05 February 2025 – PHYS 6260
Prof. John Wise
Newman Chapter 10: Random Processes and Monte Carlo Methods

10.2: Monte Carlo Integration

The last example in the previous lecture considered Rutherford scattering of α-particles from a gold atom. Although we performed a calculation that involved a random sampling of the beam, it can be solved analytically by integrating the Gaussian beam profile and determining what fraction of the particles are within an impact parameter b,

$$\begin{aligned}\frac{1}{\sigma^2}\int_0^b \exp\left(-\frac{r^2}{2\sigma^2}\right) r\,dr &= 1 - \exp\left(-\frac{b^2}{2\sigma^2}\right) \\ &= 1 - \exp\left(-\frac{Z^2e^4}{8\pi^2\epsilon_0^2\sigma^2E^2}\right). \end{aligned} \tag{1}$$

Taking the parameters that we considered in the example, we find that 0.156% of the particles will back-scatter ($\theta > \pi/2$, which is consistent with our random calculation.

We can model a system that has an analytical result by modeling it with a random processes with random numbers. This is a deep result – we can calculate answers to exact calculations with random calculations. We will first apply this idea to numerical integration, known as **Monte Carlo integration**.

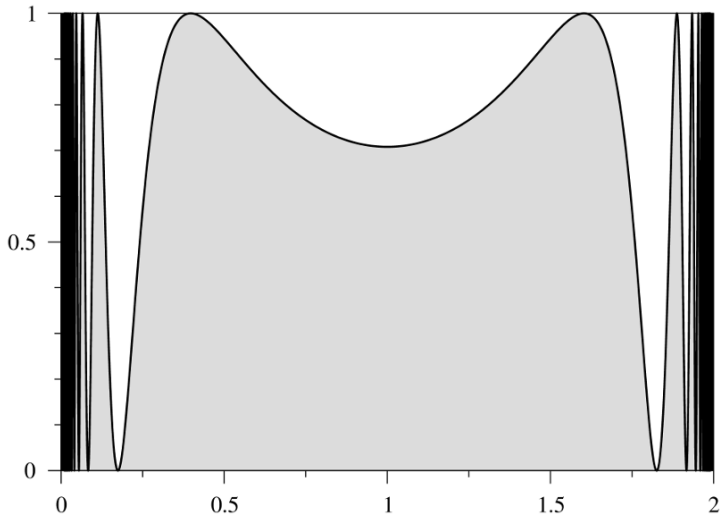

Figure 1: A pathological function

Consider an ill-behaved analytical function, that we want to integrate from 0 to 2

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

that is shown in Figure 1. One can see that it varies rapidly (approaching infinitely) at the edges of the integration limits. Using the integration methods previously discussed in the class would result in an inaccurate result because of the rapid variations.

However, we can take the following approach. We first notice that the integral is completely bounded by a 2×1 box with an area $A = 2$. We can choose N uniformly random numbers within the rectangle and determine how many, let's call that number k, lie below the function. This event should have a probability $p = I/A$ of occurring. Thus, the fraction of random points below the curve should equal to the probability, $k/N \simeq I/A$, leading to

$$I \simeq \frac{kA}{N} \tag{3}$$

Below you can see an example program that does exactly what was described above.

```
#
# To be completed in class
#
```

or in its vectorized form

```
#
# To be completed in class
#
```

When run, it should give an answer around `1.45`, which is accurate to 2 decimal places.

The problem with MC integration is that it's not accurate, even if we increase the number N of sample points. Let's analytically inspect how the error decreases as N is increased. The probability that the point lies beneath the curve is $p = I/A$ and above the curve $1-p$. Thus, the probability that k and $N-k$ of our points lie below and above the curve, respectively, is $p^k(1-p)^{N-k}$. There are ${}_NC_k$ ways to choose k out of N points. So the total probability $P(k)$ that we get exactly k below the curve is

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k}, \tag{4}$$

which is called the *binomial distribution*. The variance of this type of distribution is

$$\mathrm{var}k = Np(1-p) = N\frac{I}{A}\left(1 - \frac{I}{A}\right). \tag{5}$$

which is related to the standard deviation

$$\sigma = \sqrt{\mathrm{var}k}. \tag{6}$$

We can relate this to the error of the integral (using Equation 3) by

$$\sigma = \sqrt{\mathrm{var}k}\frac{A}{N} = \frac{\sqrt{I(A-I)}}{\sqrt{N}}. \tag{7}$$

Therefore the error decreases as $N^{-1/2}$, which is much worse than any of the integration techniques we've covered in the past.

The mean value method

The most basic MC integration method may not give accurate results at low values of N, but there are other MC methods that give improved results. The most common one is called the **mean value method**, which starts with the average value of a function $f(x)$ over the integration limits,

$$\langle f \rangle = \frac{1}{b-a}\int_a^b f(x)\,dx = \frac{I}{b-a}, \tag{8}$$

leading to

$$I = (b-a)\langle f \rangle \tag{9}$$

We can either calculate $\langle f \rangle$ by either selecting a uniformly spaced or random values of x, giving

$$I \simeq \frac{b-a}{N}\sum_{i=1}^{N} f(x_i) \tag{10}$$

This estimate is more accurate for slowly varying functions, which can calculate with its variance, $\mathrm{var}f = \langle f^2 \rangle - \langle f \rangle^2$. The variance on the sum in the above equation is N times the variance of a single term, $N\mathrm{var}f$, and the standard deviation is the square root of it. Using this definition with the integral value (Equation 10), we find the error

$$\sigma = \frac{b-a}{N}\sqrt{N\mathrm{var}f} = (b-a)\frac{\sqrt{\mathrm{var}f}}{\sqrt{N}}. \tag{11}$$

The error still behaves as $N^{-1/2}$, but the pre-factor is smaller than the basic MC integration method.

Integrals in many dimensions

Both the basic and mean value MC integration methods extend easily to multiple dimensions. With traditional integration methods, we require to have N-dimensional grids, which are fine for two and three dimensions. However, the number of sample points becomes computational intractable at some point. For instance, consider a six-dimensional space (phase space: three

spatial and three velocity components). If we require 100 points in each dimension, it would require $100^6 = 10^{12}$ points! But with the mean value method, we would simply calculate,

$$I \simeq \frac{V}{N} \sum_{i=1}^{N} f(\mathbf{r}_i), \tag{12}$$

where $\mathbf{r}_i$ are random points chosen in the volume V.

Importance sampling

MC integration is useful for pathological functions, but it can break down in diverging functions. For example, take the integral

$$I = \int_0^1 \frac{x^{-1/2}}{e^x + 1}\, dx, \tag{13}$$

which is important when calculating properties of Fermi gases (electrons in a white dwarf; neutrons in a neutron star). This integral has a finite value even though it diverges at $x = 0$.

In an MC integration, we can favor the random values of x that don't overlap with the diverging part of the function. The error of the mean value method depends on the variations of the function, which can cause problems if the function diverges. The method of **importance sampling** will draw a non-uniform random sample that avoids the diverging portion.

First, let's define a weighted mean of a function $g(x)$,

$$\langle g \rangle_w = \frac{\int_a^b w(x) g(x)\, dx}{\int_a^b w(x)\, dx}. \tag{14}$$

In this derivation, we will still use $I = \int_a^b f(x)\, dx$. By setting $g(x) = f(x)/w(x)$, we have

$$\left\langle \frac{f(x)}{w(x)} \right\rangle_w = \frac{\int_a^b w(x) f(x)/w(x)\, dx}{\int_a^b w(x)\, dx} = \frac{I}{\int_a^b w(x)\, dx}. \tag{15}$$

Solving for I,

$$I = \left\langle \frac{f(x)}{w(x)} \right\rangle_w \int_a^b w(x)\, dx. \tag{16}$$

But how do we determine the weighting function? We can equate the probability density function, which is set by the physics / math, as the normalized weight function,

$$p(x) = \frac{w(x)}{\int_a^b w(x)\, dx} \tag{17}$$

But in practice, we want to pick a $w(x)$ to factor out any diverging nature of the function $f(x)$. Now we can draw N random points from this non-uniform distribution. The average number of samples that fall between x and $x + dx$ is $Np(x)\,dx$, so for any function $g(x)$, we have

$$\sum_{i=1}^{N} g(x_i) \simeq \int_a^b Np(x)g(x)\,dx. \tag{18}$$

Now we can write the general weighted average of the function $g(x)$ as

$$\langle g \rangle_w = \frac{\int_a^b w(x)g(x)\,dx}{\int_a^b w(x)\,dx} = \int_a^b p(x)g(x)\,dx \simeq \frac{1}{N}\sum_{i=1}^{N} g(x_i), \tag{19}$$

where x_i are randomly chosen from $p(x)$. Using this with Equation (16), we find the integral

$$I \simeq \frac{1}{N}\sum_{i=1}^{N} \frac{f(x_i)}{w(x_i)} \int_a^b w(x)\,dx. \tag{20}$$

If we set $w(x) = 1$, then we recover the mean value method. Also, the error associated with the integral still behaves as $N^{-1/2}$ but it can be smaller than the mean value method if the weighting function is chosen wisely.

Let's look at the integral dealing with Fermi gases again,

$$I = \int_0^1 \frac{x^{-1/2}}{e^x + 1}\,dx, \tag{21}$$

If we choose $w(x) = x^{-1/2}$ then we remove the diverging portion of the integrand. To numerically integrate this, we would randomly choose values x_i from the distribution (see Equation 17)

$$p(x) = \frac{x^{-1/2}}{\int_0^1 x^{-1/2}\,dx} = \frac{1}{2\sqrt{x}} \tag{22}$$

and then plug them into the original integrand $f(x_i)$ and weighting function $w(x_i)$ and evaluate Equation (20) to find the integral, using that $\int_0^1 w(x)\,dx = 2$.

We can also use importance sampling to evaluate integrals over infinite domains. Note that the mean value method breaks down over such a range because the random numbers can't sample all of real space. So we can create an exponential weighting function $w(x) = e^{-x}$ to place more importance on random values that are located in regions that are sparsely sampled. Thus, the integral with importance sampling becomes

$$I \simeq \frac{1}{N}\sum_{i=1}^{N} \exp(x_i) f(x_i) \int_0^\infty e^x\,dx = \frac{1}{N}\sum_{i=1}^{N} \exp(x_i) f(x_i) \tag{23}$$

This may seem like overkill for a one-dimensional integral, as we've studied methods that allow for integrals over infinite domains by changing the variables. However, importance sampling and Monte Carlo integration become attractive when dealing with multiple dimensions, especially phase space, which we will explore in more depth in the next lecture, using statistical mechanics as an example.