Monte Carlo Integration Report

Ravish Kamath: 213893664

24 April, 2023

Introduction

In this report we will be looking at a key problem in the field of statistics, which is solving integrals, or the area under a curve. Sometimes with complicated functions, it may not be advisable to use regular calculus methods to solve integrals. What we can do is use Monte Carlo methods as another approach to solve such problems. In this report we will be looking at 3 types of Monte Carlo Integration, which are:

- Naive Monte Carlo
- Importance Sampling Monte Carlo Approach
- Stratified Sampling Monte Carlo Approach

The Naive Monte Carlo method is the most common method for estimating the expected value while the importance sampling and stratified sampling are methods to reduce our variance. We will look at how each method behaves with 3 separate functions, and do a side by side comparison through visual plots. Depending on the function, one method may be a better choice than the other two. Let us explore such occurrences.

Example 1

Goal

Integrate from $[0, 1]$ using Naive Monte Carlo:

$$\int_0^1 \sqrt{x^3 + \sqrt{x}} \tag{1}$$

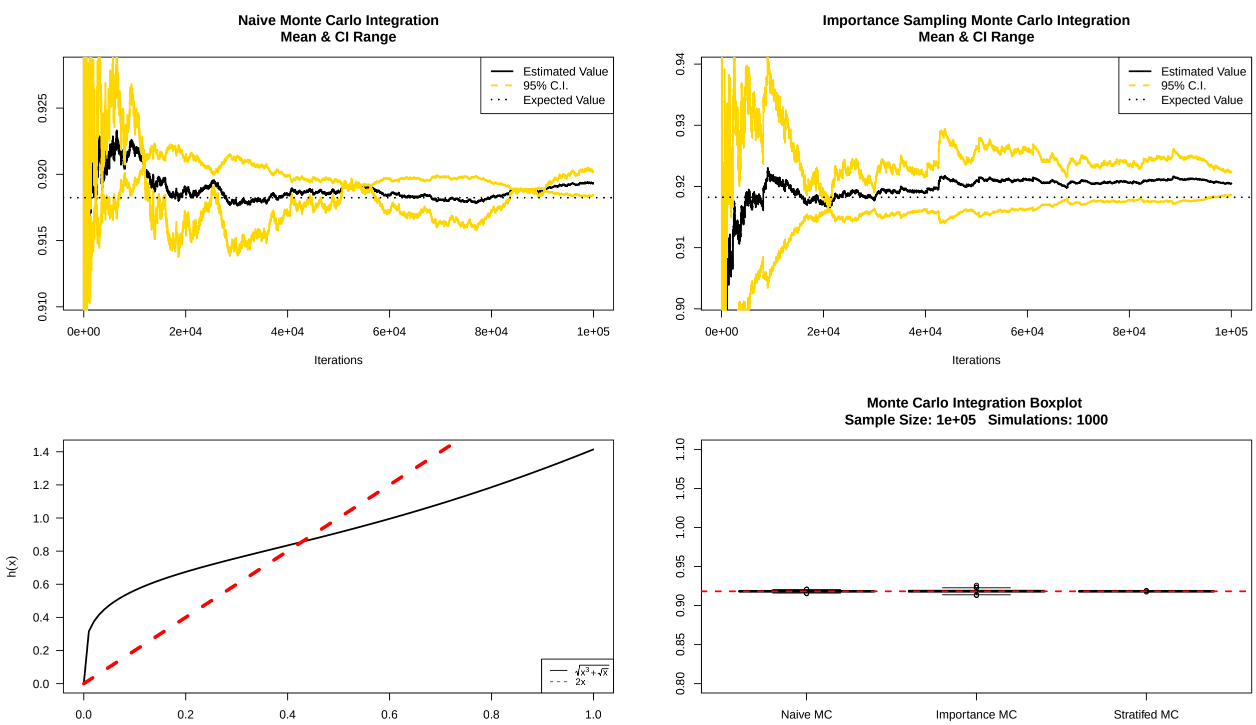

Figure 1: (1) Function Plot Convergence Test and Boxplot Comparison

Starting from the top left of Figure 1 in the previous page, we see the convergence test of our Naive Monte Carlo method. Up until the 20000, we can see heavy change in our estimated value plus a huge gap of error in our confidence interval, however as the iterations keep getting bigger, we can see that the errors tend to reduce towards to the final 100000 iteration. This suggests that it is a good estimation for our expected value, however the error could be a justified problem. Over to the next plot on the right is another convergence test for the importance sampling method. In this case, I decided to use the function $2x$ from $[0,1]$ as my density g. We can clearly see that if we do not pick a good density g, our errors are not getting any smaller, but rather getting larger than the Naive MC method. A better choice of our density g is needed to have a better estimation. We shall see in the next couple examples, a better decision on the g density. On the bottom left Figure, we can see how the original function is plotted as well as our g density for the importance sampling in the red dashes. Unfortunately we do not have a plot to simulate the convergence for the stratified method, however I decided to run a 1000 replication of each of the 3 methods with a sample of 100,000. We can see through the box plots that the stratified method is performing as well as our Naive method, but the importance sampling has greater variability.Looking at Table 1 below, we can see the errors for each of the methods. Based on the result, it looks like Naive MC method is the safest and most accurate option to solve (1) integral. Let us look at other examples, where the other methods may out shine the Naive MC.

Table 1: (1) MC Results for a Single Replication

	Estimate Value	Standard Error
Naive Monte Carlo	0.9192994	7.0e-07
Importance Sampling	0.9204672	3.2e-06
Stratified Sampling	0.9192994	3.2e-06

Example 2

Goal

Integrate from $[1,5]$ using Naive and Importance Sampling:

$$\int_0^6 e^{-5(x-3)^4}\,dx \tag{2}$$

Starting from the top left of Figure 2, although the convergence rate looks really good for the Naive Method, do take note of the y-axis length compared to the right plot. The range in both y-axis indicate that there is more variance occurring in the Naive Method, as compared to Importance Sampling. In the Importance method, I decided to use a normal distribution with a location at 3 and standard deviation of 1. This is best shown on the function plot in red. The reason for picking this particular distribution is to make sure we are sampling more around the center of $x = 3$ while the tails of the desired function have less points. Once again we do not have a particular plot for the stratified. However when performing 1000 replications of each method with sample size 100,000, we can see that the stratified method (by using 8 stratums) outstrips the other two methods. Looking at Table 2, the errors of both the importance and stratified method are significantly smaller than the Naive method.

Table 2: (2) MC Results for a Single Replication

	Estimate Value	Standard Error
Naive Monte Carlo	1.202910	2.60e-05
Importance Sampling	1.215021	1.23e-05
Stratified Sampling	1.202910	1.23e-05

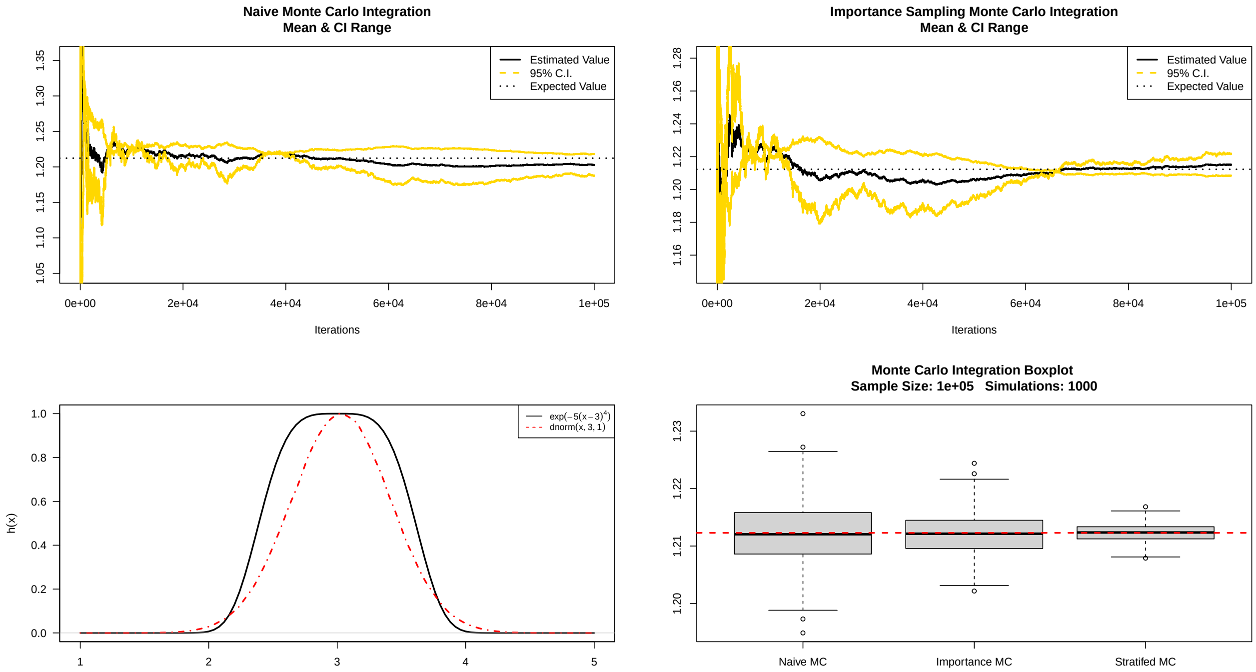

Figure 2: (2) Convergence Test

In this final example we will take a closer look at the importance sampling to find the best density that will reduce our variance.

Example 3

Goal

Integrate from $[0, 1]$ using all three Monte Carlo methods:

$$\int_0^1 \exp\left\{\frac{-x}{1+x^2}\right\} \tag{3}$$

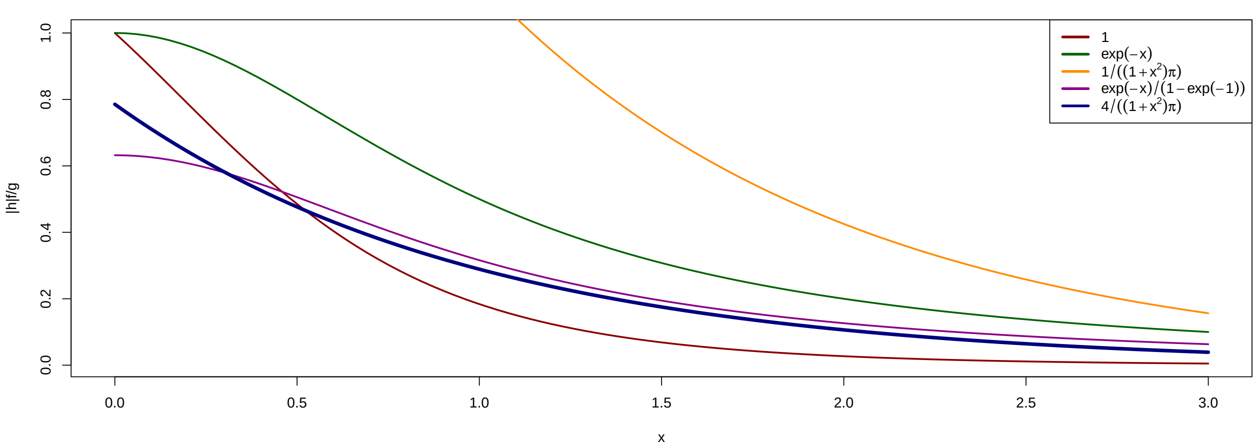

Figure 3: (3) Potential g densitys

Table 3: (3) MC Results for a Single Replication

	Estimate Value	Standard Error
Naive Monte Carlo	0.5255323	6e-07
Importance Sampling	0.5247987	1e-07
Stratified Sampling	0.5255323	1e-07

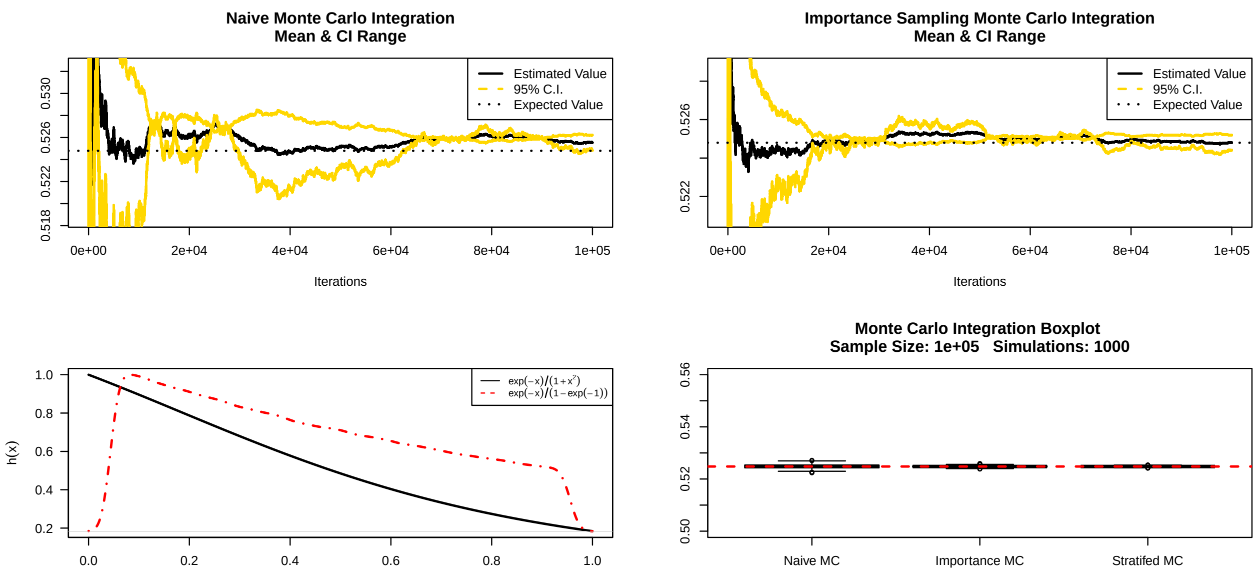

Figure 4: (6) Convergence Plots and Boxplot

As seen in Figure 4 at the top, we have the two convergence plots for the Naive and Importance Sampling. We see that there is large deviation in the confidence intervals for the Naive method. Meanwhile, the importance sampling has far less variation. Let us dive deeper into the cause for that. In order to judge whether your density g is a good choice, we can take the function $\frac{|h|f}{g}$ to be as constant as possible. Whichever is the most constant (flat line on a plot) will be the most desirable option for the importance sampling. Taking a look at Figure 3, we can see the various densities plotted. Although some did very well, the one in particular that looked the most constant was the density $\exp\{-x/(1-\exp\{-1\}\}$. Hence we picked this as the density to get our random variables and solve the estimation of the expected value. On the bottom left plot in Figure 4, we can see that the target function plot as well as the g density we chose. Finally we have a box plot of all 3 methods. We can see that the stratified, as well as the importance sampling, seem to be the best methods for getting our estimated expected value for (3).

Conclusion

As we saw in the 3 examples, each method works well depending on several factors. When looking at importance sampling, we saw that if the density g is not the right choice, we end up with worse results. In terms of stratified sampling, generally the method has very little negative attributes other than the fact that the running time may be longer than the other two. Also the implementation of stratified sampling may be more complicated than the other methods. Finally, if the other methods fail, or are too complicated and variance in the estimate is lenient, then the Naive Monte Carlo method is a perfectly fine approach to solve integrals.