Variance Reduction in Area Estimation of the Mandelbrot set

November 17, 2020

Isabelle Brakenhoff
isabelle.brakenhoff@student.uva.nl
11283912
Universiteit van Amsterdam

Jorrim Prins
jorrimprins.prins@student.uva.nl
11038934
Universiteit van Amsterdam

***Abstract*—In this report we estimate the area of the Mandelbrot set using a Monte Carlo approach in combination with a hit-and-miss transform. We investigate convergence properties of the set's area under a varying number of iterations and samples to determine the effect of these variables. Comparisons are made with estimations of the area from literature. In addition, we try to improve convergence by implementing several variance reduction techniques: two stratified sampling methods and importance sampling. We find stratified sampling to be very effective in the reduction of the estimation's variance. Importance sampling also reduces the variance, but at a lower confidence level for some parameter settings.**

I. Introduction

In many cases computers have enabled statisticians to invent new modelling structures and to solve these models without substantial simplification. In recent years, computer power has increased dramatically and access to computers has become easier and more common. With this progression, simulation methods have become increasingly feasible and therefore developed to be more refined. They have turned into a specific way of doing research and they provide another perspective; in some cases simulations even provide a theoretical basis for understanding experimental results (Landau and Binder, 2014).

The Monte Carlo (MC) method is a popular simulation technique that relies on repeated random sampling to estimate numerical quantities, often used to estimate a statistical model of a real system. The idea of MC experiments is based on the use of the Law of Large Numbers (LLN), stating that a sufficiently large random sample of size N allows for theoretical convergence to the expected value of the underlying distribution. MC methods are able to simplify complex models to a basic set of events and simultaneously incorporate the randomness of real processes. An additional advantage of MC is its scalability, and Kroese et al. (2014) note MC's ability to naturally escape local optima and effectively explore the search space due to its random nature.

A widely known mathematical problem is the estimation of the Mandelbrot set's area (Ewing and Schober, 1992; Fisher and Hill, 1993), as an analytical expression of this area is unknown. MC experiments combined with a hit-or-miss transform of the fractal are frequently used to define upper and lower bounds on the area of the set, but computational costs can become disproportionately large. Variance reduction is an effective technique to improve convergence and therefore decrease computational costs, this can be done be e.g. antithetic samples, control variets, importance sampling or stratified sampling.

We will investigate importance sampling and stratified sampling to improve the convergence rate of area calculations of the Mandelbrot set. Beginning with a definition of the fractal and the underlying recursive formula, as well as some theory on MC and our variance reduction techniques. Thereafter, we provide a description of the experiments we perform to analyse the techniques. Finally, we present results of these experiments and conclude with a discussion on the findings of this research.

II. Theoretical background

A. Mandelbrot set

The Mandelbrot set is a set of complex numbers c for which a specific recursive formula does not diverge. Whether a specific c lies within the set can be found by taking a random complex number and iterating through the following formula for a specified numbers of times:

$$f_c(z) = z^2 + c \tag{1}$$

This is a recursive relation when taking $z = 0$ as the starting value and thereafter defining z as f_c of the previous iteration. The complex number under investigation is in the Mandelbrot set when f_c diverges, which is the case when it becomes larger than 2 within the given number of iterations i. It will otherwise surely diverge to infinity, so the iteration can be stopped when $f_c \geq 2$.

Previous research has employed the counting pixel method and the finite escape algorithm described by Andreadis and Karakasidis (2015). Visualisation of the Mandelbrot set can be performed by defining a complex plane with the real components on the x-axis and complex components on the y-axis. Specify the number of points on the plane p, which will be equal to the amount of pixels in the resulting image and perform the iteration from Equation 1 on every point i

times or until it diverges. We define a plane ranging from -2 to 0.5 for the real components and from -1.1 to 1.1 for the complex components, this will picture approximately the complete Mandelbrot set. Figure 1 visualises the starting points that are within the set with the black areas. The values outside of the set are depicted by the number of iterations it takes until their divergence, ranging from white to dark purple.

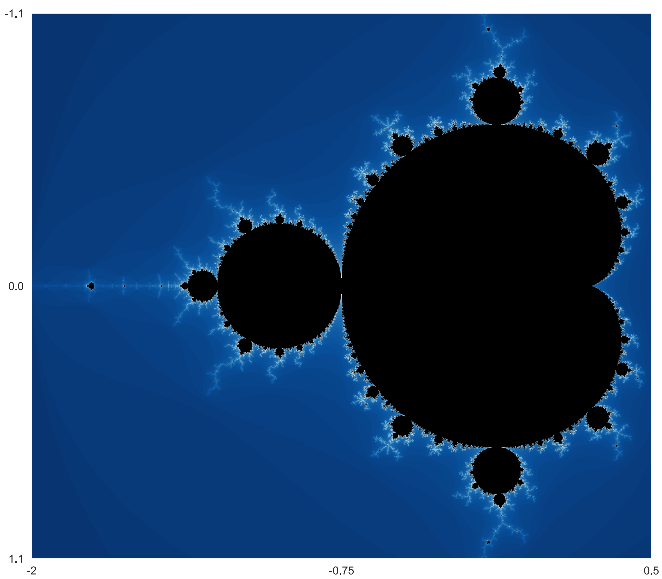

Fig. 1. Visualisation of the Mandelbrot set on a predetermined plane, with $p = 2000$ and $i = 1000$

B. Monte Carlo and the Hit-and-miss Transform

MC experiments, as discussed before, are regularly used for optimisation, numerical integration and drawing from a probability distribution (Kroese et al., 2014). This research is focused on area estimation of the Mandelbrot set and could therefore be specified as a numerical integration problem, as integration is necessary for calculating areas.

Ross (1990) take the following integral, in which Ω is the domain of integration:

$$I = \int_{\Omega} f(x)dx \tag{2}$$

This integral can be related to an expectation by using the probability density function of x, which we define as $\rho(x)$:

$$I = \int_{\Omega} f(x)dx = \int_{\Omega} \frac{f(x)}{\rho(x)}\rho(x)dx = E\left[\frac{f(x)}{\rho(x)}\right] = E\left[g(x)\right] \tag{3}$$

If we take x as independent and identically distributed random variables with probability density function $\rho(x)$, the LLN can be applied to evaluate $E\left[g(x)\right]$ as the sample mean of N samples:

$$E\left[g(x)\right] = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} g(x_k) \tag{4}$$

We want to quantify an estimation of the area (A_M) of the Mandelbrot set and therefore need to combine the MC process with a hit-and-miss transform of the random variables we generate. The process will start by generating s samples from the complex plane from Figure 1. These samples will be taken as a starting value for the iteration in Equation 1 and x_i can be defined as the hit-and-miss transform by:

$$x_i = \begin{cases} 1 & \text{if } x \text{ does not diverge within } i \text{ iterations} \\ 0 & \text{if } x \text{ diverges within } i \text{ iterations} \end{cases} \tag{5}$$

This is basically a Bernouilli distribution with $p = \frac{A_M}{A_{plane}}$, with $A_{plane} = (2.2 * 2.5) = 5.5$ in our case. The expected value of this distribution is equal to p and we therefore estimate $\hat{p} = \frac{1}{s}\sum_{k=1}^{s}\frac{x_{i,k}}{5.5}$ for large s by the LLN in Equation 4. Multiplying both sides by 5.5 would not change the result and provides an estimate of A_M.

C. Variance reduction

The accuracy of the estimation depends largely on the sample we take. The better a sample represents the entire set, the better the estimation will be, but the more computational time it will take to provide estimations. Several variance reduction techniques exist to make more reliable calculations with given computational power. We will discuss standard pure random sampling and two methods to reduce the variance, stratified sampling and importance sampling.

Regular MC experiments are performed with samples from an uniform distribution, samples are completely random on the specified interval and could be very concentrated around certain areas. This could bias individuals simulations and therefore blow up the variance of the estimation.

Stratified sampling tries to improve on this by making sure that samples are more evenly spread across the specified intervals. We specify two stratified sampling methods in *latin hypercube sampling* (LHC) and *orthogonal sampling*. For LHC sampling, we divide the complex plane in a grid of l regions and make sure that every row and every column of this grid contains exactly 1 sample. Asymptotically, the variance is less than the variance obtained using pure random sampling, with the degree of variance reduction depending on the degree of additivity in the function being integrated (Stein, 1987).

LHC sampling could still produce relatively clustered samples, orthogonal sampling is an improvement that could further distribute the samples evenly. We now define the same grid as before, but place a coarser grid on top. Every row and column of the coarser grid again needs a specific amount of samples, equal to the amount of fine regions that fit within the coarse grid. Samples are now generated completely orthogonal and will represent the specified subspace very evenly. Tang (1993) shows that for integration, which we perform in this research, the sampling scheme with orthogonal sampling offers a substantial improvement over LHC sampling.

Alternatively, we can reduce the variance by importance sampling. This technique specifies a separate distribution for the samples, that captures the importance of specific points accordingly. Even though samples are not generated uniformly over the subspace, unbiased results can be obtained by reweighting these samples by the probability they are taken. The expectation from 4 is basically replaced by $E_h\left[\frac{g(\mathbf{X})\rho(\mathbf{X})}{h(\mathbf{X})}\right]$ (Ross, 1990). Where the expectation is over distribution h of which we generate our samples, g and ρ are specified as before.

III. Experiments

As the estimation of A_M can be a computationally expensive process, we try to gain insight into its convergence behavior. We want to compare various techniques to reduce variance and improve convergence. The two main parameters of interest for this objective are obviously i and s, representing the maximum number of iterations and the number of samples respectively. We define discrete sets of possible values for i and s in $(0, 5000)$. The value for i could influence quantification of the area by labeling a starting value within the Mandelbrot set, when it would diverge after a number of iterations larger than i. As explained in Section II-B, $s \to \infty$ would present asymptotic results coming from the LLN and finite values for this parameter are therefore important.

Both parameters are of interest, but using an unbalanced combination of finite i and s could present invalid results. We want to prevent this problem from happening by looking at the difference between a specific area and the same area with equal i and highest s (or the other way around). The evolution of this statistic is initially observed with increasing i, to find a value for i that induces minimal bias into our estimations. We subsequently observe the progress of the variance for increasing s, expecting it to decrease with s. Estimations are computed for pure sampling as well as sampling through the variance reduction techniques from Section II-C. They will be compared to the area estimate by Lesmoir-Gordon and Rood (2000) that is calculated through a method similar to ours.

An larger range of s values, from $(0, 20000)$, is used with the appropriate i value that rarely induces bias. These simulations will be used to quantify the effectiveness of the variance reduction techniques.

To account for the randomness caused by number generation in our simulation method, we repeat each simulation 50 times. Areas are calculated as the average over these 50 repetitions and standard deviations, confidence intervals and hypothesis tests can be performed because of asymptotic normality of our statistic.

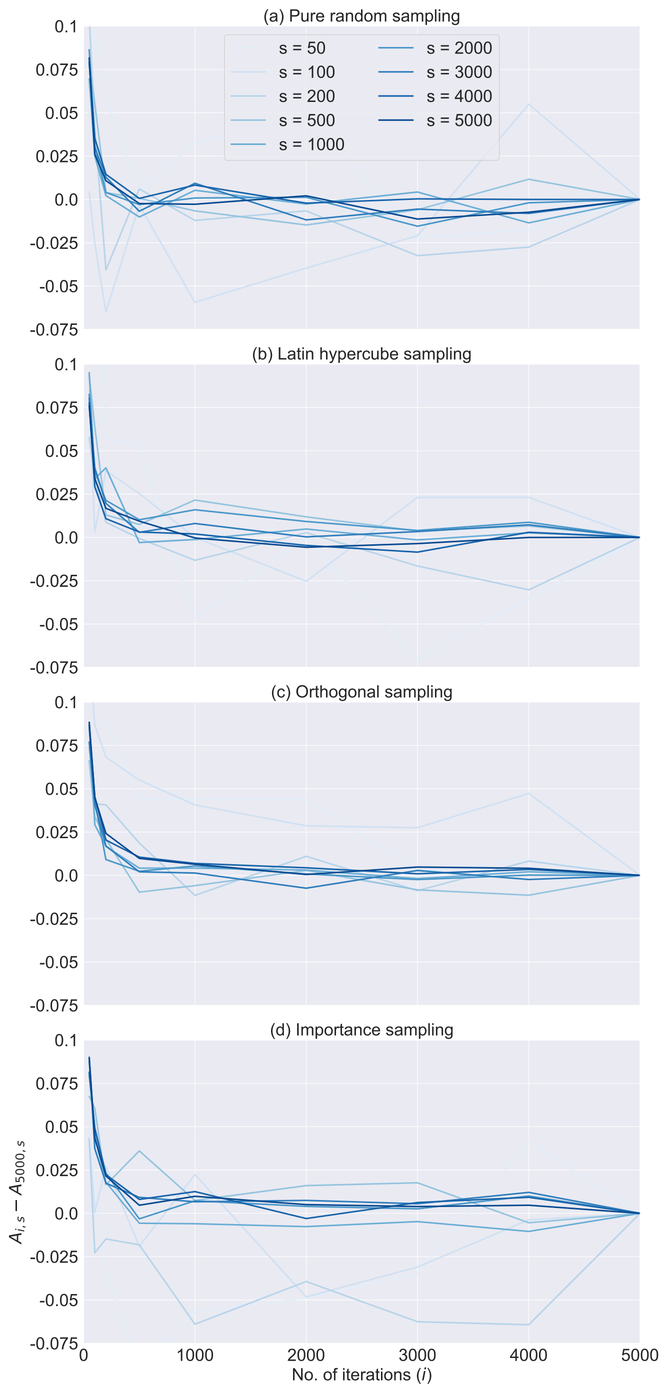

Fig. 2. Difference in area for increasing i compared to $i = 5000$ with varying sampling methods, values presented are averages over 50 replications.

IV. Results

As discussed before, we observe the progress of area differences for increasing i and s respectively. Figure 2 presents average area differences for increasing i, showing dissimilarities between the four sampling methods. As expected, the lines corresponding to the lower s values fluctuate the most and LHC and orthogonal sampling seem to keep the differences between s-values the smallest. An important result can be observed in the flattening of the curve after an i of approximately 1000. It shows that low values i overestimate the area of the Mandelbrot set, but that increasing i becomes unnecessary at a certain point.

The parameter setting for i is expected to possibly induce bias into the estimate, but should not directly influence the standard deviation of the area estimates after a certain point. The flattening curve after $i = 1000$ that we have seen in Figure 2 suggests a stable standard deviation for larger i. Figure 3 confirms this expectation under pure sampling, as all lines are approximately horizontal after $i = 1000$ and we can only see a decrease of the standard deviation by the change in s. The same result holds for the other sampling methods, as can be seen in Figure 6 in the Appendix.

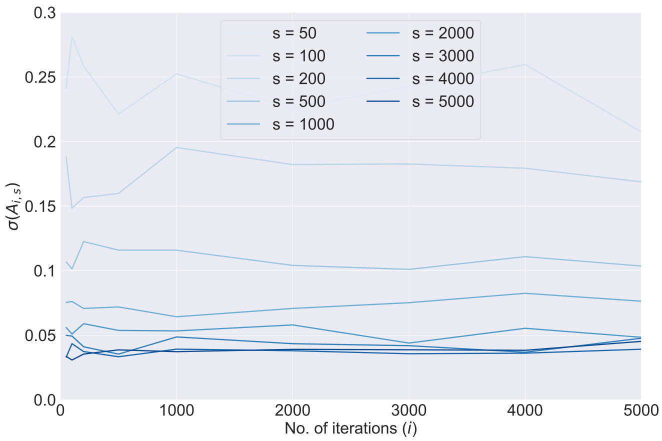

Fig. 3. Standard deviation for increasing i and pure random sampling, computed over 50 replications.

Whereas the standard deviations in Figure 3 already showed a hint of the effect of s, we visualise this effect more clearly in Figure 4. A decreasing standard deviation is obviously visible for pure random sampling and it seems to stabilise, but this would be more obvious in a plot for a larger range of s. Note the same effect for the three variance reduction methods, as shown in Figure 7 in the Appendix.

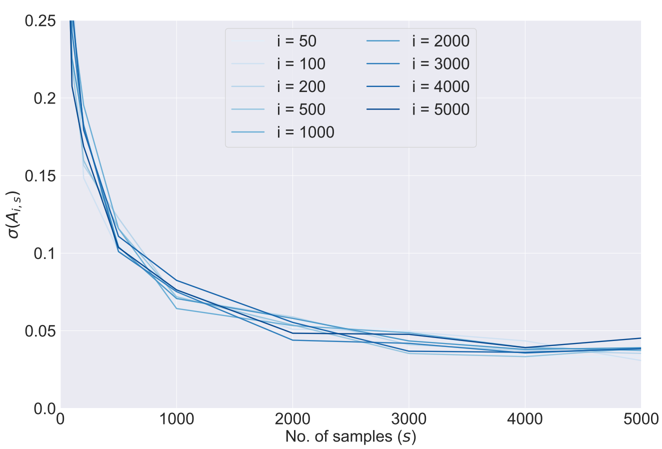

Fig. 4. Standard deviation for increasing s and pure random sampling, computed over 50 replications.

Confidence intervals for various values of i and s are presented in Table I. When s is relatively large, 5000 in this case, increasing i further than 1000 shows to have barely any effect on the confidence intervals, except when orthogonal sampling is used. Changing i has a larger effect when s is smaller, whether the effect on the confidence interval is positive or negative depends on the sampling method used. To provide better insight into the converge of these estimations, we compare the confidence intervals in the table to the area estimation by Lesmoir-Gordon and Rood (2000) which is equal to 1.50659177. The gray cells in the table correspond to confidence intervals that include this value. It is remarkable that all but one of the gray cells are the result of estimations with lower s.

TABLE I
95 PERCENT CONFIDENCE INTERVALS FOR VARYING i AND s

s	Sampling	$i = 500$	$i = 1000$	$i = 5000$
500	Pure random	(1.503, 1.523)	(1.496, 1.516)	(1.503, 1.522)
	LHC	(1.505, 1.516)	(1.519, 1.53)	(1.496, 1.509)
	Orthogonal	(1.497, 1.507)	(1.501, 1.51)	(1.506, 1.517)
	Importance	(1.526, 1.54)	(1.497, 1.512)	(1.489, 1.505)
5000	Pure random	(1.511, 1.513)	(1.511, 1.513)	(1.513, 1.516)
	LHC	(1.52, 1.521)	(1.51, 1.512)	(1.511, 1.512)
	Orthogonal	(1.513, 1.514)	(1.51, 1.511)	(1.504, 1.505)
	Importance	(1.508, 1.509)	(1.513, 1.514)	(1.503, 1.505)

We have observed the effects of various parameter settings for i and s and see that increasing i does not seem to have a significant impact for $i \geq 1000$, but s-values larger than 5000 might be interesting to explore. As computational costs are now saved by choosing a single i-value, we can increase the range of s values to 20000 in Figure 5.

The fluctuations in the area difference that are obvious in the $(0, 5000)$ range show to decrease when s increases towards 20000. The error bars show a single standard deviation above and below the sample mean for perspective. The right graph shows the estimations for the standard deviations, which certainly decrease further after $s = 5000$. The stratified sampling methods seem to be most effective in reducing the variance, hypothesis tests on the significance of their variance reduction are presented in Table II.

Table II presents variance estimates for $i = 1000$ and s equal to 1000, 5000 and 20000 respectively. It also shows F-statistics for the four sampling methods compared to pure random sampling, so these F-statistic will obviously be equal to 1 when pure random sampling is compared to itself. All three variance reduction techniques reduce the variance significantly (note that this is a one sided test) for these parameter settings at the 95 percent confidence level. On the 99 percent confidence level, importance sampling does not significantly lower the variance for $s = 1000$, all the other values are still significant.

TABLE II
F-TEST ON VARIANCE FOR VARIOUS SAMPLING METHODS,
$i = 1000$ AND VARYING s

	$s = 1000$	$s = 5000$	$s = 20000$
Pure	1.0000	1.0000	1.0000
LHC	2.6575	3.6548	2.8123
Orthogonal	4.6615	3.0113	2.5358
Importance	1.1105	1.4433	1.4224
$F_{crit,0.95}$	1.1097	1.04767	1.0235
$F_{crit,0.99}$	1.1587	1.068	1.0334

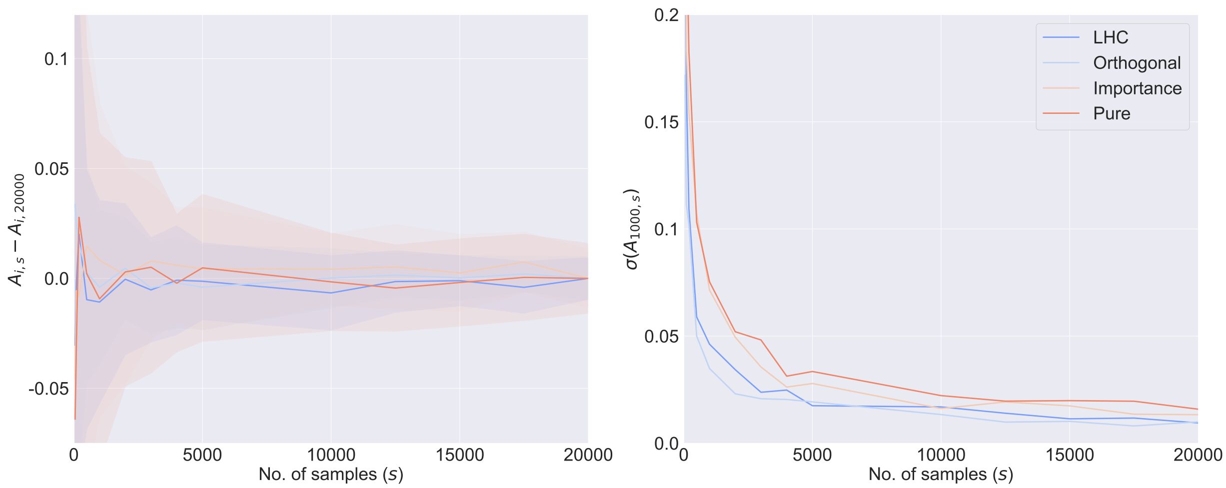

Fig. 5. Area differences and standard deviations for increasing s, $i = 1000$ and all sampling methods.

V. Conclusion

Trying to successfully estimate the area of the Mandelbrot set is not an easy task. Modern day computational power has made this estimation easier, we implement a Monte Carlo structure in combination with a hit-or-miss transform to provide estimations of the area. Variance reduction techniques like stratified and importance sampling can be helpful in reaching convergence and we explore their effects.

The estimation of the area of the Mandelbrot set depends on the number of iterations to determine whether a value is in the set and the number of samples that are used. We find that the number of iterations necessary to reduce most of the bias in the estimation lays around 1000. Higher numbers for this parameter are computationally expensive and probably not worth using. The number of samples used however, should always be of major importance as it is necessary for the asymptotic convergence of MC. We observe that an increasing number of samples clearly lowers the variance of the estimation.

Comparing our estimations to the estimation from Lesmoir-Gordon and Rood (2000) delivers a strange result. The confidence intervals from the estimations with 500 samples include the estimation from literature more often than the estimations with 5000 samples. We also test for the effectiveness of the variance reduction techniques for various sample sizes. At a 99 percent confidence level, we find that LHC and orthogonal sampling significantly reduce the variance for all sample sizes. Importance sampling does so for higher sample sizes, but does not significantly effect the variance for lower sample size. This is an important result, as variance reduction can be especially important when computational power only allows for small sample size.

Although our research provides a clear overall overview of the effect of several parameter settings and techniques, estimations of the area are not as precise as expected. The upper and lower bounds that Ewing and Schober (1992) find match with our estimations, but specific calculations from e.g. Lesmoir-Gordon and Rood (2000) seem to differ from the ones in this research. Bittner et al. (2017) redefine the upper bounds by Ewing and Schober (1992) with additional computational power, and adding this would certainly be an appropriate approach to extending the current research. Alternatively, it would be interesting to investigate other variance reduction techniques such as antithetic variables, control variets or quasi-Monte Carlo.

References

Andreadis, I. and Karakasidis, T. E. (2015). On a numerical approximation of the boundary structure and of the area of the mandelbrot set. *Nonlinear Dynamics*, 80(1-2):929–935.

Bittner, D., Cheong, L., Gates, D., and Nguyen, H. (2017). New approximations for the area of the mandelbrot set. *Involve, a Journal of Mathematics*, 10(4):555–572.

Ewing, J. H. and Schober, G. (1992). The area of the mandelbrot set. *Numerische Mathematik*, 61(1):59–72.

Fisher, Y. and Hill, J. (1993). Bounding the area of the mandelbrot set. *Availalbe at http://citeseer. ist. psu. edu/35134. html*.

Kroese, D. P., Brereton, T., Taimre, T., and Botev, Z. I. (2014). Why the monte carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(6):386–392.

Landau, D. P. and Binder, K. (2014). *A guide to Monte Carlo simulations in statistical physics*. Cambridge university press.

Lesmoir-Gordon, N. and Rood, W. (2000). *Introducing fractal geometry*. Totem Books.

Ross, S. M. (1990). *A course in simulation*. Prentice Hall PTR.

Stein, M. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151.

Tang, B. (1993). Orthogonal array-based latin hypercubes. *Journal of the American statistical association*, 88(424):1392–1397.

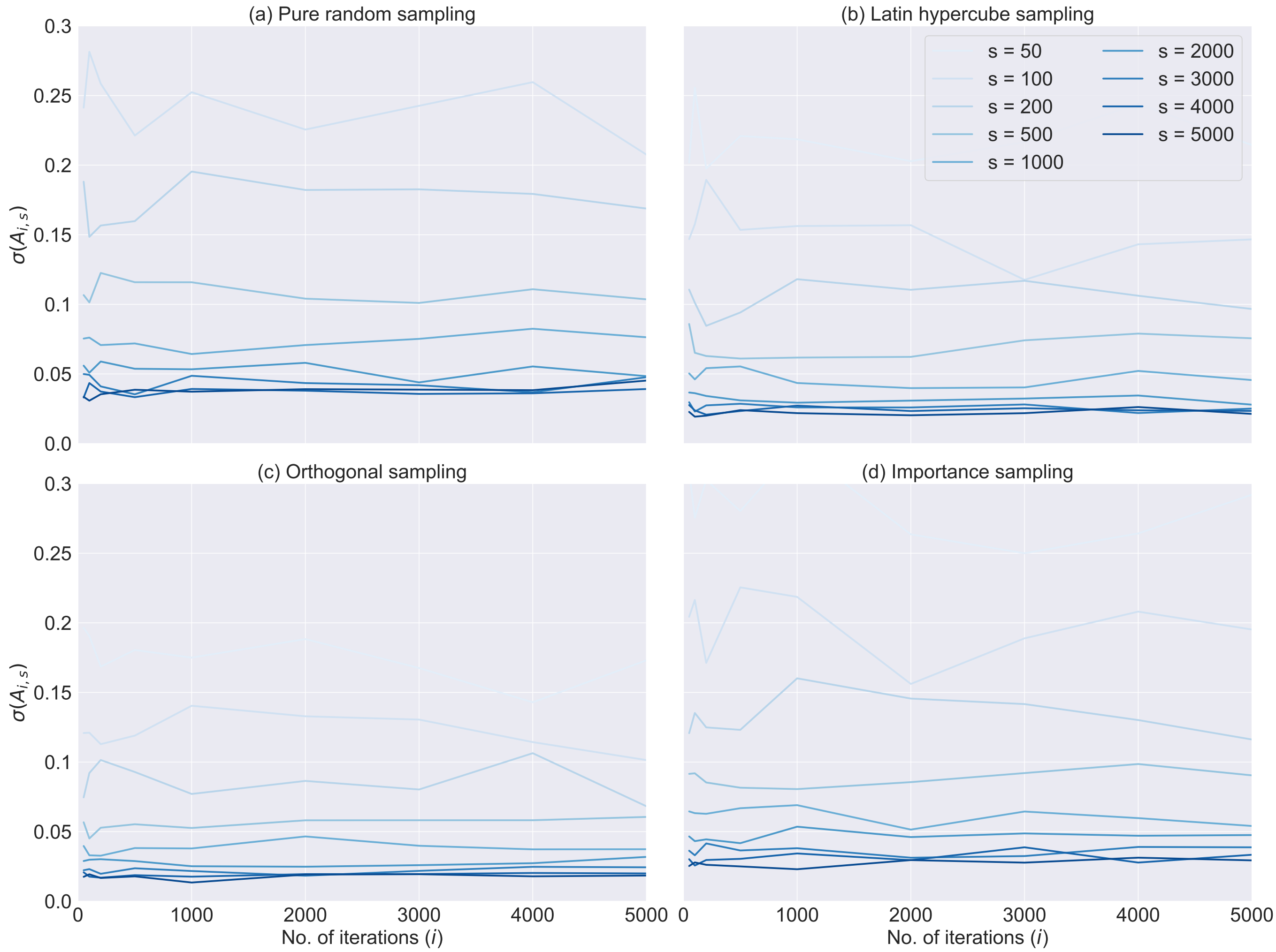

Fig. 6. Standard deviation for increasing i and varying sampling methods, computed over 50 replications.

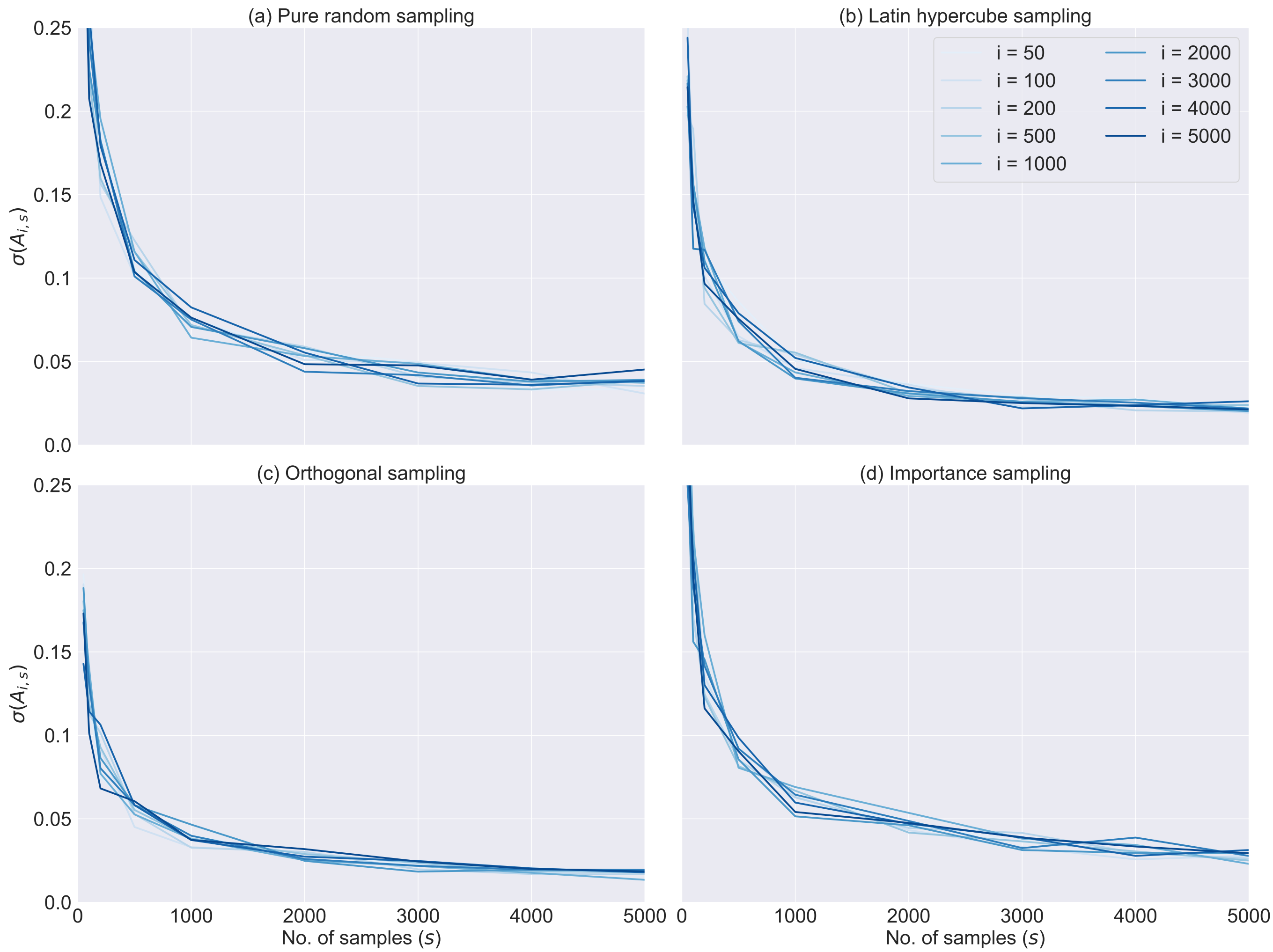

Fig. 7. Standard deviation for increasing s and varying sampling methods, computed over 50 replications.