

# Assignment 1: Algorithms Galore!

Chosen algorithm: Rejection Sampling

Computational Statistics

Student: Lucas Machado Moschen

Instructor: Luiz Max de Carvalho

November 22, 2021

**Hand-in date: 22/11/2021.**

## General guidance

- State and prove all non-trivial mathematical results necessary to substantiate your arguments;
- Do not forget to add appropriate scholarly references *at the end* of the document;
- Mathematical expressions also receive punctuation;
- Please hand in a single PDF file as your final main document.  
Code appendices are welcome, *in addition* to the main PDF document.

## Background

By now we have seen quite a few methods for computing integrals *via* Monte Carlo. Each method has its own advantages and drawbacks. It is important that we understand these properties in order to apply the methods effectively. In this assignment we will continue studying the problem of computing the average distance between two points on a disk, this time from the perspective of method comparison. That is to say that in this assignment you will experience the microcosm of comparing several methods for solving a problem for which we happen to know the right answer in closed-form.

Recall that we want to compute

$$\begin{aligned} I &= \frac{1}{\pi^2 R^4} \int_0^R \int_0^R \int_0^{2\pi} \int_0^{2\pi} \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos \phi(\theta_1, \theta_2)} r_1 r_2 d\theta_1 d\theta_2 dr_1 dr_2, \\ &= \frac{128}{45\pi} R, \end{aligned}$$

where  $\phi(\theta_1, \theta_2)$  is the central angle between  $r_1$  and  $r_2$ .

Here we will do something a bit risky: we will compare a few methods to compute  $I$  using a bunch of different methods without knowing in advance which will work best or if there are going to be any differences at all. Welcome to Science!

## Methods

Here we will list a selection of methods that will be randomly assigned to each student, along with some questions that need to be answered for that particular method.

- **Rejection sampling**

- Justify your choice of proposal distribution and show that it conforms to the necessary conditions for the algorithm to work; in particular, try to find a proposal that gives the highest acceptance probability.

- **Importance sampling**

- Justify your choice of proposal based on the variance of the resulting estimator.

- **Gibbs sampling**

- Write your full conditionals out and show that they adhere to the Hammersley-Clifford condition.

- **Metropolis-Hastings**

- Justify your choice of proposal; test different ones if you need to.

- **Static Hamiltonian Monte Carlo**

- Comment on the choice of step size ( $\epsilon$ ) and integration time ( $\tau$ ).

## Questions

1. You have been (randomly) assigned a method from the previous section. Represent  $I$  as  $\int_{\mathcal{X}} \phi(x) \pi(x) dx$  and justify your choice of  $\phi$ ,  $\pi$  and  $\mathcal{X}$ . Recall that these choices are arbitrary up to a point, but they might lead to wildly different empirical performances **and** theoretical properties for estimators of  $I$ . **Justify** your choices in light of the method you have been given to work with. Choose wisely and be rigorous in your justifications.
2. Again, starting from the eventual samples you will obtain with your method, construct a non-empty<sup>1</sup> family of estimators of  $I$  and discuss whether it is (strongly) consistent and whether a central limit theorem can be established.
3. Detail a suite of diagnostics that might be employed in your application to detect convergence or performance problems. Extra points for those who design algorithms that exploit the structure of this particular integration problem.
4. For each  $R \in \{0.01, 0.1, 1, 10, 100, 1000, 10000\}$ , perform  $M = 500$  runs from your simulation method and compute: (i) variance (ii) bias (iii) standard deviation of the mean (MCSE).
5. Can you identify one key quantity missing from the previous item? *Hint*: it bears relevance to the real world application of any computational method.

**Warning:** the questions in this assignment might seem deceptively simple; do not be fooled. I expect a lot of effort from you in making your method work the best it can. This entails loads of failed derivations and experiments, which you are encouraged to report in order to document the discovery process. Also, feel free to include answers to questions that have not been asked, if you feel they are relevant. Make loads of figures and tables and let your scientific imagination run wild! Good luck!

---

<sup>1</sup>This is a joke. It means you should come up with at least one estimator. But you might, and are even encouraged to, entertain more than one estimator.

## Rejection Sampling

To obtain a sample from distribution  $\pi$ , the idea of Rejection Sampling method is to sample from a distribution  $q$  and reject some of the samples to correct the distribution of the sampling. If for all values of  $x \in \mathcal{X}$ ,  $\pi(x) \leq Mq(x)$ , the acceptance probability is

$$\alpha = \frac{\pi(x)}{Mq(x)}.$$

The number of necessary samples until acceptance has Geometric Distribution with parameter  $1/M$  (Robert et al., 2004, p. 51). To calculate an approximation for the integral  $I$  through this method, we must define a target distribution  $\pi$  up to a normalizing constant and  $q$  such that  $\pi/q$  be bounded by  $M$  in  $\mathcal{X}$ . The presentation is in the order of the discovery process.

## A simple first derivation

The simplest method is to define  $\pi(x) \propto 1$  para  $x \in [0, R]^2 \times [0, 2\pi]^2$ . Since we have a method for sampling from  $\pi$ , we can define  $q = \pi$  and  $M = 1$ . Our Monte Carlo estimator is therefore

$$\hat{I}_1 = \frac{4}{mR^2} \sum_{i=1}^m \sqrt{\left(r_1^{(i)}\right)^2 + \left(r_2^{(i)}\right)^2 - 2r_1^{(i)}r_2^{(i)} \cos\left(\theta_2^{(i)} - \theta_1^{(i)}\right)} r_1^{(i)} r_2^{(i)},$$

where  $x^{(i)} = \left(r_1^{(i)}, r_2^{(i)}, \theta_1^{(i)}, \theta_2^{(i)}\right)$  is a sample from  $\pi(x)$ . This estimator can be simplified with  $\pi(x) \propto 1$  for  $x \in [0, 1]^4$  with

$$\hat{I}_2 = \frac{4R}{m} \sum_{i=1}^m \sqrt{\left(x_1^{(i)}\right)^2 + \left(x_2^{(i)}\right)^2 - 2x_1^{(i)}x_2^{(i)} \cos\left(2\pi\left(x_4^{(i)} - x_3^{(i)}\right)\right)} x_1^{(i)} x_2^{(i)},$$

decreasing the number of required multiplications. Another advantage of  $\hat{I}_2$  over  $\hat{I}_1$  is that the samples can be reused for different values of  $R$ . Both  $\hat{I}_1$  and  $\hat{I}_2$  have the advantage of enabling the calculation of their variance. The second moment of each term of  $\hat{I}_1$  is

$$\begin{aligned} & \frac{1}{4\pi^2 R^2} \int_0^R \int_0^R \int_0^{2\pi} \int_0^{2\pi} r_1^4 r_2^2 + r_2^4 r_1^2 - 2r_1^3 r_2^3 \cos(\theta_2 - \theta_1) \\ &= \frac{1}{4\pi^2 R^2} \frac{8\pi^2 R^8}{15} = \frac{2R^6}{15} \end{aligned}$$

and the first moment is

$$\begin{aligned} & \frac{1}{4\pi^2 R^2} \int_0^R \int_0^R \int_0^{2\pi} \int_0^{2\pi} \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta_2 - \theta_1)} r_1 r_2 \\ &= \frac{R^2}{4} I = \frac{32R^3}{45\pi}. \end{aligned} \tag{1}$$

Therefore the variance is

$$\begin{aligned} \text{Var}[\hat{I}_1] &= \frac{4^2}{mR^4} \left( \frac{2R^6}{15} - \frac{1024R^6}{45^2\pi^2} \right) \\ &= \frac{1}{m} \left( \frac{32}{15} - \frac{2^{14}}{45^2\pi^2} \right) R^2 \approx \frac{1.31}{m} R^2. \end{aligned}$$

Moreover, this estimator is unbiased by relation (1). Similar calculations for  $\hat{I}_2$  indicate the same variance, since only linear transformations are made. By the Strong law of large numbers, with the  $\hat{I}_1$  and  $\hat{I}_2$  converges to the integral  $I$  with probability 1. The central limit theorem give bounds for this convergence, stating that

$$\sqrt{m}(\hat{I}_1 - I) \xrightarrow{d} \text{Normal}(0, 1.31R^2).$$

This proposal also gives the highest acceptance probability, as required.

## Regular polygon circumscribed around a circle

When trying to find an approximation for  $\pi$ , the ratio of the circumference of a circle to its diameter, Archimedes noticed that taking the mean between the perimeters of a  $n$ -regular circumscribed polygon and a  $n$ -regular inscribed polygon, the result would converge to the circumference of the circle. Here the idea is to sample from the circumscribed polygon with  $n$  sides and use the rejection sampling to obtain the distribution  $\pi(x)$ .

First, we reparametrize integral  $I$  in Cartesian coordinates

$$\int_{\mathcal{X}} \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} \pi(x) dp_x dp_y dq_x dq_y,$$

where  $\mathcal{X}$  is the circle of radius  $R$  and  $\pi(x)$  is the uniform distribution over  $\mathcal{X}$ .

We define  $q_n$  to be the uniform distribution over the circumscribed polygon of  $n$  sides and we must develop a method to sample from  $q$ . First of all, we break the circle in  $n$  equal pieces as presented in Figure 1. A circumscribed polygon has its sides tangent to the circle at the midpoints between the red ones. Figure 2 presents the tangent points in blue.

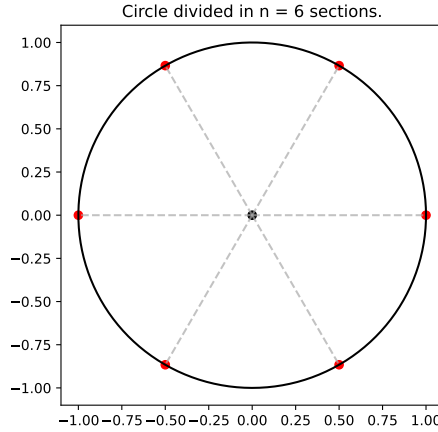


Figure 1: Circle divided in  $n = 6$  sections delimited by the red points and the grey lines.

Let  $(x_m, y_m)$  be a midpoint between two red points  $(x_1, y_1)$  and  $(x_2, y_2)$ . The tangent line over the circle passing por that point has the following expression

$$y = -\frac{x_m}{y_m}(x - x_m) + y_m,$$

which is derived with implicit differentiation for  $x^2 + y^2 = R^2$ . Moreover, the line passing through the center of the circle and the red points is given by

$$y = \frac{y_1}{x_1}x.$$

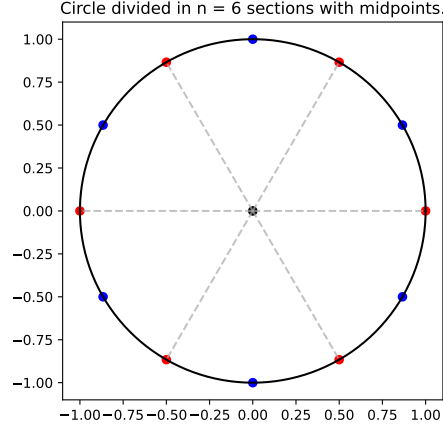


Figure 2: Circle divided in  $n = 6$  sections and midpoints in blue.

They cross at the point

$$-\frac{x_m}{y_m}(x - x_m) + y_m = \frac{y_1}{x_1}x \implies x = \frac{\frac{x_m^2}{y_m} + y_m}{\frac{x_m}{y_m} + \frac{y_1}{x_1}}, y = \frac{y_1}{x_1}x.$$

After extending the segments splitting the circle, we can build the polygon as Figure 3 presents in green. Figure 4 shows the same construction for different values of  $n$ . Notice that when  $n = 12$ , the approximations already is quite good.

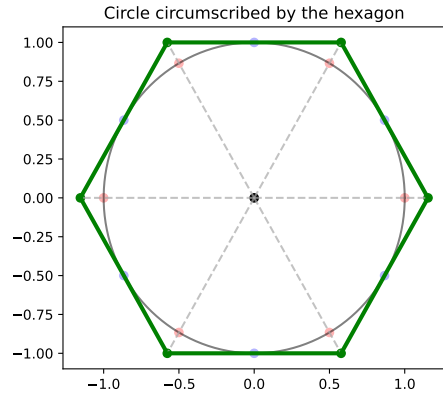


Figure 3: Hexagon circumscribing the circle.

Now that we are able to build an arbitrarily polygon with  $n$  sides, we need to know how to sample uniformly from it. Notice that the region delimited by

Different approximations of the circle by polygons

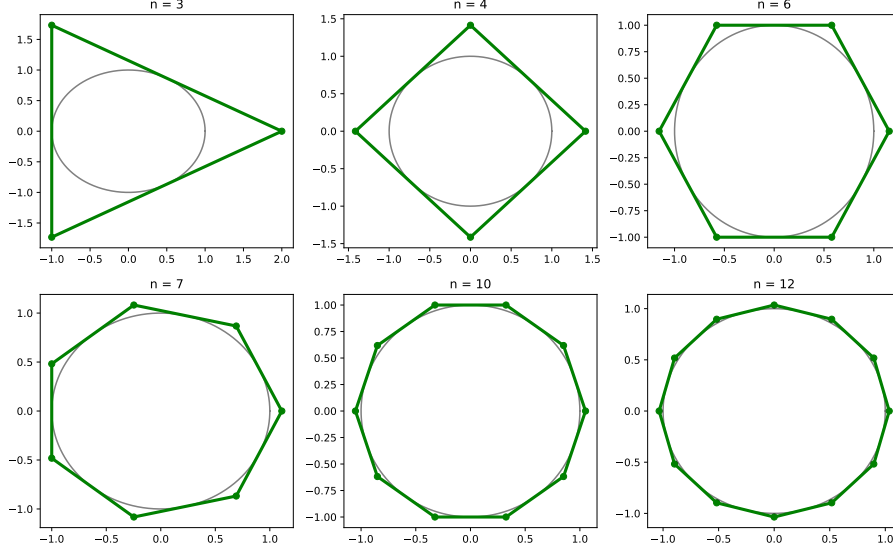


Figure 4: Different polygons approximating the circle.

the circle is included in the region delimited by the polygon. Since we separated the polygon in  $n$  sections, we perform the following algorithm:

- (i) Draw  $i \sim \text{Unif}\{1, \dots, n\}$  and pick the corresponding section delimited by the points  $(0, 0), (x_i, y_i), (x_{i+1}, y_{i+1})$  in the polygon. If  $i = n$ , we change  $i + 1$  for 1;
- (ii) Draw uniformly from the corresponding triangle.

We have to specify how to sample inside a specific triangle delimited for three points. Fortunately, this is not a hard thing. Let  $u_1, u_2 \sim \text{Unif}[0, 1]$ . The sum vector

$$v = u_1(x_i, y_i) + u_2(x_{i+1}, y_{i+1})$$

is contained in the parallelogram formed by twice the triangle. Figure 5 shows this relation for 10000 samples. To chop the part we do not want to, we have to reflect each point outside the triangle to inside the triangle. If  $u_1 + u_2 \leq 1$ , the point  $v$  is inside the triangle. Otherwise, denote  $v_1 = 1 - u_1 \text{ Unif}[0, 1]$  and  $v_2 = 1 - u_2 \text{ Unif}[0, 1]$ , and we will have  $v_1 + v_2 = 2 - (u_1 + u_2) \leq 1$ . In that sense, we are mapping each point outside to one inside. Therefore, the uniformity is kept. Figure 6 presents the result of the algorithm. Therefore, we know how to sample from  $q_n$ .

We also know to calculate the area of the circumscribing polygon of  $n$ -sides by the formula engineering fundamentals (2021)

$$A = nR^2 \tan\left(\frac{\pi}{n}\right).$$

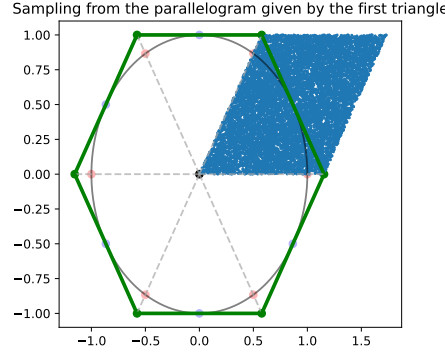


Figure 5: Different polygons approximating the circle.

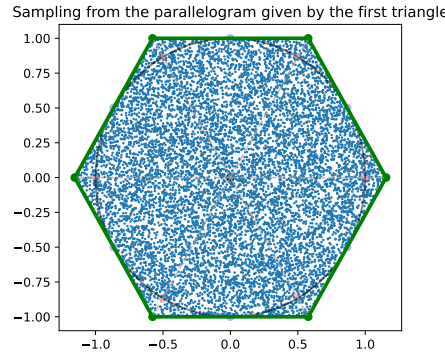


Figure 6: Sampling uniformly in the hexagon.

Therefore,

$$\pi(x) = \frac{1}{\pi^2 R^4}, q_n(x) = \frac{1}{n^2 R^4 \tan^2(\frac{\pi}{n})} \implies \frac{\pi(x)}{q(x)} = \frac{(n \sin(\pi/n))^2}{\cos^2(\pi/n) \pi^2}$$

From that, for each  $n$  we calculate

$$M_n = \frac{(n \sin(\pi/n))^2}{\cos^2(\pi/n) \pi^2},$$

the average number of samples until acceptance. Observe that

$$\lim_{n \rightarrow \infty} M_n = \frac{1}{\pi^2} \left( \lim_{x \rightarrow 0^+} \frac{\sin(x\pi)}{x} \right)^2 = \left( \lim_{y \rightarrow 0^+} \frac{\sin(y)}{y} \right)^2 = 1,$$

as expected. Figure 7 show several values for  $M_n$  for different values of  $n$ .

After sampling from  $q_n$  distribution, to obtain  $\pi$  through Rejection sampling, we must reject samples  $(x^{(i)}, y^{(i)})$  such that

$$(x^{(i)})^2 + (y^{(i)})^2 > R^2.$$

This summarizes in a simple algorithm:



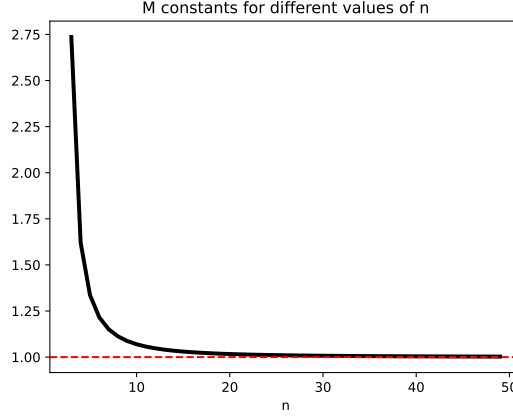


Figure 7: The red line delimits the limit value.

- (i) Fix the number of sides  $n$  and number of samples  $m$ ;
- (ii) Construct the polygon of  $n$  sides;
- (iii) Repeat  $m$  times:
  - Sample  $i \sim \text{Unif}\{1, \dots, n\}$ . Let  $T$  be the triangle delimited by the points  $(0, 0), (x_i, y_i), (x_{i+1}, y_{i+1})$ . If  $i = n$ , substitute  $i + 1$  for 0;
  - Sample  $v_1, v_2 \sim \text{Unif}[0, 1]$ . If  $v_1 + v_2 > 1$ , let  $u_1 = 1 - v_1$  and  $u_2 = 1 - v_2$ . Otherwise let  $u_1 = v_1$  and  $u_2 = v_2$ ;
  - If  $\|u_1(x_i, y_i) + u_2(x_{i+1}, y_{i+1})\|^2 > R^2$ , ignore this step and repeat it. Otherwise save the sample

$$p = u_1(x_i, y_i) + u_2(x_{i+1}, y_{i+1}).$$

With the sample from  $p^{(i)}, q^{(i)} \sim \pi(x)$  built by Rejection Sampling, we introduce

$$\hat{I}_3 = \frac{1}{m} \sum_{i=1}^m \|q^{(i)} - p^{(i)}\|.$$

This is an unbiased estimator, consistent with asymptotic distribution with similar proves to  $\hat{I}_1$  by the calculations we made in the warmup assignment (proving the both forms of integral were equivalent). The disadvantage of  $\hat{I}_3$  is the wasted samples, but it introduces a class of estimators that depend on the number of the sides of a polygon.

## Diagnosis

The samples obtained through Rejection Sampling are independent, which assures theoretical results in a simple manner. To track the number of rejected samples, we iteratively calculate  $\widehat{M}_n$  for each  $m$ . We expect that

$$\widehat{M}_n \xrightarrow{m \rightarrow \infty} M_n.$$

This generates a graphic of  $\widehat{M}_n$  for different number of samples  $m$ .

For the Monte Carlo convergence, we follow a suggestion given by (Robert et al., 2004, p. 124) to run parallel estimates and calculate the 90% empirical band. We can also compare to the normal approximation by the Central Limit Theorem. If no asymptotic property is desired, we can also use Chebyshev inequality:

$$\Pr(|\hat{I}_1 - I| \geq \alpha\sigma) \leq \alpha^{-2},$$

where  $\sigma$  is the standard deviation of this estimator. In this case we know it in closed form. If no closed form were known, the approximation would be in levels of the standard deviation, which is a general good strategy.

## Experiments

We draw 2000 samples for each of the runs and calculate variance, bias, MCSE, and  $\hat{R}$ . I notice that by theoretical results, we expect to see the variance being a value near to  $1.31R^2$  for each  $R$ . Table 1 summarizes the results for  $\hat{I}_2$ . I notice that the results are very similar to  $\hat{I}_1$  since the one is a simple transformation of the other.

R	$\hat{I}_2$			
	Variance	Bias	MCSE	R hat
0.01	$1.31 \cdot 10^{-4}$	$-5.55 \cdot 10^{-6}$	$2.63 \cdot 10^{-4}$	1.0
0.1	$1.32 \cdot 10^{-2}$	$-4.32 \cdot 10^{-5}$	$2.56 \cdot 10^{-3}$	1.0
1	$1.31 \cdot 10^0$	$-2.77 \cdot 10^{-3}$	$2.61 \cdot 10^{-2}$	1.0
10	$1.31 \cdot 10^2$	$-1.55 \cdot 10^{-2}$	$2.61 \cdot 10^{-1}$	1.0
100	$1.32 \cdot 10^4$	$-3.09 \cdot 10^{-2}$	$2.59 \cdot 10^0$	1.0
1000	$1.32 \cdot 10^6$	$2.09 \cdot 10^0$	$2.47 \cdot 10^1$	1.0
10000	$1.31 \cdot 10^8$	$4.88 \cdot 10^0$	$2.66 \cdot 10^2$	1.0

Table 1: Results of  $\hat{I}_2$  for the experiments.

I emphasize that the mean estimated of  $\hat{I}_3$  for any  $n$  are essentially the same, since the Monte Carlo estimator draws the samples from the same distribution: uniform on the circle. The only difference between them is the probability of acceptance, related to the number of rejected samples, and the time of execution, also related to the rejected samples. For that reason, we only show a unique estimate for  $\hat{I}_3$  in Table 2. Bias and variance are generally less for this estimator in comparison to  $\hat{I}_2$ , which is pretty interesting. Table 3 shows how the sampling are more efficient for  $\hat{I}_1$  and  $\hat{I}_2$  than the estimator using rejection sampling. Figure 8 shows that the ration between the number of samples and the accepted ones converge to one as predicted. Figures 9 and 10 shows the properties of consistency and asymptotic normality of the estimate  $\hat{I}_3$ .

R	$\hat{I}_3$			
	Variance	Bias	MCSE	R hat
0.01	$1.8 \cdot 10^{-5}$	$2.56 \cdot 10^{-6}$	$9.18 \cdot 10^{-5}$	1.0
0.1	$1.8 \cdot 10^{-3}$	$1.21 \cdot 10^{-5}$	$9.66 \cdot 10^{-4}$	1.0
1	$1.8 \cdot 10^{-1}$	$-7.3 \cdot 10^{-5}$	$9.79 \cdot 10^{-3}$	1.0
10	$1.8 \cdot 10^1$	$-2.67 \cdot 10^{-3}$	$9.3 \cdot 10^{-2}$	1.0
100	$1.8 \cdot 10^3$	$4.27 \cdot 10^{-2}$	$9.66 \cdot 10^{-1}$	1.0
1000	$1.8 \cdot 10^5$	$-4.95 \cdot 10^1$	$9.25 \cdot 10^0$	1.0
10000	$1.8 \cdot 10^7$	$-3.7 \cdot 10^0$	$9.27 \cdot 10^1$	1.0

Table 2: Results of  $\hat{I}_3$  for the experiments.

	$\hat{I}_1$	$\hat{I}_2$	$\hat{I}_3^{(4)}$	$\hat{I}_3^{(5)}$	$\hat{I}_3^{(6)}$	$\hat{I}_3^{(8)}$	$\hat{I}_3^{(10)}$	$\hat{I}_3^{(12)}$	$\hat{I}_3^{(20)}$
<b>N</b> ( $10^6$ )	2	2	2.55	2.32	2.2	2.11	2.07	2.04	2.02
<b>T</b> (s)	0.012	0.010	2.16	1.99	1.83	1.78	1.72	1.72	1.68

Table 3: Comparison between the number of samples  $N$  (accounting for the rejected whenever necessary) and the required time  $T$  to have 200000 accepted samples.

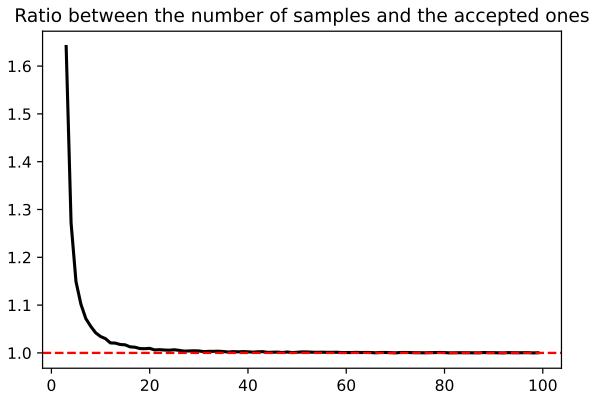


Figure 8: The red line marks the convergence value.

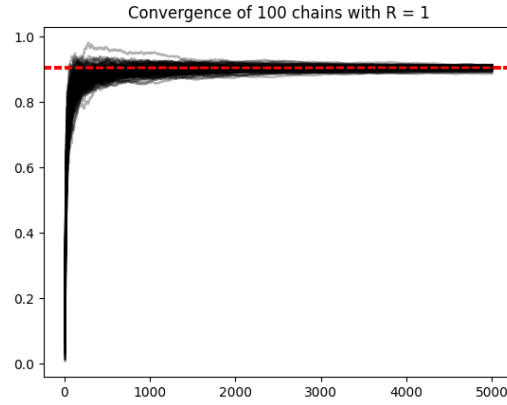


Figure 9: The red line marks the convergence value.

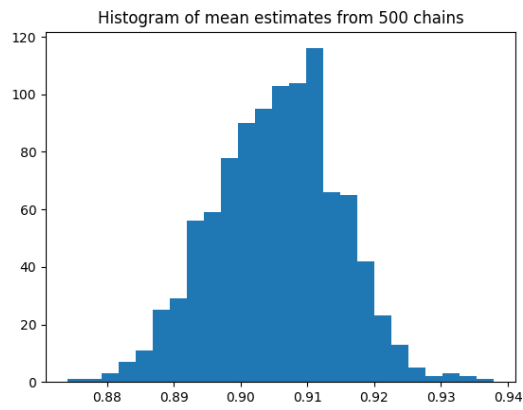


Figure 10: With  $R = 1$ , 2000 iterations and the histogram of  $\hat{I}_3$  for  $n = 10$ .

# Bibliography

engineering fundamentals, E. (2021). Web resources (sponsored by wolfram research). <https://www.efunda.com/math/areas/CircumscribePolygonGen.cfm>.

Robert, C. P., Casella, G., and Casella, G. (2004). *Monte Carlo statistical methods*, volume 2. Springer.