Chapter 6 Mathematical Basis for Function Substitution Monte Carlo

Although we have concentrated on event-based applications, the Monte Carlo method has a rich history of application to various kinds of problems in mathematics, science and engineering, especially those that present themselves with integro-differential equations. Although Monte Carlo methods are sometimes categorized as numerical methods, they do not generally produce estimates of continuous functions themselves, like most other numerical methods, but rather estimates of weighted integrals of these functions.

The main reason for this emphasis on integrals is that the theoretical basis for Monte Carlo methods is the Law of Large Numbers. Since in this law, functions appear inside integrals, the typical applications have similarly been limited to the approximation of integrals of functions (often referred to as "tallies") rather than approximations of functions themselves, which is the most common goal of traditional numerical methods.

But we can move beyond this limitation through the application of the Law of Large Numbers to each point in the continuous domain of a function to be approximated. This approach leads to Monte Carlo algorithms that deliver estimates of complete functions as output, rather than scalar estimates of integrals of these functions; we can, if desired, then insert these approximations into traditional integral tallies, but it is not necessary. Somewhat surprisingly—since this is simply a new way of looking at an old subject—this approach uncovers new possibilities for the method that were always there, but were covered up by the need to reorganize problems into integral form.

6.1 Extension of the Law of Large Numbers to functions

The traditional basis for Monte Carlo processes is the weak form of the Law of Large Numbers.

When this law is applied to a function of one variable, the expected value or mean value of the function, defined by:

$$\bar{f} = E\left[f(\hat{x}) \right] = \int_{-\infty}^{\infty} \pi(\hat{x}) f(\hat{x}) \, d\hat{x} \tag{6-1}$$

where $\pi(\hat{x})$ is a probability distribution function obeying:

$$\begin{aligned}
&\pi(\hat{x}) \geq 0 \text{ for all } \hat{x} \in \Re \\
&\pi(\hat{x}) > 0 \text{ for all } \hat{x} \in \Re \text{ for which } f(\hat{x}) \neq 0 \\
&\int_{-\infty}^{\infty} \pi(\hat{x})\, d\hat{x} = 1
\end{aligned} \tag{6-2}$$

can be represented in probability as:

$$\bar{f} = E\left[f(\hat{x}) \right] = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(\hat{x}_n) \tag{6-3}$$

if the $\hat{x}_n$ are samples taken using $\pi(\hat{x})$ and the variance is finite, i.e.,

$$\sigma^2(\hat{x}) = E\left[\left(f(\hat{x}) - \bar{f} \right)^2 \right] < \infty \tag{6-4}$$

As previously, a caret over a variable will denote a stochastic variable—i.e., a variable that is sampled from a probability density function—and a subscript on a stochastic variable will denote a particular sampled value of that variable.

A Monte Carlo algorithm consists of approximating this mean as closely as desired (i.e., through increasing N) using a finite stream of samples of the function:

$$\bar{f} \approx \frac{1}{N} \sum_{n=1}^{N} f(\hat{x}_n) \tag{6-5}$$

where:

$$\underline{\hat{x}} \equiv \left[\hat{x}_1\ \hat{x}_2\ \hat{x}_3 \ldots \hat{x}_N \right]^T, \tag{6-6}$$

a vector of samples of $\hat{x}$ selected from $\pi(\hat{x})$.

NOTE: From here on, I will generally stop writing the average of summation over the N samples—I will assume you now know what to do with a sample—and just write:

$$\bar{f} \cong f\left(\hat{x}_n \right)$$

That is, the symbol $\cong$ will be used to mean "…has the same expected value as…".

Although this is only an approximation of the desired value, one of the strengths of the Monte Carlo method is that it also delivers an estimate of the variance of $\bar{f}$:

$$\sigma^2(\underline{\hat{x}}) = \lim_{N\to\infty} \frac{1}{N-1}\left[\frac{\sum_{n=1}^{N} f^2(\hat{x}_n)}{N} - \hat{f}^2\right] \tag{6-7}$$

Key point: If we apply the Law of Large Numbers to every point *x* in the domain of a function $f(x)$, the natural extension of Equations 1 and 2 is:

$$f(x) = \int_{-\infty}^{\infty} \pi(\hat{x}) f(x,\hat{x}) d\hat{x} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x,\hat{x}_n) \tag{6-8}$$

with the result being an estimate of the entire function:

$$f(x) \approx \frac{1}{N} \sum_{n=1}^{N} f(x,\hat{x}_n) \tag{6-9}$$

with a corresponding estimate of the variance at each point in the domain:

$$\sigma^2(x,\underline{\hat{x}}) \approx \frac{1}{N-1}\left[\frac{\sum_{n=1}^{N} f^2(x,\hat{x}_n)}{N} - \hat{f}^2(x)\right] \tag{6-10}$$

I will refer to the combination of approximation and PDF in Equation 6-7 as the "pair" $\left(\pi(\hat{x}), f(x,\hat{x})\right)$--because a given $f(x,\hat{x})$ will only work with a particular $\pi(\hat{x})$.

In Equations 6-8, $f(x,\hat{x})$ is a stochastic function whose non-stochastic variable *x* is inherited from the function dependency of $f(x)$ and whose stochastic variable $\hat{x}$ is introduced for our purposes and is sampled using the probability distribution $\pi(\hat{x})$. If we are trying to approximate a function with two variables (e.g., $f(x,y)$ then we would most likely need to have a stochastic variable for each one of the original variables (e.g., $f(x,y,\hat{x},\hat{y})$), with a corresponding two dimensional PDF (e.g., $\pi(\hat{x},\hat{y})$).

But how do we identify such a pair? The approach does not says, but limits itself to the declaration that IF we can find a pair $\left(\pi(\hat{x}), f(x,\hat{x})\right)$ that has the expected value of $f(x)$, then that pair can serve as the basis of a Monte Carlo algorithm, as long as the variance is finite at all

values of *x* at which we need estimates of the function. (How the pair is found is an open question.)

Notice, first of all, that the complexity of the item delivered has increased: Equation 6-4 delivers a scalar estimate of $\bar{f}$ for each sample $\hat{x}_n$, whereas Equation 6-7 delivers a function of *x*, $f(x,\hat{x}_n)$, for each sample.

Notice also that, although there is only a simple notation change between Equations 6-4 and 6-7, we have moved where the known function is located: Equation 6-4 has it inside the integral, on the right hand side; Equation 6-7 has it on the left hand side. This strongly affects the work that we have to do to use it: Equation 6-4 clearly shows us the pair $(f(\hat{x}),\pi(\hat{x}))$ to be used in the Monte Carlo algorithm, but Equation 6-7 provides no guidance about how an appropriate pair $(f(x,\hat{x}),\pi(\hat{x}))$ is to be found, leaving us with the task of finding one. It simply assures us that, if a pair can be found that satisfies the "continuous side" of Equation 6-6 (i.e., the first equality), then that same pair can be used as the basis of a Monte Carlo estimate of $f(x)$ by averaging multiple functional samples for all values of x for which the variance is finite. The trick to designing a Monte Carlo function approximation using Equation 6-7 is to find an appropriate pair to base it on.

6.2 Random functions using the Dirac delta

Our first stochastic approximation is based on the Dirac delta. In terms of our mixed stochastic function, $f(x,\hat{x})$, we note that:

$$f(x) \cong f(x,\hat{x}) \equiv \frac{f(\hat{x})}{\pi(\hat{x})}\delta(x-\hat{x}) \tag{6-11}$$

can easily be shown to satisfy the first equality of Equation 6, i.e.,:

$$f(x) = \int_{-\infty}^{\infty} \pi(\hat{x}) f(x,\hat{x}) d\hat{x} = \int_{-\infty}^{\infty} \pi(\hat{x}) \left[\frac{f(\hat{x})}{\pi(\hat{x})}\delta(x-\hat{x}) \right] d\hat{x} = \int_{-\infty}^{\infty} f(\hat{x})\delta(x-\hat{x}) d\hat{x} = f(x)$$

as long as the probability distribution $\pi(\hat{x})$avoids singularities by having $\pi(\hat{x}) \neq 0$ at all points for which $f(\hat{x}) \neq 0$. Therefore, according to the rules we established earlier, we can use it as the basis of a Monte Carlo algorithm. But, before we do this, let's take a look at what this approximation looks like.

Example. Figure 4 shows the resulting approximation of the function $f(x) = e^{-2x} - e^{-10x}$, $0 < x < 1$ ("sample function") for 100 random values of $\hat{x}$ chosen uniformly from 0 to 1. As can be seen in the figure, the heights of the

Dirac deltas follow the shape of the curve because of the uniform distribution in x (although they are a factor of 100 smaller).

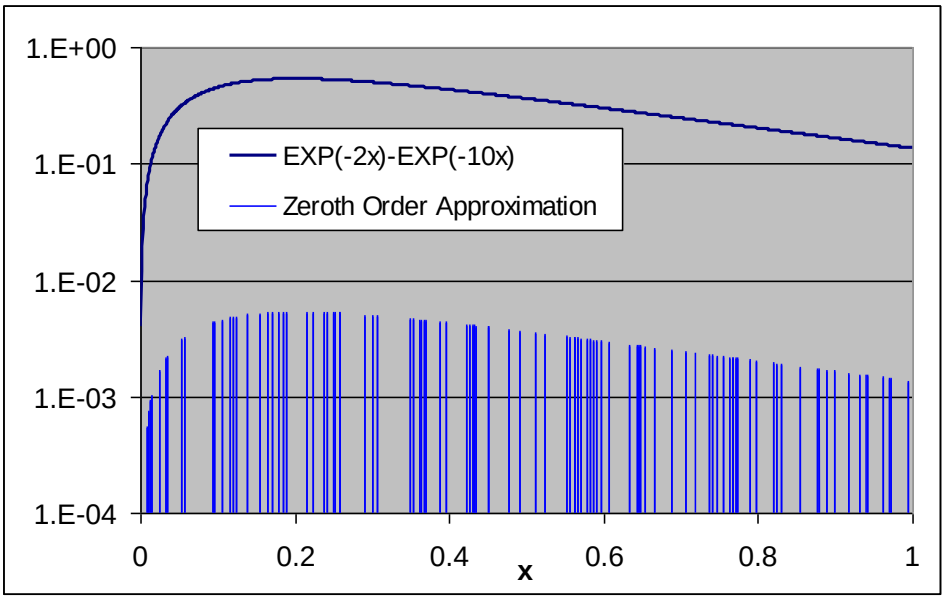

Figure 4. Dirac delta approximation of sample function ($\pi(x)$=1)

So the answer to the question "What does it look like?" is "Terrible." We expect approximations of functions (e.g., least squares fits) to look SOMETHING like the functions they are approximating!

This "approximation" cannot even satisfy the simplest requirement of a function—that it delivers a specific function value for any value of the argument within the domain. Because it contains the Dirac delta, the stochastic function given by Equation 9 can only be practically used to approximate $f(x)$ inside integrals. (As an aside, this is the approximation of $f(x)$ that is implicit in most existing Monte Carlo algorithms, which also serves to explain why the functions have to occur inside integrals.)

Then why do we do it? Because in its limited application (integration), evaluating the integral of the approximation can be thousands or millions or billions of times faster than evaluating the integral of the function itself. (If this is not true—then the thousands or millions or billions of samples that it takes to get a good approximation to the integral would not be worth it.)

In practice, the tradeoff is NOT worth it for easy integrals, but becomes practical only for either multidimensional integrations or for integrations where the integral is especially difficult—e.g., resonance cross section reaction rates.

Higher order Random Functions (TBD)

Bottom Line of This Section

The purpose of this section has been to add a "middle step" to the Law of Large Numbers that corresponds to the usual beginning point of other substitutions: Representing a functional approximation that is then substituted for the function itself in integrals, equations, etc.

But <u>why</u> do it this way? Because it allows us two advantages:

1. It gives us a function substitution methodology that we can use to develop Monte Carlo algorithms from given integrals, equations, etc.
2. By generalizing with Equation 6 (i.e., beyond the traditional Dirac delta approximation) we will be able to develop more efficient Monte Carlo algorithms or tallies. (Track length flux estimation estimation is one of these.)

For the next little while (i.e., while we are developing algorithms for solving equations), we will be sticking with the Dirac delta approximation given in Equation 6-10, and apply it to an increasingly complicated set of integral equations.

6.3 Solving integral equations

First we will apply what we have learned to the integral of know functions.

For the remaining cases in this chapter (and, likely, the majority of cases you will encounter in the real world), we can apply Equation 6-10 to the simplest function we know of—a function that is equal to 1 across its entire domain. This gives us:

$$f(x) = 1 \cong \frac{\delta(x - \hat{x})}{\pi(\hat{x})} \tag{6-12}$$

This will allow us to convert continuous variables into sampled variables in a large number of useful cases, simply by multiplying by one.

An additional identity that we will use is:

$$f(x)\delta(x - \hat{x}) = f(\hat{x})\delta(x - \hat{x}) \tag{6-13}$$

What this allows us to do is to replace any occurrence of the variable x in a term containing a Dirac delta of the form of the above equation with the selected value $\hat{x}$, leaving the only

occurrence of the continuous variable x inside the Dirac delta itself (which we be subsequently dealt with by an integral). For example, if we take:

$$f(x) = xe^{2x} \tag{6-14}$$

and convert the continuous variable to a sampled variable by multiplying by 1 (in the form of Equation 6-11), we get;

$$f(x) \cong xe^{2x}\left[\frac{\delta(x-\hat{x})}{\pi(\hat{x})}\right] = \frac{\hat{x}e^{2\hat{x}}}{\pi(\hat{x})}\delta(x-\hat{x}) \tag{6-15}$$

A final notational change is introduced (to simplify the notation) to deal with the fact that we will often encounter "left over" integrals of the sort:

$$xxx \tag{6-16}$$

$$\int_a^b \delta(x-\hat{x})\,d\hat{x}$$

You will remember that it is defined to be equation to:

$$xxx \tag{6-17}$$

$$\int_a^b \delta(x-\hat{x})\,d\hat{x} \equiv \begin{cases} 1, & \text{if } a < \hat{x} < b \\ -1, & \text{if } b < \hat{x} < a \\ 0, & \text{otherwise} \end{cases} \equiv \Delta(a,\hat{x},b)$$

We will define that last bit, $\Delta(a,\hat{x},b)$, as a "convenience function" to save keystrokes. This can be thought of a 1 if the order of the arguments is correction and 0 otherwise (we will not use the branch that comes to -1).

In practice, we will try to eliminate the $\Delta(a,\hat{x},b)$ terms altogether by picking the domain of the PDF $\pi(\hat{x})$ to be restricted to $\hat{x}\in(a,b)$, allowing us to eliminate $\Delta(a,\hat{x}_n,b)$ from the terms it is in.

Our basic procedure will be to design a sampling algorithm for a continuous variable using three steps:

1. In terms involving the variable (e.g., *x*) multiply the term by $\frac{\delta(x-\hat{x}_n)}{\pi(\hat{x}_n)}$ (legal because this has an expected value of one).
2. For all occurrences of the variable *x* not in the Dirac delta that we just added (but maybe in OTHER Dirac deltas previously introduced), replace *x* with $\hat{x}_n$.
3. If, after simplifying the integrals, we are left with simple integrals of Dirac deltas, replace with our convenience function $\Delta(a,\hat{x}_n,b)$ --where *a* and *b* are the upper and lower limits of the integral.

Examples will make this clearer as we go on.

Example 1: Reproduce the scalar form of the Law of Large Numbers using our new notation.

Answer: We will go through this step by step. The beginning of the LLN is:

$$\bar{f} = E\left[f(x) \right] = \int_{-\infty}^{\infty} \pi(x) f(x)\, dx \tag{6-18}$$

The only continuous variable is the dummy variable of integration. For Step 1 we insert the full version of Equation 6-11:

$$1 = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \frac{\delta\left(x - \hat{x}_n\right)}{\pi(\hat{x}_n)} \tag{6-19}$$

to get:

$$\begin{aligned} \bar{f} &= \int_{-\infty}^{\infty} \pi(x) f(x) \left[\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \frac{\delta\left(x - \hat{x}_n\right)}{\pi(\hat{x}_n)} \right] dx \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_{-\infty}^{\infty} \pi(x) f(x) \frac{\delta\left(x - \hat{x}_n\right)}{\pi(\hat{x}_n)}\, dx \end{aligned} \tag{6-20}$$

For Step 2, we replace the x values with $\hat{x}_n$ to get:

$$\bar{f} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_{-\infty}^{\infty} \pi(\hat{x}_n) f(\hat{x}_n) \frac{\delta\left(x - \hat{x}_n\right)}{\pi(\hat{x}_n)}\, dx \tag{6-21}$$

which simplifies to:

$$\bar{f} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(\hat{x}_n) \int_{-\infty}^{\infty} \delta\left(x - \hat{x}_n\right) dx \tag{6-22}$$

Step 3 would, formally, replace the $\int_{-\infty}^{\infty} \delta\left(x - \hat{x}\right) dx$ with $\Delta\left(-\infty, \hat{x}, \infty\right)$, but since this is obviously equal to 1, we can write our final form:

$$\bar{f} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(\hat{x}_n) \tag{6-23}$$

This is the right-hand-side of the LLN, which completes the example.

Example 2: For our second example, we reproduce the formula for the averaging method from Chapter 5 using our methodology.

Answer: In this method, we set out to solve:

$$I = \int_{a}^{b} f(x)\, dx \tag{6-24}$$

by sampling *x* uniformly in the domain *(a,b)* and then scoring $f(\hat{x}_n)(b-a)$.

To show that this sampling algorithm follows from our procedure, we quickly run through Steps 1 and 3 to get:

$$
\begin{aligned}
I &= \int_a^b f(x)\,dx \\
&\cong \int_a^b f(x)\left[\frac{\delta\left(x-\hat{x}_n\right)}{\pi\left(\hat{x}_n\right)}\right]dx \\
&= \int_a^b f(\hat{x}_n)\left[\frac{\delta\left(x-\hat{x}_n\right)}{\pi\left(\hat{x}_n\right)}\right]dx \\
&= \frac{f(\hat{x}_n)}{\pi\left(\hat{x}_n\right)}\int_a^b \delta\left(x-\hat{x}_n\right)dx \\
&= \frac{f(\hat{x}_n)}{\pi\left(\hat{x}_n\right)}\Delta\left(a,\hat{x}_n,b\right)
\end{aligned}
\tag{6-25}
$$

Since the procedure for the averaging method involves uniformly sampling x between a and b, then we must have:

$$\pi\left(\hat{x}_n\right) = \frac{1}{b-a} \tag{6-26}$$

and

$$\Delta\left(a,\hat{x}_n,b\right) = 1 \tag{6-27}$$

Substituting these two give us a score of:

$$I \cong f(\hat{x}_n)\left(b-a\right) \tag{6-28}$$

which corresponds to the averaging method score.

Example 3: For our third simple example, we "finish" Example 2 to reproduce the formula for the importance sampling method from Chapter 5 using our methodology.

Answer: In this method, we set out to solve:

$$I = \int_a^b f(x)\,dx \tag{6-29}$$

by sampling x using a guess, $h(x)$, for the unnormalized PDF.

Picking up at the point in the previous example where we had the score as:

$$I \cong \frac{f(\hat{x}_n)}{\pi(\hat{x}_n)} \Delta(a, \hat{x}_n, b) \tag{6-30}$$

Since the normalized PDF would be:

$$\pi(\hat{x}_n) = \frac{h(\hat{x}_n)}{\int_a^b h(x)\,dx} = \frac{h(\hat{x}_n)}{I_h} \tag{6-31}$$

and we would again sample x between a and b, therefore

$$\Delta(a, \hat{x}_n, b) = 1 \tag{6-32}$$

then substitution would give us:

$$I \cong \frac{f(\hat{x}_n)}{h(\hat{x}_n)} I_h \tag{6-33}$$

which corresponds to the importance sampling method score.

Now that we have caught up to the event-based methods, let's go after an integral equation.

Example 4: Develop an algorithm to sample the solution of the integral equation:

$$f(x) = \int_0^x u^2\, du + 4 \tag{6-34}$$

Answer: This problem has two continuous variables, *x* and *u*, rather than the one in the previous three examples. We can get after them in whatever order we prefer. For this example, we will go after *u* first, by following our rules to get:

$$\begin{aligned} f(x) &= \int_0^x u^2\, du + 4 \\ &\cong \int_0^x u^2 \left[\frac{\delta(u-\hat{u}_n)}{\pi(\hat{u}_n)} \right] du + 4 \\ &= \frac{\hat{u}_n^{\,2}}{\pi(\hat{u}_n)} \int_0^x \delta(u-\hat{u}_n)\, du + 4 \\ &= \frac{\hat{u}_n^{\,2}}{\pi(\hat{u}_n)} \Delta(0,\hat{u}_n,x) + 4 \end{aligned} \tag{6-35}$$

Next we deal with the continuous variable *x* by again following our procedure:

$$\begin{aligned} f(x) &\cong f(x) \left[\frac{\delta(x-\hat{x}_n)}{\pi(\hat{x}_n)} \right] \\ &= \frac{f(\hat{x}_n)}{\pi(\hat{x}_n)} \delta(x-\hat{x}_n) \\ &= \frac{\dfrac{\hat{u}_n^{\,2}}{\pi(\hat{u}_n)} \Delta(0,\hat{u}_n,\hat{x}_n) + 4}{\pi(\hat{x}_n)} \delta(x-\hat{x}_n) \end{aligned} \tag{6-36}$$

The resulting formal solution is:

$$f(x) \cong w_n \delta(x - \hat{x}_n) \tag{6-37}$$

where

$$w_n = \frac{\frac{\hat{u}_n^{\,2}}{\pi(\hat{u}_n)} \Delta(0, \hat{u}_n, \hat{x}_n) + 4}{\pi(\hat{x}_n)} \tag{6-38}$$

The preceding example results in an algorithm that we will refer to as a "general" algorithm, since it still has the PDF(s) in the solution weights. Before it can be implemented in a solution, these PDFs must be chosen; how they are chosen depends on how the samples will be used.

Using the preceding example results, if we were planning to integrate f(x) from 9 to 10, then the result of this integration would be:

$$\begin{aligned} I &= \int_9^{10} f(x)\, dx \\ &\cong w_n \int_9^{10} \delta(x - \hat{x}_n)\, dx \\ &= w_n \Delta(9, \hat{x}_n, 10) \\ &= \frac{\frac{\hat{u}_n^{\,2}}{\pi(\hat{u}_n)} \Delta(0, \hat{u}_n, \hat{x}_n) + 4}{\pi(\hat{x}_n)} \Delta(9, \hat{x}_n, 10) \end{aligned} \tag{6-39}$$

A logical (but maybe not the most efficient) way to proceed would be to sample x uniformly between 9 and 10, and sample u uniformly between 0 and the chosen value of x. Then the deltas would disappear and the PDFs would be:

$$\pi\left(\hat{x}_n\right)=1$$
$$\pi\left(\hat{u}_n\right)=\frac{1}{\hat{x}_n} \tag{6-40}$$

This would result in a score of:

$$I \cong \hat{x}_n \hat{u}_n^{\,2}+4 \tag{6-41}$$

On the other-hand, if we were just interested in the value of f(10), there would be no reason to really sample x at all. We always "sample" x at the value 10, and sample u uniformly between 0 and 10 to get:

$$\pi\left(\hat{x}_n\right)=\delta\left(x-10\right)$$
$$\pi\left(\hat{u}_n\right)=\frac{1}{10} \tag{6-42}$$

This would result in a score of:

$$f\left(10\right) \cong 10\hat{u}_n^{\,2}+4 \tag{6-43}$$

Developing integral equations from differential equations

We really don't encounter integral equations as often as differential equations.

But, we can often convert a differential equation into an integral equation. This is best shown with a couple of examples.

Example 5: Sample the differential equation and boundary condition:

$$\frac{df(x)}{dx} = 2e^x, \quad f(0) = 4 \tag{6-44}$$

Answer: The basic technique is to integrate the differential equation FROM the point where the boundary condition is known TO the desired values of x. So, if we integrate the equation from 0 to some value x, we get:

$$\begin{aligned} \int_0^x \frac{df(u)}{du}\,du &= \int_0^x 2e^u\,du \\ f(x) - f(0) &= 2\int_0^x e^u\,du \\ f(x) &= f(0) + 2\int_0^x e^u\,du \\ f(x) &= 4 + 2\int_0^x e^u\,du \end{aligned} \tag{6-45}$$

If we follow through like the previous example by working with x first and then u (for a slightly different approach resulting in the same answer), our development chain would look like this:

$$f(x) = 4 + 2\int_0^x e^u \, du$$

$$\Rightarrow f(x)\frac{\delta(x-\hat{x}_n)}{\pi(\hat{x}_n)}$$

$$= \frac{f(\hat{x}_n)}{\pi(\hat{x}_n)}\delta(x-\hat{x}_n)$$

$$= \frac{4 + 2\int_0^{\hat{x}_n} e^u \, du}{\pi(\hat{x}_n)}\delta(x-\hat{x}_n) \qquad (6\text{-}46)$$

$$= \frac{4 + 2\int_0^{\hat{x}_n} e^u \frac{\delta(u-\hat{u}_n)}{\pi(\hat{u}_n)} du}{\pi(\hat{x}_n)}\delta(x-\hat{x}_n)$$

$$= \frac{4 + 2\frac{e^{\hat{u}_n}}{\pi(\hat{u}_n)}\Delta(0,\hat{u}_n,\hat{x}_n)}{\pi(\hat{x}_n)}\delta(x-\hat{x}_n)$$

6.4 Solving linked sets of integral equations

There is nothing particularly new in this module except for practice and growing confidence in your ability to work longer problems, where one Monte Carlo approximation feeds into another one.

That is, if you have a string of linked integral equations (or can get them by integrating differential equations):

$$\begin{aligned} a(u) &= \int_{-\infty}^{u} f(t)\,dt \\ b(v) &= \int_{-\infty}^{v} g(u')\,a(u')\,du' \end{aligned} \tag{6-47}$$

you attack it in pieces that "cascade" into each other:

$$\begin{aligned} a(u) &= \int_{-\infty}^{u} f(t)\,dt \\ &\cong a(u)\frac{\delta(u-\hat{u}_n)}{\pi(\hat{u}_n)} \\ &= \frac{a(\hat{u}_n)}{\pi(\hat{u}_n)}\delta(u-\hat{u}_n) \\ &= \frac{\int_{-\infty}^{\hat{u}_n} f(t)\,dt}{\pi(\hat{u}_n)}\delta(u-\hat{u}_n) \\ &= \frac{\int_{-\infty}^{\hat{u}_n} f(t)\frac{\delta(t-\hat{t}_n)}{\pi(\hat{t}_n)}\,dt}{\pi(\hat{u}_n)}\delta(u-\hat{u}_n) \\ &= \frac{\frac{f(\hat{t}_n)}{\pi(\hat{t}_n)}\Delta(-\infty,\hat{t}_n,\hat{u}_n)}{\pi(\hat{u}_n)}\delta(u-\hat{u}_n) \end{aligned} \tag{6-48}$$

And then:

$$
\begin{aligned}
b(v) &= \int_{-\infty}^{v} g(u')a(u')du' \\
&\cong \frac{b(\hat{v}_n)}{\pi(\hat{v}_n)}\delta(v-\hat{v}_n) \\
&= \frac{\int_{-\infty}^{\hat{v}_n} g(u')a(u')du'}{\pi(\hat{v}_n)}\delta(v-\hat{v}_n)
\end{aligned}
\tag{6-49}
$$

At this point, we might be tempted to sample the u' variable, but we don't because we already have an approximation for a(u') that has a Dirac delta in it! So we just substitute it to get:

$$
\begin{aligned}
b(v) &\cong \frac{\int_{-\infty}^{\hat{v}_n} g(u')a(u')du'}{\pi(\hat{v}_n)}\delta(v-\hat{v}_n) \\
&= \frac{\int_{-\infty}^{\hat{v}_n} g(u')\left[\frac{f(\hat{t}_n)}{\pi(\hat{t}_n)}\Delta(-\infty,\hat{t}_n,\hat{u}_n)\frac{}{\pi(\hat{u}_n)}\delta(u'-\hat{u}_n)\right]du'}{\pi(\hat{v}_n)}\delta(v-\hat{v}_n) \\
&= \frac{\frac{g(\hat{u}_n)\frac{f(\hat{t}_n)}{\pi(\hat{t}_n)}\Delta(-\infty,\hat{t}_n,\hat{u}_n)}{\pi(\hat{u}_n)}\Delta(-\infty,\hat{u}_n,\hat{v}_n)}{\pi(\hat{v}_n)}\delta(v-\hat{v}_n) \\
&= \frac{g(\hat{u}_n)\,f(\hat{t}_n)\Delta(-\infty,\hat{t}_n,\hat{u}_n)\Delta(-\infty,\hat{u}_n,\hat{v}_n)}{\pi(\hat{t}_n)\pi(\hat{u}_n)\pi(\hat{v}_n)}\delta(v-\hat{v}_n)
\end{aligned}
\tag{6-50}
$$

(Yes, I admit that the algebra is no fun. But the resulting algorithm is interesting to code and see the results converge to the right answer.)

Besides problems that physically cascade (like radioactive decay chains), there are two other common situations that generate cascades: higher-order differential equations and recurrence

integral equations (where the function on the left-hand side appears on the right as well). The second of these we will deal with in the next section.

Higher-order differential equations can be made to cascade when the linked equation set comes from reducing the nth order equation to n first-order equations (with the n boundary conditions dispersed among the resulting linked equations).

Example: Create a Monte Carlo algorithm to solve:

$$y'''(x) = f(x)\ ;\ y(0)=1, y'(0)=2; y''(0)=3 \tag{6-51}$$

Answer: We define:

$$\begin{aligned} w(x) &= f'(x) \\ u(x) &= w'(x) = f''(x) \end{aligned} \tag{6-52}$$

Use of these functions changes the original equation into:

$$\begin{aligned} u'(x) &= f(x)\ ;\ u(0)=3 \\ v'(x) &= u(x)\ ;\ v(0)=2 \\ y'(x) &= v(x)\ ;\ y(0)=1 \end{aligned} \tag{6-53}$$

Integrating each of these first-order differential equations gives us:

$$\begin{aligned} u(x) &= 3+\int_0^x f(x')\,dx' \\ v(x) &= 2+\int_0^x u(x')\,dx' \\ y(x) &= 1+\int_0^x v(x')\,dx' \end{aligned} \tag{6-54}$$

We use our methodology to sample the *u(x)* term, substituting this sample into the integral of the *v(x)* term, and then substitute THIS sample into the y(x) term. The algebra is tedious, but doable. The result is:

$$
\begin{aligned}
f(x) &\cong w_f \delta(x - \hat{x}_f);\ w_f \equiv \frac{f(\hat{x}_f)}{\pi(\hat{x}_f)} \\
u(x) &\cong w_u \delta(x - \hat{x}_u);\ w_u \equiv \frac{3 + w_f \Delta(0, \hat{x}_f, \hat{x}_u)}{\pi(\hat{x}_u)} \\
v(x) &\cong w_v \delta(x - \hat{x}_v);\ w_v \equiv \frac{2 + w_u \Delta(0, \hat{x}_u, \hat{x}_v)}{\pi(\hat{x}_v)} \\
y(x) &\cong w \delta(x - \hat{x});\ w \equiv \frac{1 + w_v \Delta(0, \hat{x}_v, \hat{x})}{\pi(\hat{x})}
\end{aligned}
\tag{6-55}
$$

I could (if I wanted to) substitute all these subscripted weights to get the (final) w weight in terms of $f(x)$ and all the PDFs, but I won't—since I usually code all the intermediate weights just like the above equations (to minimize the likely algebraic mistakes I would make combining them).

Of course, like we said before, this algorithm leaves you with $y(x)$ in terms of Dirac deltas—which are only useful if put into integrals—, so you must ultimately be interested in some integral of $y(x)$ to make this a useful exercise. But, I will leave that part out of this Example.

6.5 Neumann decomposition

Sampling from recurring equations introduces a complexity. We cannot use the above procedure because, if we try to follow it we will find ourselves unable to sample the occurrences of the function on the right-hand side of the equation. That is, the procedure requires that we sample from $f(x)$ (i.e., on the right-hand side) in order to sample from $f(x)$ (i.e., on the left-hand side).

However, all is not lost. For **linear** occurrences of $f(x)$ on the right-hand side, we can "bootstrap" a solution by representing $f(x)$ as an infinite Neumann series:

$$f(x) = f_0(x) + f_1(x) + f_2(x) + \ldots, \tag{6-56}$$

substituting this series for $f(x)$ on BOTH sides of the equation, dividing the resulting equation into a infinite set of coupled equations for the $f_i(x)$, and then sampling -- in turn -- the $f_i(x)$. TOGETHER the samples for the individual $f_i(x)$ would combine to form a single sample for $f(x)$.

Of course, this procedure has an infinite number of steps for each sample of $f(x)$, so it will have to be adjusted, but—before worrying about that—let us first look at an example of how the procedure so far would shape up.

Example: Develop an infinite sampling procedure for the recurring equation:

$$\frac{df(x)}{dx} = f(x),\ f(0) = 1,\ 0 < x < 3 \tag{6-57}$$

Answer: Integrating the differential equation over x from 0 to x (and applying the boundary condition) gives us the recurring integral equation:

$$f(x) = 1 + \int_0^x f(u)\,du,\ 0 < x < 3 \tag{6-58}$$

If we insert the infinite Neumann series for the function on both sides, we get:

$$f_0(x) + f_1(x) + \ldots = 1 + \int_0^x f_0(u)\,du + \int_0^x f_1(u)\,du + \ldots \tag{6-59}$$

This can be decomposed into the following coupled equations:

$$f_0(x) = 1 \tag{6-60}$$

$$f_1(x) = \int_0^x f_0(u)\,du$$

…

$$f_n(x) = \int_0^x f_{n-1}(u)\,du$$

...

Since the function $f(x)$ is the infinite sum of these $f_n(x)$, the procedure to sample from $f(x)$ is:

3. Sample from $f_0(x)$ by:
   a. Choosing an $\hat{x}_{i0}$ between 0 and 3. This involves using the normalized probability distribution $\pi_0(x) = 1/3$.
   2. Our sample of $f_0(x)$ is found from:

$$f_0(x) \cong w_{i0}\delta(x - \hat{x}_{i0}) \tag{6-61}$$

with:

$$w_{i0} = \frac{f_0(\hat{x}_{i0})}{\pi(\hat{x}_{i0})} = \frac{1}{1/3} = 3 \tag{6-62}$$

4. Sample from $f_1(x)$ using the sample of $f_0(x)$ by:
   1. Choosing an $\hat{x}_{i1}$ between 0 and 3. Again, the normalized distribution is 1/3.
   2. Setting $\hat{u} = \hat{x}_{i0}$
   3. Our sample of $f_1(x)$ is found from:

$$
\begin{aligned}
&f_1(x) \cong w_{i1}\delta(x-\hat{x}_{i1}) \\
&w_{i1} = \frac{f_1(\hat{x}_{i1})}{\pi(\hat{x}_{i1})} = \frac{\int_0^{\hat{x}_{i1}} f_0(u)\,du}{1/3} \cong 3\int_0^{\hat{x}_{i1}} \left[w_{i0}\delta(x-\hat{x}_{i0})\right] du \\
&\quad = 3w_{i0}\int_0^{\hat{x}_{i1}} \left[\delta(x-\hat{x}_{i0})\right] f_0(u)\,du \\
&\quad = \begin{cases} 3w_{i0} & \text{if } 0 < \hat{x}_{i0} < \hat{x}_{i1} \\ 0 & \text{Otherwise} \end{cases}
\end{aligned}
\tag{6-63}
$$

`

5. Sample from $f_n(x)$ in the analogous manner using the sample of $f_{n-1}(x)$:

$$
\begin{aligned}
&f_{in}(x) \cong w_{in}\delta(x-\hat{x}_{in}) \\
&w_{in} = \begin{cases} 3w_{i,n-1} & \text{if } 0 < \hat{x}_{i,n-1} < \hat{x}_{in} \\ 0 & \text{Otherwise} \end{cases}
\end{aligned}
\tag{6-64}
$$

Now that we have developed the "infinite procedure", let us make some observations.

6. The IF TEST in step 2.C is necessary because if the Dirac delta sample for $f_0(x) \cong w_{i0}\delta(x-\hat{x}_{i0})$ lies outside the range of the integral, the integral -- and therefore our sample of $f_1(x)$—goes to zero because of the definition of Dirac delta integration.
7. The procedure is infinite in theory, but not infinite in practice because as soon as we pick a value of $\hat{x}_{in}$ that is SMALLER than the one before it, then w_{in} will go to zero. Once this happens, of course, we can ignore the rest of the Neumann steps because their weights will be zero as well.
8. We must remember that it is not a single sample of $f_0(x)$ or $f_1(x)$, etc., that constitutes our sample of the function $f(x)$, but ALL OF THEM together. Therefore, the i'th sample of $f(x)$ is, formally:

$$
f(x) \cong \sum_{n=0}^{\infty} w_{in}\delta(x-\hat{x}_{in}) \tag{6-65}
$$

9. Therefore, if we improve our approximation of $f(x)$ by taking *N* samples, the combined best result would be:

$$f(x) \cong \frac{1}{N}\sum_{i=1}^{N}\sum_{n=0}^{\infty} w_{in}\delta\left(x-\hat{x}_{in}\right) \tag{6-66}$$

As a practical matter, point 3 means that our coding must collect data in "sample bins" -- i.e, which collect data from individual Neumann terms within a single sample -- and, at the end of the sample, contribute from the "sample bins" to the overall "solution bins".

Chapter 6 Exercises

6-1. Write and run a code to sample the differential equation and boundary condition:

$$\frac{df(x)}{dx} = \frac{2}{x}, \quad f(2) = 1$$

Use your code to find:
a. The value of the function at x=4; and
b. The integral of the function from 2 to 4.

6-2. Solve for $\int_1^2 y(x)dx$, given the equation:

$$y''(x) = x^2 \; ; \; y(0) = 3, y'(0) = 0$$

6-3. Assume that at t=0, you have a single atom of species A. A decays to B with a half-life of 1 second and B decays to C with a half-life of 0.5 seconds. (C is stable.) Estimate the expected number of atoms of C at t=3 seconds?

Hint: In case you are rusty with the equations of decay, the integral equations you should start with are:

$$A(t) = e^{-\lambda_A t}$$

$$B(t) = \int_0^t \lambda_A A(t') e^{-\lambda_B (t-t')} dt'$$

$$C(t) = \int_0^t \lambda_B B(t') dt'$$

6-4. Use the drunken sailor algorithm to solve a diffusion problem. On a 5 cm x 5 cm grid with boundary conditions 1 (left edge), 2 (bottom), 3 (right), and 4 (top), use Monte Carlo (on a 0.5 cm grid) to estimate the value of the function at the point (3.5, 1.5).

Answers to selected exercises

Chapter 6

6-1. a. $2.38 \pm \frac{.28}{\sqrt{N}}$

b. $3.55 \pm \frac{1.59}{\sqrt{N}}$

6-2. $3.52 \pm \frac{3.93}{\sqrt{N}}$

6-3. 0.77 (Standard deviation depends on PDF used)

6-4. $2.5 \pm \frac{0.801}{\sqrt{N}}$