Chapter 4 Sampling of Integrals

4.1 Four techniques for estimating integrals

Our next set of mathematical tools that we will develop involve Monte Carlo integration. In the grand scheme of things, our study so far of sampling from distributions has provided us with the tools we need to **run** a simple Monte Carlo simulation; what the study of integration of functions will give us relates to **keeping score** in a simulation. In addition, we will use this lesson to introduce – in the simplest context I can think of – the first techniques of variance reduction.

This lesson is divided into 2 main parts:

1. Our basic mathematical framework for performing the integration; and
2. Development of four particular methods using the framework.

Mathematical framework for integration

We will be considering the use of Monte Carlo methods to sample the value of a basic integral of the form:

$$I = \int_a^b f(x)\,dx \qquad (4\text{-}1)$$

We are going to use this as a very simple example of a complete Monte Carlo simulation. (There are much better ways of solving this integral.)

Our basic mathematical framework is a two step procedure whereby we:

- Sample x (somehow) between a and b, giving us an $\hat{x}_i$; and then
- Score a guess for the integral solution, I_i, based on that $\hat{x}_i$.

Different methods for solving this integral are possible and can be characterized by how they perform these two tasks.

All of them are based on a (more or less) direct application of the Law of Large Numbers, which can be over-simplified as:

$$\overline{f} = \int_a^b f(x)\pi(x)\,dx = \lim_{N\to\infty} \frac{\sum_{i=1}^{N} f(\hat{x}_i)}{N} \qquad (4\text{-}2)$$

where the samples $\hat{x}_i$ are chosen using the PDF $\pi(x)$.

Note particularly the fact that there are TWO functions in the integrand (actually one function and one distribution) and the two have different ROLES in the subsequent Monte Carlo approximation:

3. The $\pi(x)$ is used as the basis for choosing the $\hat{x}_i$ value; and
4. The $f(x)$ is used (evaluated at the chosen value) for the score.

The different methods we will study involve taking what is inside the integral and "breaking it up" different ways to satisfy these two roles.

In particular, the integral that I presented at the beginning of the lesson does not HAVE a PDF in it; so, you have to supply one. If we denote a normalized probability distribution between a and b as $\pi(x)$, we can write the integral as:

$$I = \int_a^b \left[\frac{f(x)}{\pi(x)} \right] \pi(x)\, dx \tag{4-3}$$

From this framework, we can "flesh out" our two step approach into:

- Sample x using π(x) between a and b (which I designate using the notation $I = \int_a^b \ldots \pi(x)\, dx$) ; and then
- Score a guess for the integral solutions as $I_i = \dfrac{f(\hat{x}_i)}{\pi(\hat{x}_i)}$ (or as some probability mixing strategy that has this expected value).

Four particular methods for performing the integration

We will now go over four particular variations on this theme, each of them characterized mathematically by particular choices of *π(x)* and the method of scoring:

- Rejection method
- Averaging method
- Control variates method
- Importance sampling method

Rejection method

This is a similar approach to the use of rejection methods in picking from a distribution. It is a "dart board" method in which we estimate the area under a functional curve by containing the curve in a rectangular "box", picking a point randomly in the box, and scoring 0 if it misses (i.e., is above the curve) or the full rectangular area if it hits (i.e., is below the curve).

As before, we have to specify an upper bound of the function, $f_{\max}$, and then proceed by:

5. Choosing a value of $\hat{x}_i$ uniformly between a and b.
6. Choosing a value of $\hat{f}_i$ uniformly between 0 and $f_{\max}$.
7. Scoring $I_i = f_{\max}(b-a)$ if $\hat{f}_i < f(\hat{x}_i)$ and scoring $I_i = 0$ otherwise.

In terms of our mathematical framework, this is equivalent to using:

$$\pi(x) = \frac{1}{b-a} \tag{4-4}$$

(for a uniform distribution between a and b) and scoring with a probability mixing strategy of scoring:

$$\frac{f_{\max}}{\pi(\hat{x}_i)} \text{ with probability } \frac{f(\hat{x}_i)}{f_{\max}} \tag{4-5}$$

or scoring 0 with probability $1-\frac{f(\hat{x}_i)}{f_{\max}}$.

This mixed scoring strategy obviously has the desired expected value of $I_i = \frac{f(\hat{x}_i)}{\pi(\hat{x}_i)}$.

Example 1: Find $I = \int_0^2 x^2 dx$ using a rejection method. (The answer is 8/3.)

Answer: The maximum value of this function in the domain is 4, so our procedure is to:

8. Choose a value of $\hat{x}_i$ uniformly between 0 and 2.
9. Choose a value of $\hat{f}_i$ uniformly between 0 and 4.
10. Score 8 if $\hat{f}_i$ is less than $f(\hat{x}_i)$; otherwise score 0.

A FORTRAN computer program was written to solve this problem. The results are:

Number of histories	$\hat{I}$	S_I
10,000	2.7080	0.03786
100,000	2.6571	0.01192

1,000,000	2.6628	0.00377

Averaging method

This is a much more straight-forward approach to the problem because it uses the function $f(x)$ directly. The procedure for this method is to:

11. Choose a value of $\hat{x}_i$ uniformly between a and b.
12. Score $I_i = f(\hat{x}_i)\cdot(b-a)$.

In terms of our mathematical framework, this is equivalent to again using:

$$\pi(x) = \frac{1}{b-a} \tag{4-6}$$

(for a uniform distribution between a and b) and scoring with a direct use of

$$I_i = \frac{f(\hat{x}_i)}{\pi(\hat{x}_i)} \tag{4-7}$$

Example 2: Again find $I = \int_0^2 x^2 dx$ using an averaging method.

Answer: The procedure is to:

13. Choose a value of $\hat{x}_i$ uniformly between 0 and 2.
14. Score $2\hat{x}_i^2$.

A FORTRAN computer program was written to solve this problem. The results are:

Number of histories	$\hat{I}$	S_I

10,000	2.6747	0.02393
100,000	2.6615	0.00753
1,000,000	2.6629	0.00238

The estimated standard deviations are less than for the previous rejection example.

Control variates method

This method is the first of two methods that utilize a user-supplied second function, $h(x)$, which is chosen to be a "well behaved" approximation to $f(x)$. What makes these methods so powerful is that they allow the user to take use of *a priori* knowledge about the function $f(x)$. In the control variates method, the integral solution "begins" as the integral of the known function:

$$I_h = \int_a^b h(x)\,dx \tag{4-8}$$

and uses the Monte Carlo approach to find an additive correction to this user-supplied guess.

The procedure for this method is to:

15. Choose a value of $\hat{x}_i$ uniformly between a and b.
16. Score $I_i = (f(\hat{x}_i) - h(\hat{x}_i)) \cdot (b-a) + I_h$.

Notice that there is NO variance introduced through the I_h part of the score, but only in the difference between $f(x)$ and $h(x)$. Obviously, then a good guess will result in a small difference and, therefore a small variance; in the limit of a perfect guess, $h(x) = f(x)$, there would be no correction and no therefore no variance. (Note that there would also be no variance introduced if $h(x) = f(x) + \text{constant}$ since, again, each guess would be identical and correct.)

In terms of our mathematical framework, this is equivalent to again using:

$$\pi(x) = \frac{1}{b-a} \tag{4-9}$$

(for a uniform distribution between a and b) and scoring with:

$$I_i = \frac{f(\hat{x}_i) - h(\hat{x}_i) + E(h(x))}{\pi(\hat{x}_i)}$$

$$= \frac{f(\hat{x}_i) - h(\hat{x}_i) + \frac{\int_a^b h(x)dx}{\int_a^b dx}}{\pi(\hat{x}_i)} \quad . \qquad (4\text{-}10)$$

$$= \frac{f(\hat{x}_i) - h(\hat{x}_i) + \frac{I_h}{b-a}}{\pi(\hat{x}_i)}$$

$$= \frac{f(\hat{x}_i) - h(\hat{x}_i)}{\pi(\hat{x}_i)} + I_h$$

Example 3: Again find $I = \int_0^2 x^2 dx$, this time using a control variates method with $h(x) = \frac{x^3}{2}$. (Note that $h(x)$ matches $f(x)$ at the endpoints and is similarly uniformly increasing.)

Answer: Noting the integral of $h(x)$ over the domain (0,2) is 2. With this value known, the procedure is to:

17. Choose a value of $\hat{x}_i$ uniformly between 0 and 2.
18. Score $I_i = 2(\hat{x}_i^2 - \frac{\hat{x}_i^3}{2}) + 2$.

A FORTRAN computer program was written to solve this problem. The results are:

Number of histories	$\hat{I}$	S_I
10,000	2.6643	0.00408
100,000	2.6670	0.00128
1,000,000	2.6649	0.00041

The estimated standard deviations are lower than for either of the two previous examples.

Importance sampling method

The final method is the importance sampling method. This technique is similar to the control variates method, in that it takes advantage of *a priori* knowledge about the function $f(x)$, but differs from it in that its correction is multiplicative rather than additive.

In terms of our mathematical framework, the importance sampling method uses the approximate function $h(x)$ as the (unnormalized) probability distribution from which the variables $\hat{x}_i$ are drawn:

$$\pi(x) = \frac{h(x)}{\int_a^b h(x)\,dx} = \frac{h(x)}{I_h}, \qquad (4\text{-}11)$$

)

With the resulting score of:

$$I_i = \frac{f(\hat{x}_i)}{\pi(\hat{x}_i)} = \left[\frac{f(\hat{x}_i)}{h(\hat{x}_i)}\right] I_h \qquad (4\text{-}12)$$

As with control variates, a "perfect" guess of $h(x) = f(x)$ would result in a zero variance solution, this time because, again, every guess would be the correct one, I_h. (Note that any guess that satisfies $h(x) = f(x) \times \text{constant}$ would do just as well, since it would normalize to the same $\pi(x)$.)

Example 4: One last time, find $I = \int_0^2 x^2 dx$, this time using an importance sampling method with the same guess that worked so well before, $h(x) = \frac{x^3}{2}$.

Answer: Since the integral of h(x) over the domain (0,2) is 2, the resulting probability distribution from which to pick the $\hat{x}_i$ will be:

$$\pi(x) = \frac{x^3}{4} \qquad (4\text{-}13)$$

Following the direct procedure for choosing from this distribution, we first determine the CDF, which is:

$$\Pi(x) = \int_0^x \frac{x'^3}{4} dx' = \frac{x^4}{16} \tag{4-14}$$

We then set this CDF to the uniform deviate:

$$\xi = \frac{x^4}{16} \tag{4-15}$$

and invert to get the formula:

$$\hat{x}_i = 2\sqrt[4]{\xi_i} \tag{4-16}$$

As before, a FORTRAN computer program was written to solve this problem. The results are:

Number of histories	$\hat{I}$	S_I
10,000	2.6860	0.01036
100,000	2.6700	0.00308
1,000,000	2.6651	0.00095

The estimated standard deviations are greater than those for the control variates problem (Example 3), but less than those for the other two methods.

Chapter 4 Exercises

4-1. Design and code a Monte Carlo estimate of the integral:

$$I = \int_0^3 e^x dx$$

using the rejection method.

4-2. Design and code a Monte Carlo estimate of the integral:

$$I = \int_0^3 e^x dx$$

using the averaging method.

4-3. Design and code a Monte Carlo estimate of the integral:

$$I = \int_0^3 e^x dx$$

using the control variates method. Use a guess function of $h(x) = 1 + 0.707x^3$.

4-4. Design and code a Monte Carlo estimate of the integral:

$$I = \int_0^3 e^x dx$$

using the importance sampling method. Use a guess function of $h(x) = 1 + 0.707x^3$.

4-4. Design, code, and run a Monte Carlo estimate of the integral (using averaging method)

$$I = \int_0^2 \int_0^2 \int_0^2 (x + y + z)^2 dx\, dy\, dz$$

for:

a. pseudorandom numbers
b. Halton sequence using bases 2, 3, and 5 for the three dimensions.

Present a plot of error (not standard deviation of the mean) vs. N.

Answers to selected exercises

Chapter 4

4-1. $19.08 \pm \frac{28.02}{\sqrt{N}}$

4-2. $19.08 \pm \frac{15.46}{\sqrt{N}}$

4-3. $19.08 \pm \frac{1.232}{\sqrt{N}}$

4-4. $19.08 \pm \frac{2.889}{\sqrt{N}}$

4-4. a. $80 \pm \frac{48.92}{\sqrt{N}}$