Monte Carlo Integration

In this study, by implementing the regular Monte Carlo, importance sampling, and stratified sampling with both optimal and non-optimal approaches, the following function was used to estimate I.

$$\varphi(x) = (4 - x^2)e^{(3-x^2)}\ ;\ \ x \in [0,1]$$

Hence, by taking the integral of 0 to 1, we can calculate the theoretical value of I, which is:

$$\mathrm{I} \ = \ \int_0^1 \varphi(x)\, dx \ = \ 56.19551\ldots$$

The results of the methods will be both numerically and visually demonstrated in the next part.

1.1) Regular Monte Carlo

The regular Monte Carlo method, known for its simplicity, was executed as a first approach. This method is a straightforward implication to give us an approximate idea of the estimated result. Without large n values, it lacks precision. Also, as the functions get more complicated, they can produce a high variance. To give a numeric understanding of the regular Monte Carlo for estimation, the method is run once with **n=1000**, which resulted in the following:

$$\hat{\mathrm{I}} = 55.47978$$

$$\mathrm{Vâr}(\hat{\mathrm{I}}) = 0.352643$$

1.2) Importance Sampling

Importance Sampling is a tricky technique that can either increase the precision by order of magnitude or decrease it by order of magnitude. Thus, one needs to think mathematically before implementing it. The rule of thumb is to make the $\varphi(x)$ contant as much as possible. Since

$$(4 - x^2) = \ (2 - x)(2 + x)$$

Using neither of the parts could possibly yield more precision. However, we will require the part to be PDF, which means the integral should sum up to **1**.

$$\int_0^1 (2 - x)\, dx = \frac{3}{2}$$

$$\rightarrow \mathrm{g(x)} = (2 - x)\frac{2}{3}$$

Therefore, by integrating the **g(x),** the new equation is now:

$$\mathrm{I} \ = \ \int_0^1 \frac{\varphi(x)}{g(x)} g(x)\, dx$$

The inverse cdf theorem is used to create a random generator of x where x follows a g distribution.

$$x = -\sqrt{4 - 3u} + 2; U \sim Uniform[0,1]$$

Again, to understand the difference briefly, with **n = 1000**, I is estimated.

$$\hat{I} = 56.60613$$

$$\text{Vâr}(\hat{I}) = 0.07573875$$

1.3) Stratified Sampling

With importance sampling, there is an indication of improvement. And, as we have theoretically proved in class, we guarantee better or equal precision by implementing stratified sampling. Hence, we would expect an improvement during the simulation study as each method is implemented.

Moreover, stratified sampling helps with precision by dividing the defined intervals into subsections.

$$S_1 U S_2 U \dots U S_m = S;\ S = [0,1]$$

For easier calculations, the following sample sizes are usually chosen.

$$n_i = n * a_i$$

$$where\ a_i = P(X \in S_i)$$

$$g_i(x) = \frac{g(x)}{a_i} I_{[a_j, b_j]}(x)$$

Therefore, the following results are calculated using the **n = 100 & m = 10** once with the above sample size allocation formula.

Stratified Sampling – Regular Monte Carlo

$$\hat{I} = 56.40815$$

$$\text{Vâr}(\hat{I}) = 0.02797143$$

Stratified Sampling – Importance Sampling

$$\hat{I} = 56.28738$$

$$\text{Vâr}(\hat{I}) = 0.008956549$$

Furthermore, although estimated I is not precise for both techniques, there is a dramatic decrease in the estimated variance. To decrease even further, the optimized sample allocation formula can be used, which is:

$$n_i = n * \frac{a_i \sigma_i}{\sum_{j=0}^{m} a_j \sigma_j}$$

The challenging part about the optimization is that the standard deviation is unknown. Thus, by first using the $n_i = n * a_i$ we have estimated the standard deviation and reallocated the sample among each subsection. As a result, the following estimations are found.

Stratified Sampling – Regular Monte Carlo

$$\hat{I} = 56.26408$$

$$\hat{Var}(\hat{I}) = 0.02725889$$

Stratified Sampling – Importance Sampling

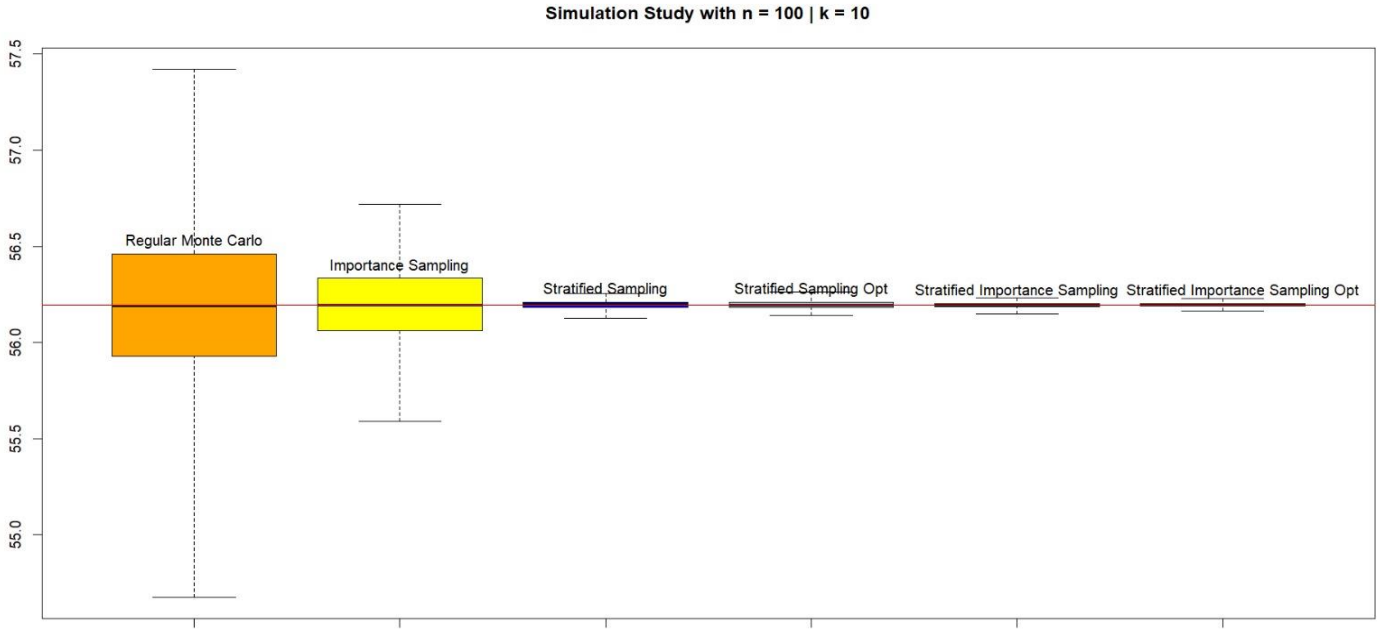
$$\hat{I} = 56.33721$$

$$\hat{Var}(\hat{I}) = 0.006563842$$

We can observe a slight improvement as we proceed with each process. The next section will discuss visual comparison of the methods in more depth.

2.1) Simulation Study

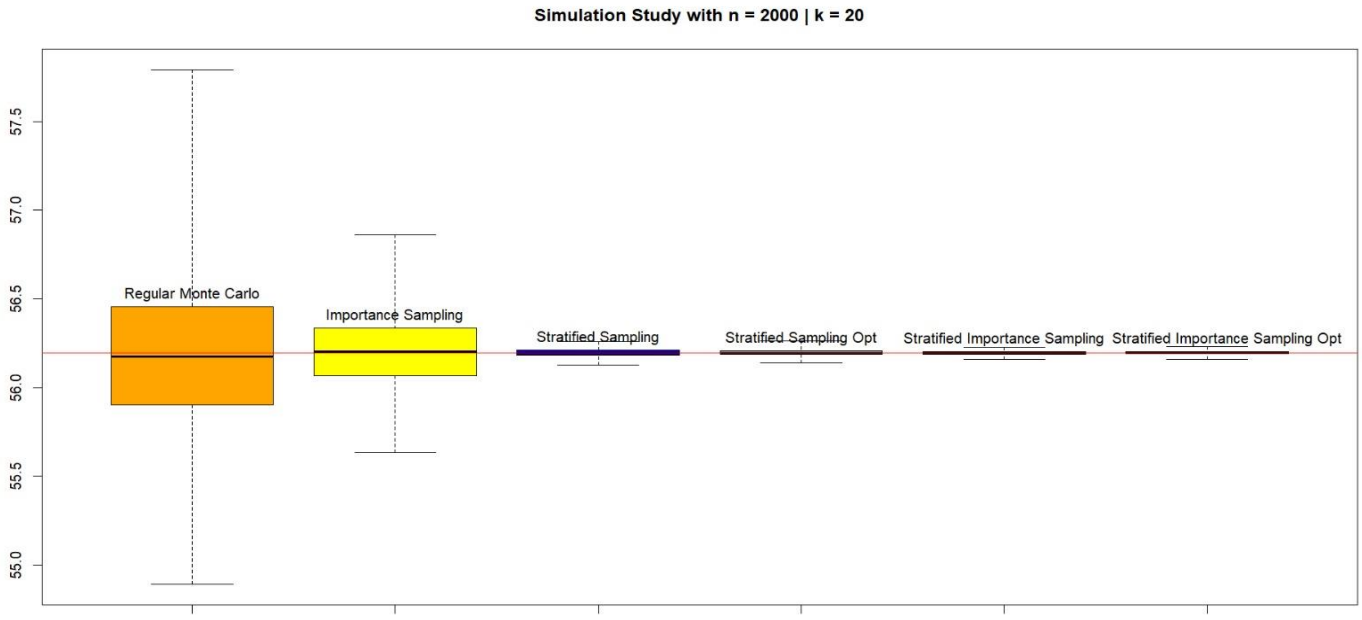
As there is a random component, even running the methods several times would not provide a clear, trustful outcome. Hence, a simulation study is a great way to examine the underlying procedures. By taking **B = 1000**, meaning running each method 1000 times and inserting it into a vector, we plot the box plot of each vector. For **n = 100 & k = 100**, the results are as follows:



The plot supports our intuition of **smaller variance** as the method proceeds. Our chosen **g** function appears to decrease the variation dramatically as we go from **regular Monte Carlo** to **Importance Sampling**. Also, the stratification method substantially increases the precision, and

the optimized version has a small but effective impact on the estimation. Hence, we can observe that **regular Monte** Carlo has the higher result deviation, whereas **stratified importance optimized sampling** has the smallest. Lastly, all of the implementations appear to constitute the theoretical value in the middle successfully.

Now let's observe under the scope of greater **n** and **k**, **2000** and **20**, respectively.



The gap for the regular Monte Carlo tilted upwards slightly. However, other than that, there is no dramatic change as we go from **n = 100 & k = 10** to **n = 2000 & k = 20.**

In conclusion, by choosing the **g** distribution carefully, we have successfully decreased the variance and observed the significant impact that stratified sampling can have with simple coding. However, particularly for this example, going from **stratified sampling** to **optimized stratified sampling** did not have a major improvement. Therefore, the optimization of the sample size allocation might not be implemented in the cases that are harder to achieve.