Monte Carlo Importance Sampling

Varun Subudhi (2011188)

National Institute of Science Education and Research (NISER)

Abstract: Monte Carlo is an integration technique of approximating an expectation by the sample mean of a function of simulated random variables. In this project, we try to estimate the volume of a 3-Dimensional shape known as Ellipsoid with its equation while employing importance sampling to the reduce the variance of Monte Carlo approximations.

Introduction:

Monte Carlo Simulation is a mathematical method for calculating the odds of multiple possible outcomes occurring in an uncertain process through repeated random sampling.

This computational algorithm makes assessing risks associated with a particular process convenient, thereby enabling better decision-making. this statistical technique uses randomness to solve probabilistic problems. Also, it can perform sensitivity analysis and correlation between input variables. It finds its application in prediction and forecasting models in business, supply chain, project management, finance, science, engineering, particle physics, artificial intelligence, astronomy, meteorology, sales forecasting, and stock pricing.

Importance sampling is a variance reduction technique that can be used in the Monte Carlo method. The idea behind importance sampling is that certain values of the input random variables in a simulation have more impact on the parameter being estimated than others. If these important values are emphasized by sampling more frequently, then the estimator variance can be reduced. Hence, the basic methodology in importance sampling is to choose a distribution which encourages the important values. This use of biased distributions will result in a biased estimator if it is applied directly in the simulation. However, the simulation outputs are weighted to correct for the use of the biased distribution, and this ensures that the new importance sampling estimator is unbiased.

In this project, we use this technique to estimate the volume of ellipsoid and also correlate between the accuracy in estimation and number of random points generated for estimation.

The Problem:

Suppose that we need to compute the expected value:

$$E[g(X)]$$

of a function of a random vector X by Monte Carlo integration.

The standard way to proceed is to produce a computer-generated sample of realizations of n independent random vectors $X_1,\ldots,X_n$ having the same distribution as X.

Then, we use the sample mean:

$$\bar{g}_{X,n} = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

to approximate the expected value.

Thanks to the propositions in the previous section, we can compute an alternative Monte Carlo approximation of E[g(X)] by extracting n independent draws $Y_1,...,Y_n$ from the distribution of another random vector Y (in what follows we assume that it is discrete, but everything we say applies also to continuous vectors).

Then, we use the sample mean:

$$\bar{g}_{Y,n} = \frac{1}{n}\sum_{i=1}^{n}\frac{p_X(Y_i)}{p_y(Y_i)}g(Y_i)$$

as an approximation.

The reason why we use importance sampling is that we can often choose Y in such a way that the variance of the approximation error is much smaller than the variance of the standard Monte Carlo approximation.

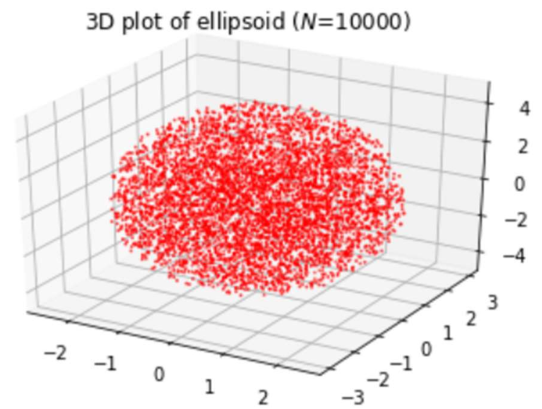
For our project, the basic idea is to construct a cuboid of dimensions (2a, 2b, 2c), enclose the ellipsoid within it and then generate random points within our constructed cuboid volume and note down the number of points that happen to fall within the ellipsoidal volume.

We then calculate the following ratio as:

$$\mathit{fraction\ of\ points\ within\ ellipsoid}$$

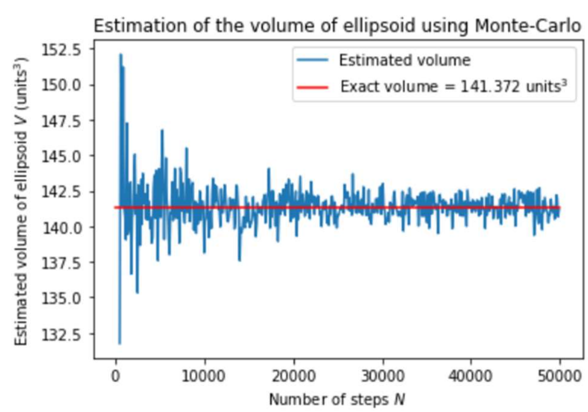
$$= \frac{\mathit{no. of\ points\ within\ ellipsoid}}{\mathit{total\ no. of\ points}}$$

The product of this ratio and the volume of the constructed hypothetical cuboid yields an approximate volume of the ellipsoid required.

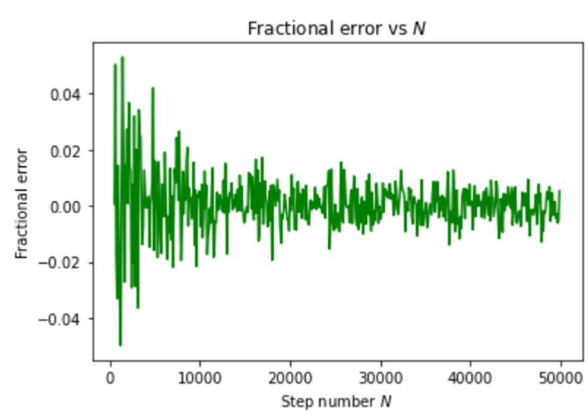


Results & Conclusion:

We can see that as more random points are generated; our estimated volume of the ellipsoid becomes more accurate.



Now, we plot the fractional errors in each estimation vs. the total number of random points as:



It may be considered that this method may be computationally inefficient and with large amount of variables bounded to different constraints, it requires a lot of time and a lot of computations to approximate a solution using this method and the quality of output is as good as the input and heavily relies on the input parameters & constraints.

However, this method was very easy to implement and it provided a good approximation solution to our sample mathematical problem and can be proved to be a success.

References:

a. https://www.wallstreetmojo.com/monte-carlo-simulation/

b. https://www.statlect.com/asymptotic-theory/importance-sampling

c. https://github.com/paniash/p342-lab/tree/master/project/monte-carlo

d. https://ib.berkeley.edu/labs/slatkin/eriq/classes/guest_lect/mc_lecture_notes.pdf

e. https://towardsdatascience.com/monte-carlo-method-explained-8635edf2cf58