Practice 3

Monte Carlo method to estimate sphere volume

The aim of this practice is to use both **crude Monte Carlo** and **Importance Sampling** methods to obtain an estimation of the volume of a D-dimensional sphere.

Methodology

In order to do that, we will follow these steps:

1. Generate D random variables according to a uniform distribution, such that $(x_1, x_2, \ldots, x_D) \in [0,1]^D$.
2. Check that the random sampled points are inside the sphere. That is, if $\sum_{i=1}^{D} x_i^2 < 1$ then we count it as a hit.
3. We repeat steps 1. and 2. N times and keep counting the number of hits.
4. Finally, approximate the volume of a D dimensional sphere as

$$V \approx (2R)^D \frac{N_{\text{hits}}}{N_{\text{tries}}}$$

   where the $(2R)^D$ factor comes from the fact that sampling D uniform variables in $(0,1)$ we only account for the region of the sphere where all variables are positive. So, in order to obtain an approximation of the volume of a sphere of radius R (centered at the origin), we have to multiply by $(2R)^D$.

5. Then, we can compare this estimation with the exact volume of a D-dimensional sphere. In the case of a circle (2D sphere), the volume (or the area) is simply:

$$A = V_2 = \pi R^2$$

   while in 3 dimensions it is,

$$V_3 = \frac{4}{3}\pi R^3$$

In general, we can follow a recurrence rule for the volume of a D-dimensional sphere that will allow its computation in approximately $D/2$ steps:

$$V_D(R) = \begin{cases} 1 & \text{if } n = 0 \\ 2R & \text{if } n = 1 \\ \frac{2\pi}{D} R^2 \times V_{D-2} & \text{otherwise} \end{cases}$$

3.1 First Part: Estimate area of a circle

Consider a circle of unit radius, $R = 1$ in two dimensions. We can generate N random numbers for each coordinate $0 < x < 1$ and $0 < y < 1$ using uniform random distribution. Count the number of times that a pair (x, y) lies inside the circle. Namely, check that $x_i^2 + y_i^2 < 1$ for each sample $i = 1, \ldots, N$ of the iterations and use this number to estimate the area of a circle.

Finally, compare it with the exact area of a circle and plot the relative error as well as the statistical error.

3.1.1 Results

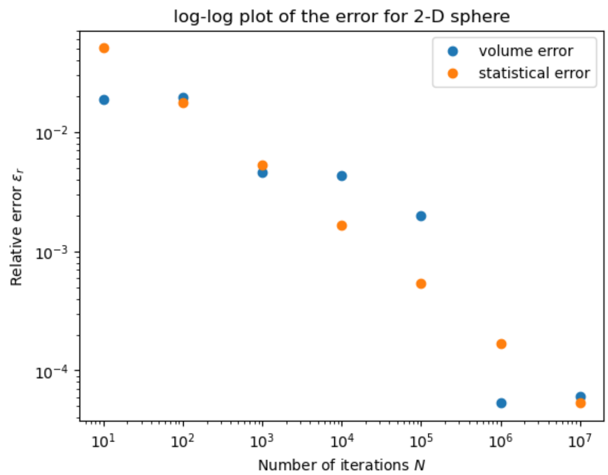

Figure 3.1: Error of the volume estimation for the 2D sphere.

From figure 3.1, we can see that the crude Monte Carlo method can approximate the area of the circle quite well (up to 4 digits of precision) and the statistical error is a good measure of the exact error.

3.2 Second Part: Volume of a sphere in 3D

Consider now a sphere of unit radius $R = 1$ in 3 dimensions. Repeat the same procedure as in 2D, but now generate an additional random variable $0 < z < 1$ and check whether the triplet (x, y, z) is inside the sphere. In other words, for each iteration i, count how many times the condition $x_i^2 + y_i^2 + z_i^2 < 1$ is fulfilled.

Compare it with the exact result of the volume, and report the statistical error.

3.2.1 Results

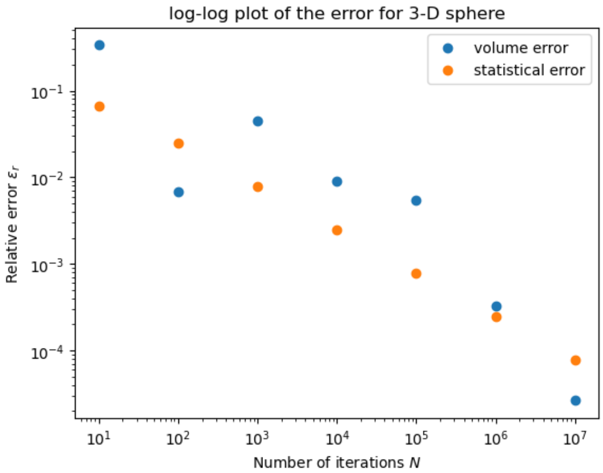

Figure 3.2: Error of the volume estimation for the 3D sphere.

Once again, Figure 3.2 shows that our estimation with the crude Monte Carlo method allows us to calculate the volume of a sphere with good precision. In addition, we can say that the statistical error correctly estimates (on average) the actual error of our approximation.

3.3 Third Part: Volume of a sphere in D dimensions

Now let us generalize to the D-dimensional case. Generate D random numbers $0 < x_j < 1$, $j = 1, \dots, D$ for N times. Calculate the probability that a point $(x_1, \dots, x_D)$ is inside the sphere, i.e., $\sum_{i=1}^{D} x_i^2 < 1$. Use this result to approximate the volume of the sphere and compare it with the exact result, as well as the statistical error.

What is the largest space dimensionality, in which this method can be reliably used? What is the ratio between the volume of the sphere of diameter $2R$ and a cube with side $2R$ in D dimensions?

3.3.1 Results

First, let us try to estimate the volume of a 10-dimensional sphere.

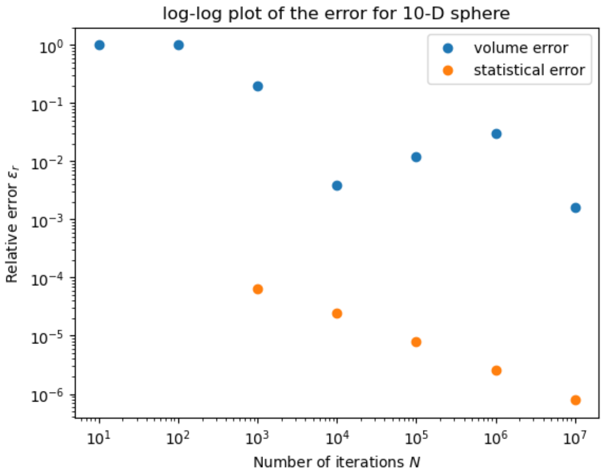

Figure 3.3: Error of the volume estimation for the 10D sphere.

Here, we can observe a discrepancy between the statistical error and the error of our approximation. Specifically, for $N = 10$, 100, we can see that our error is of the order of the unit. This is because with little tries, the number of times we generate a point inside the sphere is 0, and thus we cannot approximate the volume. However, increasing the number of tries seems to somewhat approximate the volume, although with bad precision.

To see what is happening, let us plot the ratio between a hypersphere and a hypercube for increasing values of D:

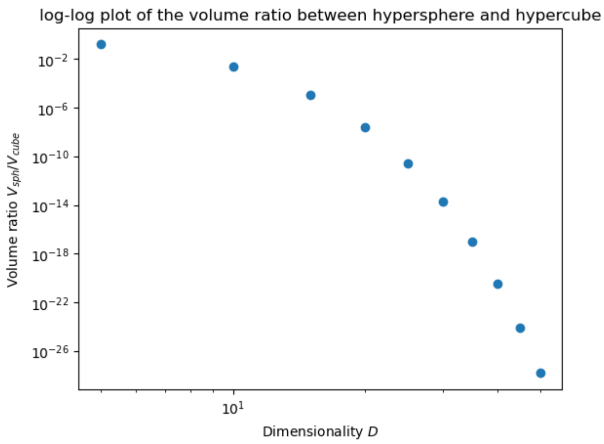

Figure 3.4: Ratio between volume of cube and sphere in D dimensions.

Notice that the ratio between sphere and cube volume decays faster than an exponential with the dimensionality. This ratio can give us a clue of what would happen if we were to use crude Monte Carlo to approximate the volume of the sphere: For $D = 10$, the ratio becomes $1/1000$, meaning that we would need ~ 1000 tries on average to generate a single point inside the sphere. This result is confirmed by Figure 3.3 as we needed $N = 1000$ iterations in order to have a non-zero approximation of the volume.

To answer the question of what would be the largest space dimensionality to use crude Monte Carlo, we can assume that our maximum number of tries should be $N = 10^8$, since at this point the computation would be too slow. Then, looking at Figure 3.4, we see that $D > 20$ we need (on average) more than 10^8 tries to get a hit. We can check this is the case by trying to calculate the volume of a sphere in $D = 25$ dimensions using crude Monte Carlo:

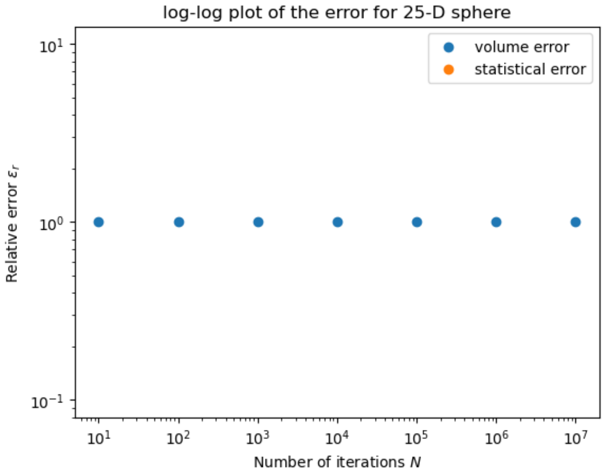

Figure 3.5: Error of the volume estimation for the 25D sphere.

It is clear from Figure 3.5 that we need to use a different method if we want to estimate volumes in higher dimensions.

3.4 Fourth Part: Volume of a sphere in 100 dimensions

The crude Monte Carlo method does not work well if the sphere volume is significantly smaller than the volume of the cube in which we generate the random variables. A possible way out is to sample more frequently the region close to the origin by using an **Importance Sampling** technique:

1. We generate the random numbers according to a normal distribution.
2. We adjust the mean value to 0 (around the origin), and let the variance be a free parameter σ_p:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_p} e^{-x^2/2\sigma_p^2}$$

3. This degree of freedom can be used to minimize the variance and the resulting statistical error. But most importantly, we can adjust it in order to sample more points inside of the sphere in high dimensionality spaces.

One can find the optimal value of this free parameter by considering the important region of interest as the edge of the sphere, $\sum_{i=1}^{D} x_i^2 = R^2$, which after averaging equals to $\langle \sum_{i=1}^{D} x_i^2 \rangle = D\sigma_p^2 = R^2$. This leads to an optimal value of the free parameter $\sigma_p = R/\sqrt{D}$.

Now let us see if we can estimate the volume of a 100-dimensional sphere with good precision.

3.4.1 Results

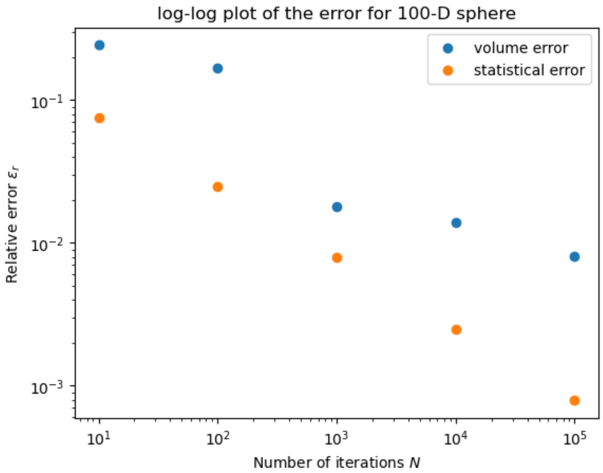

Figure 3.6: Error of the volume estimation for the 100D sphere.

As we can see, using Importance Sampling, we are able to estimate the volume of the 100-dimensional sphere with good precision and using less number of tries.