide here an interpretation of the sample mean imethod in terms of Lebesgue integration. We consider, as before, the integral

$$I = \int_{x=a}^{b} g(x)\, dx \ .$$

More generally, we might consider a domain of integration D in d dimensions; in our simple case here, $D \equiv (a, b)$ and $d = 1$. We first recall that the deterministic integration techniques introduced in a previous nutshell build on the idea of *Riemann integration*. In the Riemann integral, we partition the domain of integration into small segments δx and sum the contributions $g(x)\delta x$; we then consider the limit $\delta x \to 0$. A typical numerical approximation suspends the limit: we consider a Riemann sum for finite δx. The accuracy of such a scheme, for example the rectangle rule, will depend on δx, the dimension of the domain, d, the complexity of the domain D, and the regularity of the integrand g. In contrast, Monte Carlo integration builds on the idea of *Lebesgue integration*, as we now describe.



(a) definition of differential area $dA(y; dy)$ (b) definition of area density $\rho(y)$

Figure 6: Illustration of Lebesgue integration.

To introduce the Lebesgue integral, we partition the *range* of the function into small segments dy and "sum" $y\, dA(y; \delta y)$. Here the differential area $dA(y; \delta y)$ is defined by

$$dA(y; dy) \equiv \{x \in (a, b) \,|\, y \le g(x) \le y + dy\} \ ;$$

roughly, $dA(y; dy)$ is the area (in our case, length) of the subdomain for which the function $g(x)$ takes on values in $(y, y+dy)$, as depicted in Figure 6(a). We may then reformulate I, an integration with respect to the *domain of the function*, as an equivalent integral, I^{Leb}, with respect to the *range of the function*:

$$I \equiv \int_{x=a}^{b} g(x)\, dx \quad = \quad I^{\text{Leb}} \equiv \int_{y=y_{\min}}^{y_{\max}} y\, dA(y; dy) \ . \tag{10}$$

We may further break the interval $[y_{\min}, y_{\max}]$ into (say) $N-1$ equisized segments $S_j, 1 \leq j \leq N-1$, in terms of which we can write (10) as

$$I^{\text{Leb}} = \sum_{j=1}^{N-1} \int_{S_j} y \, dA(y; dy) \tag{11}$$

$$= \sum_{j=1}^{N-1} \langle y \rangle_j \; A_j \; , \tag{12}$$

where

$$\langle y \rangle_j \equiv \frac{1}{A_j} \int_{S_j} y \, dA(y; dy), \quad A_j \equiv \int_{S_j} dA(y; dy) \; , \tag{13}$$

for $1 \leq j \leq N-1$. The expression (12) shall be the point of departure for our numerical formulation.

We digress briefly. We may view (10) as a Riemann-Stieljes integral. We may also define, when appropriate, an area density

$$\rho(y) \equiv \frac{dA}{dy} \; ,$$

such that $dA(y) = \rho(y)dy$, as depicted in Figure 6(b). (For our one-dimensional example, the density near the smooth (local) maxima and minima of $g(x)$ develops an inverse square-root singularity.) We can thus write

$$I^{\text{Leb}} = \int_{y=y_{\min}}^{y_{\max}} y \, dA(y) = \int_{y=y_{\min}}^{y_{\max}} y\rho(y)dy \; , \tag{14}$$

which we can understand as a kind of mean with respect to a density ρ. The expression (14) is convenient for "Riemannian" visualization of the Lebesgue integral.

We now develop a numerical approximation to I^{Leb} of (12). We first generate n i.i.d. random variates of random variable X uniformly distributed over (a, b): $(X \rightarrow x_i)_{i=1\,\ldots,n}$. We next generate, from our sample $(x_1, x_2, \ldots, x_n)$, n i.i.d. random variates of random variable $Y = g(X) : (Y \rightarrow y_i \equiv g(x_i))_{i=1,\ldots,n}$. We then form our approximation $\hat{I}_n^{\text{Leb}}$ to I^{Leb} of (12) as

$$\hat{I}_n^{\text{Leb}} \equiv \sum_{j=1}^{N-1} \left(\left(\frac{1}{\#(S_j)} \sum_{k=1}^{\#(S_j)} y_{k;j} \right) \left(\frac{\#(S_j)}{n} (b-a) \right) \right) \; . \tag{15}$$

Here $\#(S_j), 1 \leq j \leq N-1$, is the number of random variates $y_i, 1 \leq i \leq n$, in interval S_j, and $y_{k;j}, 1 \leq k \leq \#(S_j), 1 \leq j \leq N-1$, is a relabeling of the $y_i, 1 \leq i \leq n$, such that the $y_{k;j}$ reside in S_j. We can easily motivate (15): the term

$$\frac{1}{\#(S_j)} \sum_{k=1}^{\#(S_j)} y_{k;j} \tag{16}$$

is a plausible approximation to $\langle y \rangle_j$ of (12) (as defined in (13)); the term

$$\frac{\#(S_j)}{n} (b-a) \tag{17}$$

is a plausible (in expectation, or limit $n \to \infty$) approximation to A_j of (12) (as defined in (13)).

We now note that

$$\hat{I}_n^{\text{Leb}} = \frac{(b-a)}{n} \sum_{j=1}^{N-1} \sum_{k=1}^{\#(S_j)} y_{k;j} \;=\; \frac{(b-a)}{n} \sum_{i=1}^{n} y_i \;=\; \hat{I}_n^{\text{sample mean}} \,. \tag{18}$$

(We note that the segments $S_j, 1 \le j \le N-1$, are purely an artifact of interpretation. We could also choose for segments the Voronoi diagram induced by the sample $y_i, 1 \le i \le n$.) We may thus interpret Monte Carlo sample mean integration as Lebesgue integration with (effectively) Bernoulli estimation of the differential area dA. We note that the sampling procedure preferentially selects y — and hence reduces the A_j estimation error — associated with the larger areas A_j; these larger areas, in turn, will *often* support the more substantial contributions to I^{Leb}. But not always: a function which is vanishingly small over a large region of D and very large over a small region of D will not be efficiently integrated by the (sample mean) Monte Carlo method.

We can also view deterministic methods from this Lebesgue perspective. The important difference between deterministic methods and Monte Carlo methods arises in the approximation of A_j of (13): for deterministic methods, the error in the approximation of A_j will typically converge as $\mathcal{O}(n^{-d})$, hence very slowly for larger dimensions d; in contrast, for Monte Carlo methods, the error in the approximation to A_j (by (17)) will converge as $\mathcal{O}(n^{-1/2})$ — also slowly, but now independent of the spatial dimension, d. Note that n here refers to the number of function evaluations $x \to g(x)$. We conclude that if we wish to break the curse of dimensionality it does not suffice to replace the d-dimensional "Riemann" domain with the one-dimensional "Lebesgue" range; we must also accurately approximate the integrand of the Lebesgue integral.

2.086 Numerical Computation for Mechanical Engineers
Fall 2014