Stochastic Simulation: Assignment 1

C. Carissimo 10890661

November 19, 2017

1 Introduction

This report summarizes the results from assignment 1 for the course Stochastic Simulation at the University of Amsterdam. The assignment required the exploration of the area of the Mandelbrot set, a mathematical set calculated by the iteration of a complex function over different complex numbers in the complex plane. The Mandelbrot set is defined as the set of all numbers c such that $f(z) = z^2 + c$ does not diverge when iterated from $z = 0$. The area of the Mandelbrot set will be calculated by Monte Carlo sampling, and different Monte Carlo methods will be compared in terms of their convergence to the real area of the Mandelbrot set.

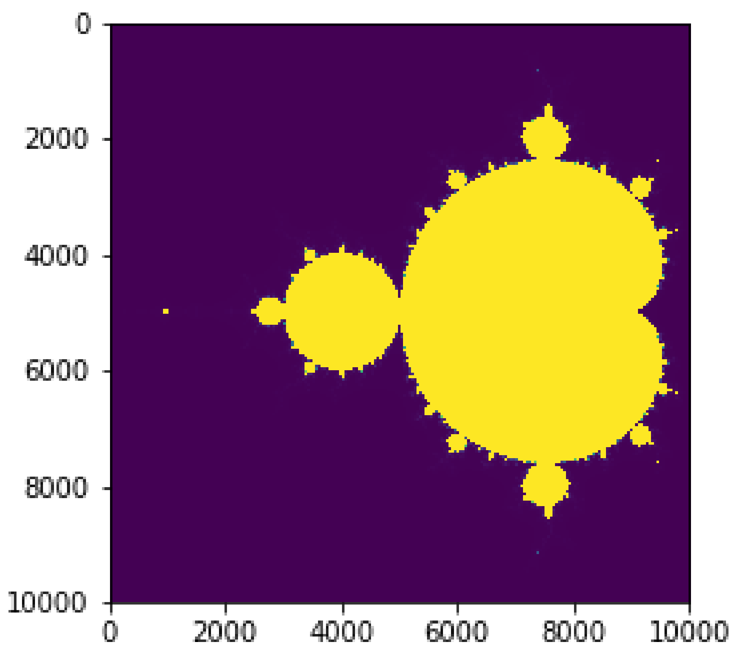

Figure 1: Mandelbrot set, where points in yellow are the points contained within the set, and thus the points that do not diverge when iterated.

2 Methods

A representation of Mandelbrot set can easily be computed and visualized with a few lines of code. The fundamental limiting factors in calculating the Mandelbrot set are:

1. a range of complex numbers c, which we will define as $c \in \mathbb{C}$, $-2 < Re(c) < 0.5$, $-1.25 < Im(c) < 1.25$.
2. the density of the range, which we will define at 10,000 points per axis, for a total of 100,000,000 points $\in \mathbb{C}$.
3. and the maximum number of iterations for each point, which we will take as 2048.

These decisions were made based on the time required to compute the Mandelbrot set. This set of constraints will be taken as the best possible computation of the Mandelbrot set for this assignment. The Mandelbrot set will be calculated discretely and saved in a 2 dimensional, $10000 * 10000$ array. Each element of the array $a_{i,j}$ will reflect a number $c \in \mathbb{C}$ in our defined range of complex numbers. The best possible approximation of the area of the Mandelbrot set given this computation is defined as the fraction of all $a_{i,j} = 2048$ divided by the total number of points, multiplied by the area of the complex plane used for computation. This area is equal to 1.5080599375. This best estimate of the area of the Mandelbrot set will be used as the benchmark to assess the accuracy of all the other Monte Carlo methods, and we will call this area A_M. The Monte Carlo methods that will be tested are pure random sampling, latin hypercube sampling and orthogonal sampling.

2.1 Estimation

All three methods are hit and miss Monte Carlo integration methods. Random points are drawn from a joint uniform distributions and are checked against the Mandelbrot set. The samples can either hit or miss the Mandelbrot set. H is the random variable that counts the number of hits for n samples, where $H_i = 1$ if sample i hits and $H_i = 0$ if sample i misses.

$$H = \sum_{i=1}^{n} H_i \tag{1}$$

The expected value of H_i is p, where p is the probability that a random sample will hit $(P(h = 1))$. Therefore, the expected value of H is:

$$E[H] = E[\sum_{i=1}^{n} H_i] = \sum_{i=1}^{n} E[H_i] = np \tag{2}$$

The variance of H is:

$$Var[H] = \sum_{i=1}^{n} Var(H_i) = np(1-p) \tag{3}$$

When we calculate the area of the Mandelbrot set, we use an estimate of p determined by our estimated random variable H divided by n. In terms of p, the area of the Mandelbrot set is:

$$A_M = P[h = 1]A_{square} = pA_{square} \tag{4}$$

To estimate H we take use repeated sampling and take the mean of our samples $\bar{H}$, which is an unbiased estimator of H. For large enough n samples ($n > 100$), $E[\bar{H}] = p$ as $\bar{H}$ is approximately normally distributed with mean p and standard deviation $\sigma/\sqrt{n}$. With these properties of our random variable H, we can define confidence intervals for an alpha level $\alpha = 0.05$. The probability that p lies within boundaries defined by our alpha levels is:

$$P[\bar{H} - 1.96 * \frac{\sigma}{\sqrt{n}} < p < \bar{H} + 1.96 * \frac{\sigma}{\sqrt{n}}] \approx 1 - \alpha \tag{5}$$

According to Slutsky's theorem we can substitute σ for our sample standard deviation S, which we can calculate from our sample. The confidence intervals become:

$$P[\bar{H} - 1.96 * \frac{S}{\sqrt{n}} < p < \bar{H} + 1.96 * \frac{S}{\sqrt{n}}] \approx 1 - \alpha \tag{6}$$

2.2 Pure Random Sampling

Integration by pure random sampling means that two random numbers in the appropriate complex and real ranges will be generated for each sample and checked against the Mandelbrot set. Figure[] shows the convergence of the estimated area as a function of iterations.

2.3 Latin Hypercube Sampling

The latin hypercube model first requires the subdivision of our space into a number of subspaces equal to the number of samples you wish to take. Then, you sample randomly a random point from all subspaces with two constraints on the sampling grid: that no two subspaces from which you sample share the same column, and equally that they do not share the same row.

2.4 Orthogonal Sampling

Orthogonal sampling takes latin hypercube sampling (LHS) a step further. Just as in LHS, the space is subdivided into subspaces with the same row and column constraints on the sampling. In addition, the subspaces are also subdivided into subspaces and the same constraints are applied to the sampling within that subspace. In some sense, orthogonal sampling is latin hypercubes inside latin hypercubes.

3 Pure Random Sampling

With pure random sampling we experiment with the rate of convergence of the estimated area of the Mandelbrot set as a function of iterations. Figure[] shows the convergence of the area as a function of iterations. We begin the iterations at about 100 iterations. The best estimation of the are given pure random sampling will be taken as the estimation with the most iterations, which was 2000 iterations. The mean $\bar{H}$ for that estimation is 1.509025.

For the pure random sampling the confidence intervals:

$$P[1.50343571 < \bar{H} < 1.51461429] \approx 1 - \alpha = 0.05 \tag{7}$$

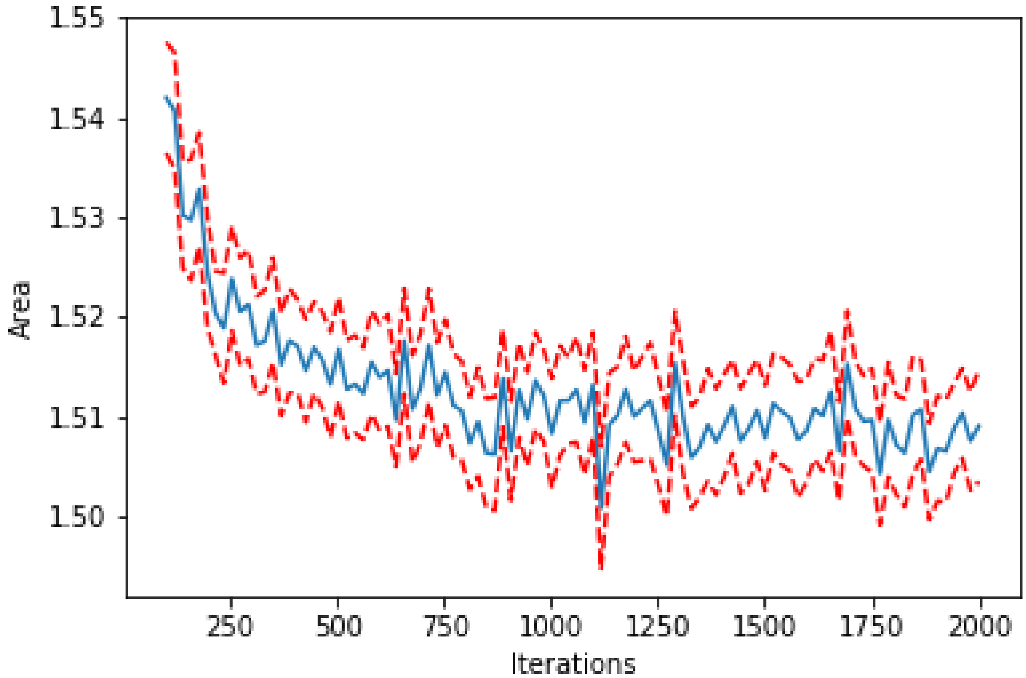

Figure 2: Plot of the estimated area of the Mandelbrot set as a function of iterations for Pure Random Sampling.

4 Latin Hypercube Sampling

The best estimation given by pure random sampling will be taken as the estimation with the most iterations, which was 2000 iterations. The mean $\bar{H}$ for that estimation is 1.5062.

For the pure random sampling the confidence intervals:

$$P[1.50288671 < \bar{H} < 1.50951329] \approx 1 - \alpha = 0.05 \tag{8}$$

5 Orthogonal Sampling

The best estimation given by pure random sampling will be taken as the estimation with the most iterations, which was 2000 iterations. The mean for that estimation is 1.5062. For the pure random sampling the confidence intervals:

$$P[1.50288671 < \bar{H} < 1.50951329] \approx 1 - \alpha \tag{9}$$

6 Antithetic Variables

Latin hypercube sampling was the sampling method that proved to have the least sample standard deviation, and thus the smallest confidence interval. Building upon LHS to obtain

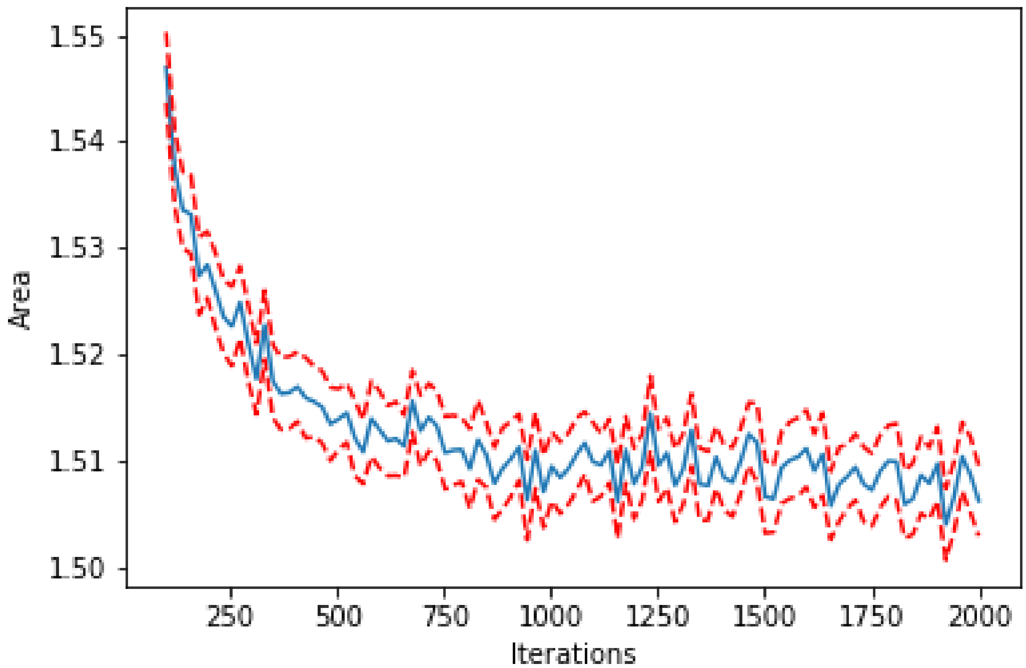

Figure 3: Plot of the estimated area of the Mandelbrot set as a function of iterations for Latin Hypercube Sampling.

a lower sample variance and a more accurate estimation we tried to implement Antithetic Variables to reduce the variance further. Antithetic variables works as follows: we sample a point randomly, with a random x and y coordinate, and then we generate a new random point by taking our random x and y coordinates and subtracting them from the maximum ranges in the x and y direction. This leaves us with a new random point in the x, y plane. We can test for the inclusion of this point in the Mandelbrot set as well. The usefulness of this application, is that these two points are correlated, so when we calculate the join variance the correlation between these two points will influence our sample variance. The significant relationship between these random points is that they are likely to be negatively correlated, which would lead the correlation term to be negative, and could reduce the sample variance.

$$P[1.50653329 < \bar{H} < 1.51122921] \approx 1 - \alpha = 0.05 \tag{10}$$

7 Conclusion

In Conclusion, this final attempt proved to be the most accurate estimation of the area of the Mandelbrot set. Surprising was the fact that orthogonal sampling did not significantly improve the accuracy of estimation of the Mandelbrot set. Not surprising was the fact that pure random sampling had the least accuracy.

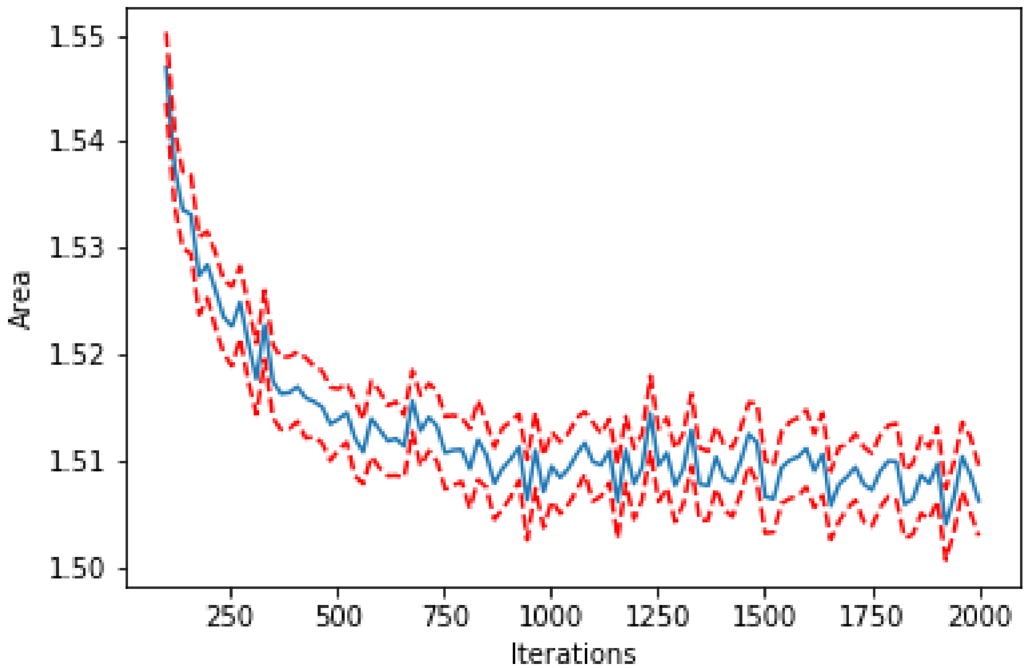

Figure 4: Plot of the estimated area of the Mandelbrot set as a function of iterations for Orthogonal Sampling.

References

[1] Ross, S. M. (2013). Simulation. Prentice Hall PTR.

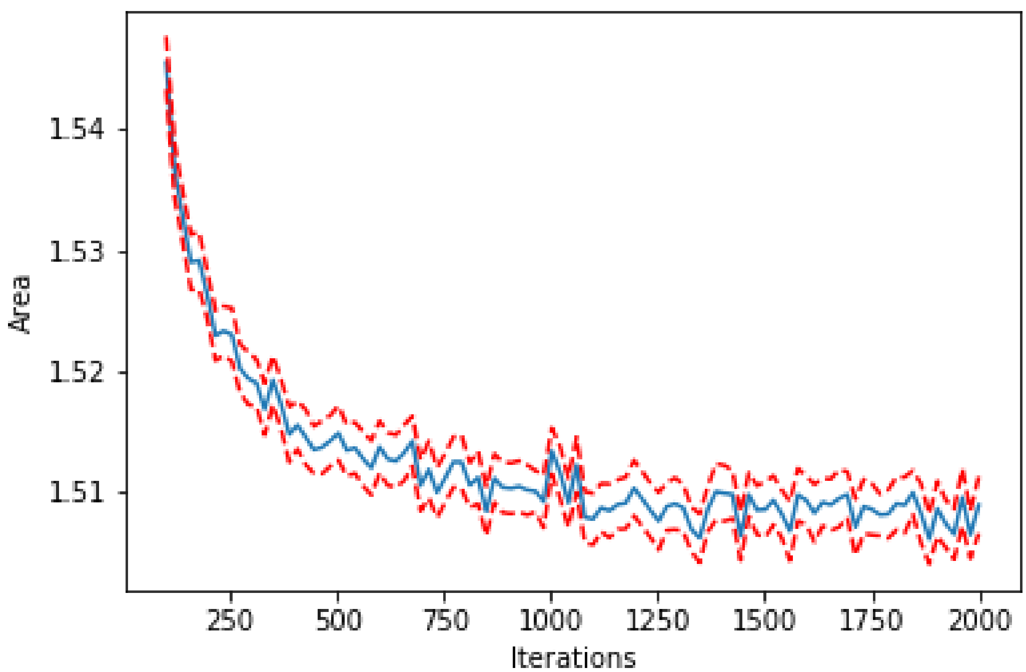

Figure 5: Plot of the estimated area of the Mandelbrot set as a function of iterations for Latin Hypercube Sampling with the inclusion of antithetic variables.