Monte Carlo integration

Looking at the Rutherford scattering example, we can determine the probability of scattering through more than 90° (i.e. the probability of b being less than bmax) as:

$\int_0^{b_{max}} p(r)dr = \frac{1}{\sigma^2} \int_0^{b_{max}} exp(-\frac{r^2}{2\sigma^2})rdr$

where

$b_{max} = \frac{Ze^2}{2\pi\epsilon_0 E}.$

Then,

$\int_0^{b_{max}} p(r)dr = \frac{1}{\sigma^2} \int_0^{b_{max}} exp(-\frac{r^2}{2\sigma^2})rdr = 1 - exp(-\frac{b_{max}^2}{2\sigma^2})$

$\int_0^{b_{max}} p(r)dr = 1 - exp(-\frac{Z^2e^4}{8\pi^2\epsilon_0^2E^2\sigma^2})$ (1)

Replacing the values from the previos example:

$Z = 79$

$e = 1.602 \text{ x } 10^{-19}$

$E = 7.7 \text{ x } 10^6 e$

$\epsilon_0 = 8.854 \text{ x } 10^{-12}$

$a_0 = 5.292 \text{ x } 10^{-11}$

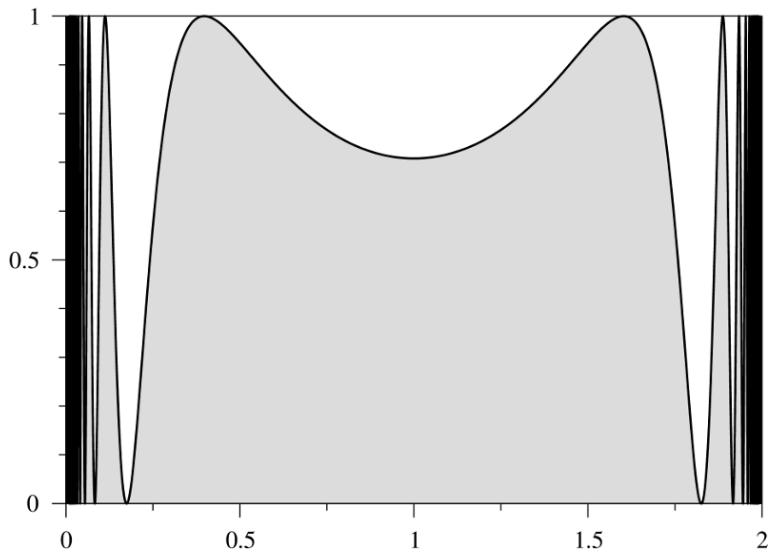
$\sigma = a_0/100$

we get a fraction of reflected particles equal to 0.156%, in good agreement with the result of the Python program. Now, look at this in a different way. If equation 1 and our Python computer program calculate the same thing, it implies that we can, if we want, calculate the value of the integral in equation 1, an integral that has a known exact value, approximately by simulating a random process using random numbers. This is actually quite a deep result: we can calculate the answers to exact calculations by doing random calculations. Normally, we think of the process the other way around: we are interested in some physical process that has a random element, but instead we write down an exact, non-random calculation that gives an answer for its average behaviour. But there is no reason in principle why we cannot make an argument in reverse: **start with an exact problem, such as the calculation of an integral, and find an approximate solution to it by running a suitable random process on the computer**. This leads us to a novel way of performing integrals, which works as follows:

Suppose we want to evaluate the integral

$I = \int_0^2 sin^2\left[\frac{1}{x(2-x)}\right] dx$

The shape of such an integrand is this:



This is a challeging integral because the function varies a lot as it approches the edges. On the other hand, since the entire function fits inside a 2x1 rectangle, **the value of the integral is perfectly finite and must be less than 2**, so in principle at least there is a well-defined answer for this integral. Here, neither basic methods like the Trapezium rule or more advanced ones like the Gaussian quadrature method are going to work well. They will not be able to capture the infinitely fast variation of the function at the edges.

How to tackle this problem? If the shaded area under the curve is I and given that the area of the bounding rectangle is A = 2, if we choose a point uniformly at random in the rectangle, the probability that that point falls in the shaded region (under the curve as oppose to over the curve) is p = I/A. So here is the scheme:

- we generate a large number N of random points in the bounding rectangle,
- check each one to see if it´s below the curve and keep a count of the number that are. Let's call this number k.
- Then, the fraction of points below the curve, which is k/N, should be approximately equal to the probability p. That is:

$$\mathbf{I} \simeq \frac{\mathbf{kA}}{\mathbf{N}}$$

This technique is called **Monte Carlo integration**, after the famous casino town in Monaco. It uses a random process to calculate the answer to an exact, nonrandom question. This integration method is particularly useful for problems like this one where the integrand is pathological and also for multiple integrals in high dimensions.

Example:

Write a program to calculate the afformentioned integral using Monte Carlo integration.

$I = \int_0^2 sin^2 \left[\frac{1}{x(2-x)} \right] dx$

The main disadvantage of the MC method is that it does not give very accurate answers. For simple integrals where we can use the trapezium rule or Simpson's rule, it normally gives worse results than those methods. How much worse?

- The probability that a single random point falls below the curve is p = I/A
- The probability that it falls above the curve is 1 - p
- Thus, the probability that a particular k of our points fall below and the remaining N-k fall above is: $p^k(1-p)^{N-k}$
- There are C_k^N ways to choose the k points out of N total, so the total probability P(k) that we get exactly k points below the curve is (binomial distribution):

$$P(k) = C_k^N p^k (1-p)^{N-k}\,.$$

Then, the variance is:

$$\sigma_k^2 = Np(1-p) = N\frac{I}{A}\left(1-\frac{I}{A}\right)$$

This gives us an estimate of the expected variation or error in the value of k. Then, the expected error on the integral I itself is:

$$\sigma_I = \sigma_k \frac{A}{N} = \frac{\sqrt{I(A-I)}}{\sqrt{N}}$$

This means, the error varies with N as $N^{-1/2}$, which means the accuracy improves the more random samples we take. However, $N^{-1/2}$ is not a very impressive rate of improvement. If we improve the number of samples by a factor of 100, the size of the error will go down by a factor of 10. By contrast, with the trapezium rule the error goes down by a factor of 10 000, and with Simpson's rule by a factor even greater, of 100 000 000. Clearly, if we can use regular integration methods, we should. **MC integration should be used only for cases where other methods break down.**

Mean value method

Even among MC methods, the method described above is not very good. There are significant better ways to perform the calculation, the most common of which is the mean value method. Let's take a general integration problem:

$I = \int_a^b f(x)dx$

The average value of f(x) in the range from a to b is by definition:

$\langle f \rangle = \frac{1}{b-a}\int_a^b f(x)dx = \frac{I}{b-a}$

Thus,

$I = (b-a)\langle f \rangle$

If we can estimate $\langle f \rangle$, then we can estimate I. A simple way of estimating $\langle f \rangle$ is just to measure f(x) at N points $x_1, \ldots, x_N$ choosen uniformly at random between a and b and calculate

$\langle f \rangle \simeq \frac{1}{N}\sum_{i=1}^N f(x_i)$

Then,

$$\mathbf{I \simeq \frac{b-a}{N} \sum_{i=1}^{N} f(x_i)}$$

This is the fundamental formula for the mean value method. As with the previous method it gives only an approximate estimate of the integral. How accurate is it? We can estimate the error using standard results for the behaviour of random variables. The variance for the values $f(x_i)$ can be calculated as:

$\sigma_f^2 = \left\langle f^2 \right\rangle - \left\langle f \right\rangle^2 = \frac{1}{N} \sum_{i=1}^{N} f(x_i)^2 - \left[\frac{1}{N} \sum_{i=1}^{N} f(x_i) \right]^2$

Therefore, the variance for N values of $f(x_i)$ is $N\sigma_f^2$, and the standard deviation of I is:

$\sigma = \frac{(b-a)}{N} \sqrt{N} \sigma_f$

$\sigma = (b-a) \frac{\sigma_f}{\sqrt{N}}$

Once again, the error goes as $N^{-1/2}$. However, the leading constant is smaller in this case, which means the mean value method is always more accurate then our previous method.

Integrals in many dimensions

In addition to the integration of pathological functions, Monte Carlo integration is used for performing high-dimensional integrals. Performing two-dimensional integrals by standard methods requires a 2D grid of points, performing 3D integrals requires a 3D grid of points, and so forth; i.e. the number of points on the grid can become very large and the standard integration methods can become very slow. On the other hand, Monte Carlo integration gives reasonably good results for much smaller number of points, a few thousand is often adequate. The mean value method generalizes straightforwardly. The integral of a function f($\mathbf{r}$) over a volume V in a high-dimensional space is given by:

$$I \simeq \frac{V}{N} \sum_{i=1}^{N} f(\mathbf{r_i}) \qquad (1)$$

where the points $\mathbf{r_i}$ are picked uniformly at random from the volume V.

An important application of this type of integral is in finance, where speed is necessary for trading decisions.

Example: volume of a hypersphere

Estimate the volume of a sphere of unit radius in ten dimensions using the Monte Carlo integration method. First, consider the equivalent problem in two dimensions: the area of a circle of unit radius.

$I = \int \int_{-1}^{+1} f(x,y) dx dy$

where

$$f(x,y) = \begin{cases} 1 & x^2 + y^2 \leq 1 \\ 0 & otherwise \end{cases}$$

In order to calculate this integral we can generate a set of N random points (x,y) where both x and y are in the range from -1 to +1. Then, the 2D version of equation 1 is:

$$I \simeq \frac{4}{N} \sum_{i=1}^{N} f(x_i, y_i)$$

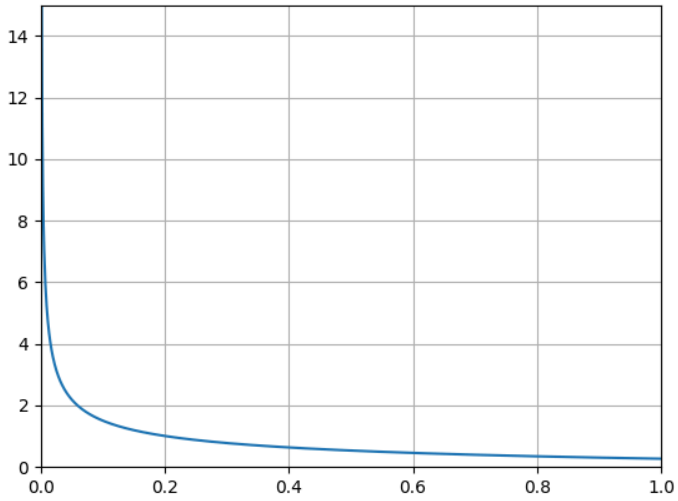
Generalize this method to the ten-dimensional case and write a program to perform a MC calculation of the volume of a sphere in ten dmensions.

Importance sampling

MC integration does not work well with functions that contain a divergence. Consider the integral:

$$I = \int_0^1 \frac{x^{-1/2}}{e^x + 1} dx$$

Even though the integrand diverges at x = 0, the integral is perfectly finite in value.



But if you try to do it using the mean value method, you will run into problems because que value of $f(x_i)$ diverges when $x_i \to 0$. So, occasionally you will get a very large contribution to the sum in the equation that defines the mean value method. This means that the estimated value of the integral can vary widely from one run to the next. Another way of saying this is that the error (σ) on the estimate of the integral can become very large. **We can get around this problems by drawing our points x_i non-uniformly from the integration interval**. This technique is called *importance sampling* and it works as follows:

For any general function, we can define a weighted average over the interval from a to b thus:

$$\langle g \rangle_w = \frac{\int_a^b w(x) g(x) dx}{\int_a^b w(x) dx}$$

where $w(x)$ is any function we choose.

Now consider again the general 1D integral:

$I = \int_a^b f(x)dx$

Setting $g(x) = f(x)/w(x)$ we have:

$$\left\langle \frac{f(x)}{w(x)} \right\rangle_w = \frac{\int_a^b w(x)f(x)/w(x)dx}{\int_a^b w(x)dx} = \frac{\int_a^b f(x)dx}{\int_a^b w(x)dx} = \frac{I}{\int_a^b w(x)dx}$$

$$I = \left\langle \frac{f(x)}{w(x)} \right\rangle_w \int_a^b w(x)dx$$

$$\mathbf{I = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{w(x_i)} \int_a^b w(x)dx} \qquad (2)$$

where the points x_i are chosen from the probability distribution:

$$\mathbf{p(x) = \frac{w(x)}{\int_a^b w(x)dx}}$$

This equation is analogous to the mean value equation, but it allows us to calculate the value of the integral from a weighted average, rather than a standard uniform average. If we choose w(x) = 1, then we recover the equation for the mean value method we learnt before.

The formula allows us to calculate an estimate of the integral I by calculating not the sum $\sum_{i=1}^{N} f(x_i)$, but instead the modified sum $\sum_{i=1}^{N} \frac{f(x_i)}{w(x_i)}$, where $w(x)$ is any function we choose. This is useful because it allows us to choose a w(x) that gets rid of pathologies in the integrand f(x). The price we pay is that we have to draw our samples x_i from a nonuniform distribution of random numbers instead of a uniform one, which makes the programming more complex since we have to use for example the transformation method.

The error σ can be calculated using again the properties of variances of random variables. The result is:

$$\sigma = \frac{\sigma_w(f/w)}{\sqrt{N}} \int_a^b w(x)dx$$

where

$$\sigma_w(f/w) = \left\langle (f/w)^2 \right\rangle_w - \left\langle (f/w) \right\rangle_w^2$$

Example:

Use the importance sampling method to evaluate the integral:

$I = \int_0^1 \frac{x^{-1/2}}{e^x+1} dx$

Choose $w(x) = x^{-1/2}$, then

$f(x)/w(x) = \frac{1}{e^x+1}$

which is finite and well-behaved over the integration domain. Use the transformation method to draw random x values in the range from 0 to 1 from the following prob. distribution:

$$p(x) = \frac{x^{-1/2}}{\int_0^1 x^{-1/2} dx} = \frac{1}{2\sqrt{x}}$$

Then, use equation 2 to get the value for the integral. Note that

$$\int_0^1 w(x) dx = \int_0^1 x^{-1/2} dx = 2$$

Sample N = 1 000 000 random points. You should get a value of 0.84

In [ ]: