University of Amsterdam

Faculty of Science,
Amsterdam, The Netherlands

Stochastic Simulation

The Mandelbrot Set

Authors:

R. A. Peerboom
10791523
roman.peerboom@student.uva.nl

F. M. C. de Vries
11710799
floris.de.vries@student.uva.nl

Abstract

This paper researched the convergence behaviour of the area of the Mandelbrot set and the potential usage of several sampling techniques to estimate the area more precisely. Consistently, errors in the estimation of the set decreased until around 200 iterations. After that it stayed almost the same across the different sampling methods. Latin hypercube sampling successfully decreased the variance compared to random sampling, orthogonal sampling reduced the variance even further. Orthogonal sampling combined with a control variate that consisted of the fraction of bounded sampled points returned the lowest variance.

November 17, 2020

Contents

1 Introduction

The Mandelbrot set is the total set of complex numbers for which the following function remains bounded:

$$f(z_{n+1}) = z_n^2 + c \tag{1}$$

In this function, z_0 starts at 0 and c is a complex number. A complex number is the sum of a real number and an imaginary part. This imaginary part consists of the imaginary unit, the square root of -1, multiplied by a factor (Heath, 2018). For different values of c, $f(z_n)$ behaves differently. For some values, it will remain in orbit, for others it tends to a fixed point and for a last group it goes to infinity. The formula $f(z_n)$ is said to remain bounded when it does not tend to infinity. So, the Mandelbrot set divides the complex numbers between those who tend to infinity and those who do not (Dolotin et al., 2006). This can be depicted in a, quite beautiful, image in which the pixels with complex numbers that tend to infinity have a different colour than those who do not. The pixels that are dark in this picture are the ones that belong to the Mandelbrot set.

Apart from being an interesting visual, it is also interesting mathematically. Thus far, no one has been able to prove what the exact size of the set is (Mitchell, 2001). In this paper, we will try to get an as precise as possible estimate of the area of the Mandelbrot set. We are also interested in the convergence behaviour of the estimates as the number of iterations and samples increase. We will test this using several different sampling techniques. Whether or not the convergence behaviour happens is difficult to determine as we do not know the exact area. Because of this and because we work with stochastic processes, we have to work with confidence intervals. To reduce the confidence interval it is necessary to reduce the variance. So, the goal of this research is to find sampling techniques that estimate the area of the Mandelbrot set with small variances.

2 Methods

2.1 Mandelbrot set

A first step to a better understanding of the Mandelbrot set is to create an image of it. We did this by for every point on a plane (54.000 x 36.000 pixels) converting the pixel coordinate to a complex value, where the x-coordinate is the real part and the y-coordinate the imaginary part.

For every iteration, this complex number is transformed by the previously mention function $f(z_{n+1}) = z_n^2 + c$. Every iteration, the result of this is compared with the threshold for infinity, which is 2 in our case. Although this does not seem high, it is high enough, because if the value of z exceeds 2 once, it will always tend to infinity (Subathra and Jayalalitha, 2019). If the value of z exceeds the threshold during one of the iterations, the number of iterations that this took is saved. The colour of the pixel from which the complex number was derived is then changed according to the number of iterations. The more iterations, the darker the pixel. If after a thousand iterations the threshold has not been reached, the relevant pixel will be completely black.

2.2 Sampling Methods

2.2.1 Theory

The area of the Mandelbrot set can be determined by calculating every complex value in a plane, but this uses a lot of computational power. A solution for this problem is sampling (quasi-)random points and estimating the area based on these samples. This type of approach is called a Monte Carlo method.

In this paper, we will compare different sampling methods by using them to determine the area of the Mandelbrot set with changing amounts of samples and iterations. The three methods of interest are pure random sampling, Latin hypercube sampling, and orthogonal sampling. Pure random sampling is choosing totally random numbers, without taking in account the results of the previous or forthcoming samples. Latin hypercube is a more ordered way of sampling. The main idea is that the samples are more evenly distributed among the n-dimensional search space. The variance of this sampling method is either lower or equal to that of random sampling. In case of a monotone function it is always lower(Ross, 2013), but that is not the case in this paper. Orthogonal sampling is related to Latin hypercube sampling but with one additional constraint. The search space is divided in sub spaces, in our case there are four sub spaces, in which there are the same amount of samples. Literature suggests that Orthogonal sampling provides better uniform coverage than Latin hypercube sampling (Burrage et al., 2015).

2.2.2 Experimental Method

For every sampling method, we will estimate the area of the Mandelbrot set for different values of the number of iterations and two sample sizes. The number of iterations will be all the values starting from 1, with steps of 2 until 999 is reached. Increasing the step size from 1 to 2 was done to reduce the computational power needed. The used sample sizes were 529 and 1024. We chose these sizes because the sample amounts needed to be squares of an integer for orthogonal sampling. In order to rightfully compare the sampling methods with each other we kept the sample sizes consistent throughout the process.

Generating random numbers was done with numpy.random(). Because the search space is two-dimensional (x and y), the stratifying sampling methods will be designed to take exactly one sample for every row and column. For every sampled point, it was calculated whether or not the point belongs to the Mandelbrot set, i.e. if it remains bounded. The area of the Mandelbrot set was then calculated by multiplying the fraction of bounded points with the total area wherein the points were sampled, which was 6.

Apart from comparing the quality of the results for different amounts of iterations and sample sizes, we also compared them extensively for the one specific case in which $i = 1000$ and $s = 1024$. This was simulated 1000 times. This was done to get a more comprehensive view of the difference in quality between the methods with high values for i and s. The variances of the different methods were compared using a Levene's test. Levene's test was chosen because it is one of the most robust tests for variances and it has a relatively high powerLim and Loh, 1996.

2.3 Control Variates

There are other ways than using different sampling methods to improve the estimations of the Mandelbrot set. One of those ways is to reduce the variance by using control variates. We will try to further improve our orthoganal sampling method by introducing the amount of sampled points that were bound by infinity as control variate. This amount is denoted by t and is transformed as follows:

$$f(t_i) = \frac{1}{1 + t_i} \tag{2}$$

The area of the Mandelbrot set, denoted by variable X, is then calculated as follows:

$$X^* = X_i + c(f(t_i) - \mu_t) \tag{3}$$

The value for c that decreases the variance of X the most is given by this formula:

$$c = -\frac{COV(X,t)}{VAR(t)} \quad (4)$$

Using these formulas, the variance of X should decrease by $100 * Corr(X,t)^2$ percent (Ross, 2013). We chose the amount of bounded points as control variate because it has a high correlation with X, so a significant reduction of the variance can be expected.

3 Results

3.1 Plotting the Mandelbrot set

Figure 1 displays a visualization of the Mandelbrot set. The black pixels correspond with complex values that are bounded, and thus part of the Mandelbrot set. The shades of the other pixels depends on the amount of iterations needed to reach the threshold for infinity. The fact shades of red are used is trivial, it could have been any other color.

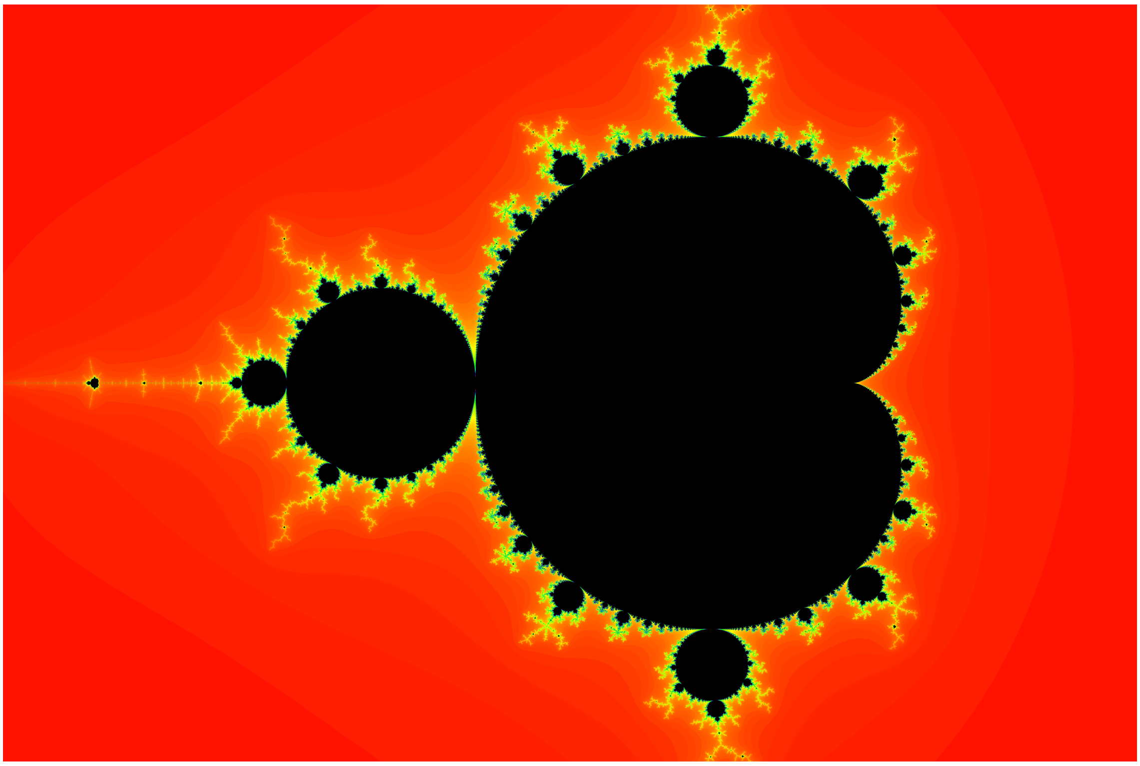

Figure 1: A visualization of the Mandelbrot set.

3.2 The influence of the maximum number of iterations

To investigate the influence of the maximum number of iterations for each point to reach the threshold of infinity, the error of the Mandelbrot area was plotted as a function of this maximum number of iterations (j). The error is defined as the difference of each $A_{j,s}$ minus $A_{i,s}$, where i is the maximum value of j. Two different samples were used to have some understanding of the influence of this parameter.

3.2.1 Random sampling

The error of the estimation of the Mandelbrot set as a function of j obtained with random sampling is displayed in Figure 2. For j from 0 to 200 the error decreases from a large positive error towards zero. After 200 there does not seem to be a strong continuing decrease of the error.

The mean of the smaller samples size seems to fluctuate more around zero and also the confidence interval of the smaller sample size is larger compared to the bigger sample size. This means that the smaller sample size has a larger variance, which is what we would expect, since the variance scales with $\frac{1}{\sqrt{n}}$.

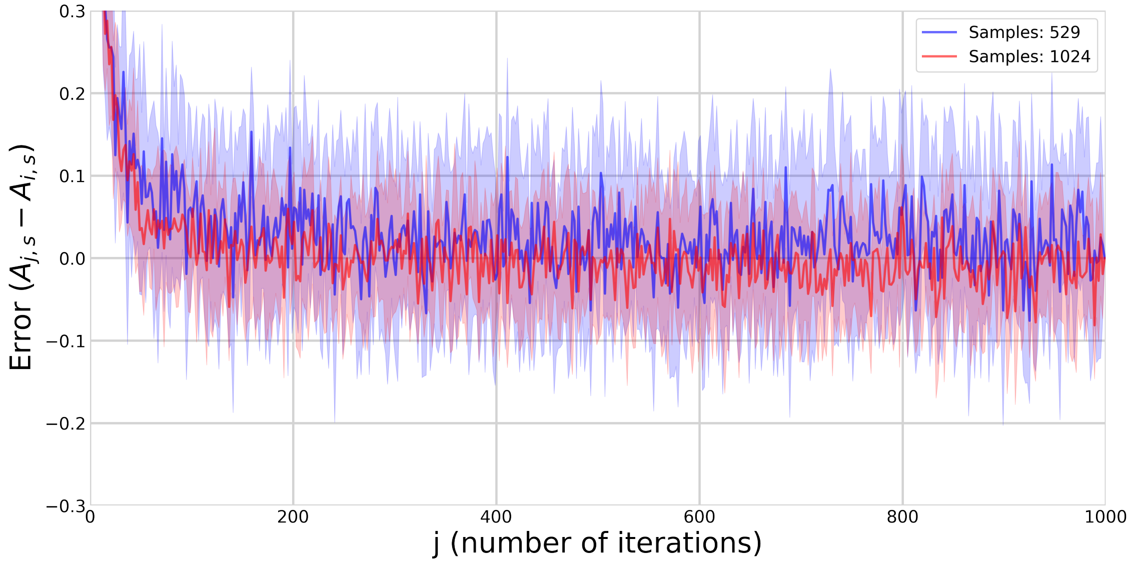

Figure 2: The mean error with a 95% confidence interval from ten simulations of the area of the Mandelbrot set estimated as function of j (A_j) with random sampling for two samples sizes, where the error is calculated as the difference of $A_{j,s} - A_{i,s}$, for i being the largest j, being is 999 in this case ($A_{i,529} = 1.494$, $A_{i,1024} = 1.522$).

3.2.2 Latin hypercube & orthogonal sampling

To investigate the influence of different sampling methods, the sampling error was also investigated for Latin hypercube sampling (Figure 3) and orthogonal sampling (4). The same two findings as in Figure 2 are present. Firstly, the error seems to decrease for values of j below 200 and after that point stays relatively constant. Secondly, as expected, the fluctuations of the smaller sample are again larger, indicated by greater fluctuations of the mean and larger confidence intervals.

In Figure 3 the error for the group with a sample size of 1024 is for almost every j above zero. This can probably be attributed to the fact that $A_{i,1024}$ is 1.489, which is the smallest of all $A_{i,s}$ estimations. Therefore, $A_{j,s} - A_{i,s}$ is for almost every j above zero. This phenomenon is also present to a lesser extend in the other groups, since $A_{i,s}$ will never be precisely equal to the real Mandelbrot area.

If we compare the confidence intervals of the different sampling methods, it is visible that random sampling gives the largest confidence interval and orthogonal the smallest. This is as expected, since Latin hypercube sampling takes more homogeneous samples compared to random sampling and therefore has a smaller variance. Orthogonal sampling samples more homogeneously than Latin hypercube and therefore the confidence interval shrinks further.

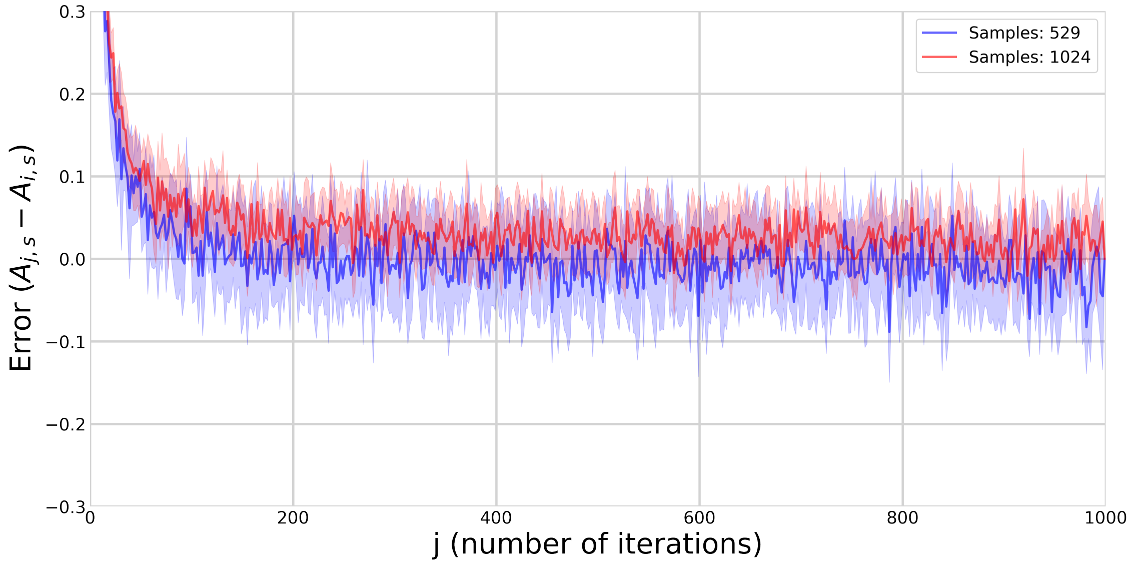

Figure 3: The mean error with a 95% confidence interval from ten simulations of the area of the Mandelbrot set estimated as function of j (A_j) with Latin hypercube sampling for two samples sizes, where the error is calculated as the difference of $A_{j,s} - A_{i,s}$, for i being the largest j, which is 999 in this case ($A_{i,529} = 1.523$, $A_{i,1024} = 1.489$).

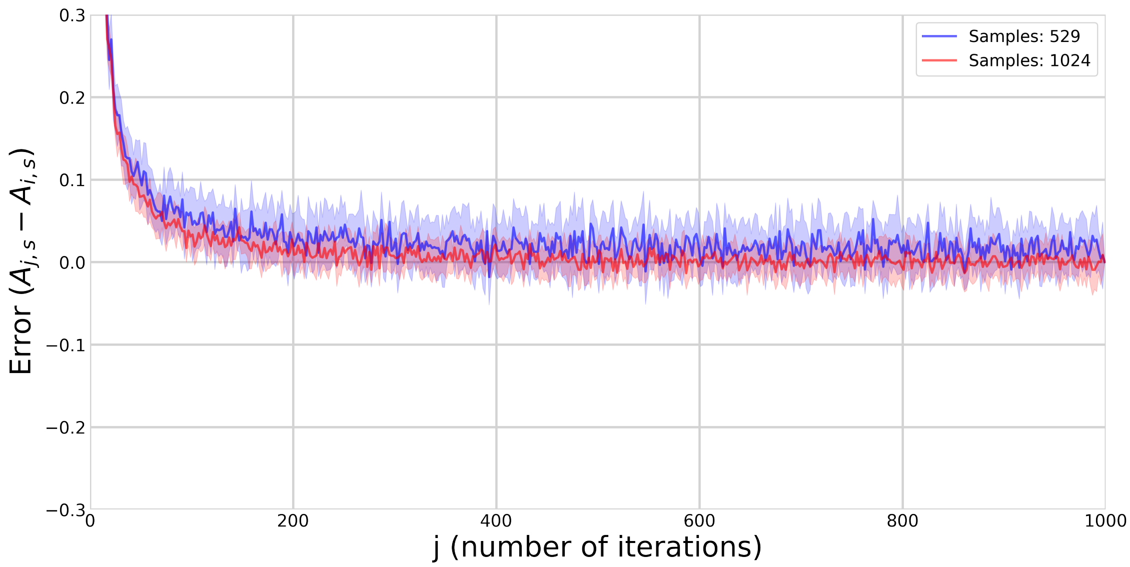

Figure 4: The mean error with a 95% confidence interval from ten simulations of the area of the Mandelbrot set estimated as function of j (A_j) with orthogonal sampling for two samples sizes, where the error is calculated as the difference of $A_{j,s} - A_{i,s}$, for i being the largest j, which is 999 in this case ($A_{i,529} = 1.495$, $A_{i,1024} = 1.511$).

3.3 Estimating the area of the Mandelbrot set for different sampling methods

To compare the quality of the estimation of the Mandelbrot area for different sampling methods, 1000 estimations were done for each sampling method. The maximum number of allowed iterations in these simulations was 1000. For each estimation 1024 points were taken. The results are visualised as box plots in Figure 5. Levene's test was performed to determine if the variance was significantly different. In line with our earlier findings the orthogonal sampling gives a smaller variance than the Latin hypercube sampling ($p < 0.001$), which in turn gives a smaller range compared to random sampling ($p < 0.001$).

To reduce the variance of the estimation with orthogonal sampling even further we tried to combine it with the control variates method. As control variate we chose the fraction of sampled points that did not go to infinity. Because the expected value was unknown, we estimated it in two ways. One way was estimating it using the mean of Mandelbrot area estimated by the earlier performed three sampling methods. This method led to a small, non significant reduction in the variance ($p > 0.99$). This is probably because the estimation of the expected value of the fraction of points that did not go to infinity was not accurate enough.

The second option was to use the best estimation of the Mandelbrot area as described by (Mitchell, 2001) to determine the expected value of the fractions of points that did not go to infinity. This means we used the answer to obtain the answer, and therefore this method is labeled as cheating. This method led to a much bigger and significant reduction in the variance for orthogonal sampling($p < 0.001$).

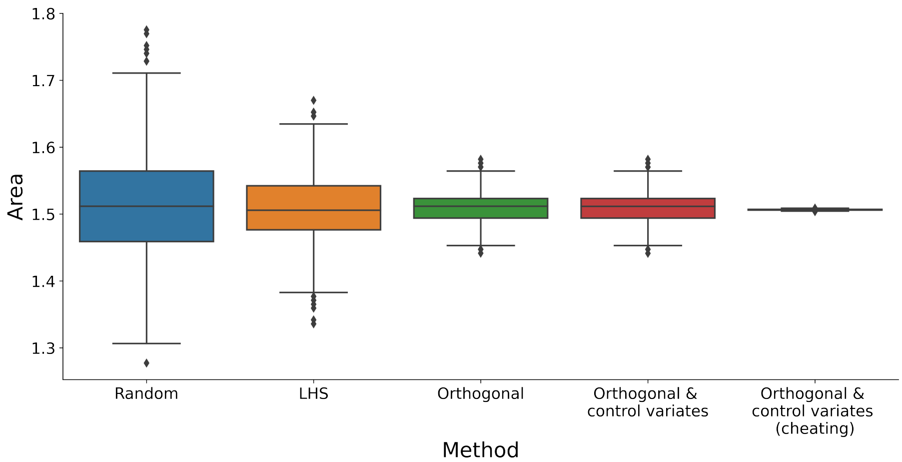

Figure 5: Box plots displaying the estimation of the area of the Mandelbrot set with different (sampling) methods

4 Discussion

To reduce the error of the estimation of the Mandelbrot area the maximum number of iterations was varied. After the maximum number of iterations reached 200, the error did not seem to decrease much further for our two samples sizes and different sampling methods. Possibly, this is because only a small fraction of the points need more iterations than 200 to reach the threshold for infinity and with our current sample sizes of 529 & 1024 this fraction of points is too small to make a significant difference.

To increase the convergence rate of the estimation of the Mandelbrot area we stated earlier that it is necessary to reduce the variance. Significant reduction of the variance was reached with Latin hypercube sampling compared to random sampling, and even more with orthogonal sampling. This confirms the statements posed by earlier researchers, as mentioned in the theory section of this paper.

The addition of control variates to orthogonal samplings showed potentially promising results, in the situation where foreknowledge about the answer was used. It would be interesting to investigate if the control variates method is also useful if there is no knowledge about the answer. To investigate this, it would be necessary to apply this method with a control variate that has an value that correlates with the value for the Mandelbrot area and wherefore the expected value is known, so it does not need to be estimated. A possibility could be the random value that is used to generate the x- or y-value.

Bibliography

Burrage, K., Burrage, P., Donovan, D., & Thompson, B. (2015). Populations of models, experimental designs and coverage of parameter space by latin hypercube and orthogonal sampling. *arXiv preprint arXiv:1502.06559*.

Dolotin, V. et al. (2006). *The universal mandelbrot set: Beginning of the story*. World Scientific.

Heath, M. T. (2018). *Scientific computing: An introductory survey, revised second edition*. SIAM.

Lim, T.-S., & Loh, W.-Y. (1996). A comparison of tests of equality of variances. *Computational Statistics & Data Analysis*, *22*(3), 287–301.

Mitchell, K. (2001). A statistical investigation of the area of the mandelbrot set. https://www.fractalus.com/kerry/articles/area/mandelbrot-area.html

Ross, S. M. (2013). *Simulation*. Academic Press.

Subathra, G., & Jayalalitha, G. (2019). Compactness in mandelbrot sets. *The International journal of analytical and experimental modal analysis*, *11*, 120–125.