



Calculating mathematical constants with Monte Carlo simulations

**A project submitted to
Department of Statistics
St. Xavier's College (Autonomous), Mumbai
by**

**Bhoris Dhanjal
Khushi Chauhan**

2020-2021

St. Xavier's College (Autonomous), Mumbai
Department of Statistics
Project by

Sr. No.	Name of Student	Class	UID	Roll No.
1	Bhoris Dhanjal	SYBSc	202020	314
2	Khushi Chauhan	SYBSc	202019	313

Title of the Project:
Calculating mathematical constants with Monte Carlo
simulations

Declaration by the Student

We hereby declare that the project entitled “Calculating mathematical constants with Monte Carlo simulations” was completed and written by us is the result of original project work and has not formed earlier the basis for the certificate (any award) or similar title of this or any other school/college or examining body.

Date: July 15, 2021

Name and Signature of Student:

Bhoris Dhanjal

Khushi Chauhan

Acknowledgements

We would like to express our deepest thanks to our professors, Prof. Myrtle Fernandes and Prof. Ayesha Dias and the rest of the Department of Statistics at St. Xavier's, for encouraging us to explore this topic.

We would further like to thank our friends and family for offering emotional support.

We further extend our gratitude towards the vast trove of information we gleaned from the internet, via papers and various online forums on StackExchange (CrossValidated, Mathematics, Mathematica & TeX).

Contents

Declaration by the Student	ii
Acknowledgements	iii
1 Introduction	1
2 Numerical Monte Carlo integration	2
2.1 Overview of Monte Carlo method	2
2.2 Naive Monte Carlo integration	2
2.3 Error of Naive Monte Carlo integration	3
2.4 An example of naïve Monte Carlo Integration	4
3 Estimating Pi	5
3.1 Elementary Method	5
3.2 Buffon's Needle	6
3.3 Buffon's Noodle	9
3.3.1 A note on the simulation of noodles	9
4 Estimating Euler's number	11
5 Estimating Euler-Mascheroni constant and Stieltjes constants	12
5.1 Euler-Mascheroni constant	13
5.2 Stieltjes constants	14
6 Estimating Phi	16
7 Methods of integral variance reduction	16
7.1 Antithetic variates	17
7.2 Control variates	18
8 Quasi Monte-Carlo integration	20
8.1 Quasi-random sequences (low-discrepancy sequences) . .	20
8.2 Error in Quasi Monte-Carlo integration	21
8.3 Examples of Quasi-Random sequences	21
8.3.1 Van der Corput and Halton sequence	21
8.3.2 A cursory view of (t,s) sequences	24
8.3.3 Niederreiter sequence	24
8.3.4 Sobol sequence	25
8.4 Computing Apéry's constant	26
References	27

Calculating mathematical constants with Monte Carlo simulations

Bhoris Dhanjal
Khushi Chauhan

July 15, 2021

Abstract

This project primarily serves to showcase Monte Carlo simulations. We first derive the naïve Monte Carlo estimator for numerical integration. Following which we use Monte Carlo simulations to estimate a few well known mathematical constants. Variance reduction techniques and Quasi-Monte Carlo integration will be examined at the end.

1 Introduction

In this project we will explore Monte Carlo simulations. The objectives of this project are as follows:

- Provide a conceptual overview of the Monte Carlo method.
- Describing and demonstrating naïve numerical Monte Carlo integration.
- Using various Monte Carlo simulations to calculate π , e , γ_n , ϕ , and $\zeta(3)$.
- Describing variance reduction techniques for Monte Carlo integration.
- Demonstrating Quasi-Monte Carlo integration and comparing it with naïve numerical MC integration.

The Mathematica notebook (comprising of all code used for computations and plots) and L^AT_EX source code for this document can be viewed at GitHub with the following link.

🔗 <https://github.com/BhorisDhanjal/MonteCarloMathsConstants>

2 Numerical Monte Carlo integration

In this section we will describe the algorithm for numerical Monte Carlo integration. Before this we will provide a short overview of the concept of Monte Carlo simulations.

2.1 Overview of Monte Carlo method

Monte Carlo simulations are a general class of computational algorithms that rely on random number sampling ¹. They are used to estimate numerical results for problems for which analytical solutions are either very difficult to obtain or impossible.

2.2 Naive Monte Carlo integration

We can demonstrate the theory behind naïve ² Monte Carlo integration in a simplified manner as below [1, 2].

Theorem 2.1. $\frac{1}{N} \sum_{i=1}^N f(X_i)$ is an unbiased estimator for $\int_{[0,1]^s} f(x) dx$

For points $x = (x_1, x_2, \dots, x_s)$ consider a function $f(x)$ over the unit hypercube $[0, 1]^s$ in \mathbb{R}^s .

$$\mathcal{I}[f] = \int_{[0,1]^s} f(x) dx$$

The approximation of $\mathcal{I}[f]$ is given by the expectation of $f(X)$ for some uniformly distributed random variable $X \in [0, 1]^s$.

$$\mathcal{I}[f] = E[f(X)] = E$$

Where an estimation for expectation is given by the sample mean,

$$\hat{\mathcal{I}} = \frac{1}{N} \sum_{i=1}^N f(X_i)$$

By the Law of Large numbers [3] we can state that,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(X_i) \rightarrow \mathcal{I}[f]$$

¹Note that for the majority of this project we will be utilizing pseudo-random numbers.

²We call it "naïve" since it utilizes uniformly distributed random variates, more sophisticated methods are discussed in the last section of this project.

Furthermore we can state that $\hat{\mathcal{I}}$ is an unbiased estimator since $E[\hat{\mathcal{I}}] = \mathcal{I}[f]$ for all N

Therefore, we can estimate $\mathcal{I}[f]$ with,

$$\mathcal{I}[f] \approx \hat{\mathcal{I}} = \frac{1}{N} \sum_{i=1}^N f(X_i) \quad \square$$

In general, for a function $f(x)$ over any volume V in \mathbb{R}^n we can say [4],

$$\mathcal{I}[f] = \int_V f(x) dx \approx V \frac{1}{N} \sum_{i=1}^N f(X_i) \quad (2.1)$$

We can simplify the above expression for the case of a one dimensional integral over some closed interval $[a, b]$ where $a, b \in \mathbb{R}^1$ as,

$$\int_a^b f(x) dx \approx (b - a) \frac{1}{N} \sum_{i=1}^N f(X_i) \quad (2.2)$$

2.3 Error of Naive Monte Carlo integration

Theorem 2.2. *The standard error of the estimator is inversely proportional to the square root of the number of random points chosen.*

Consider the variance for above described estimator $\hat{\mathcal{I}}$,

$$\begin{aligned} \sigma_{\hat{\mathcal{I}}}^2 &= \sigma^2 \left(\frac{1}{N} \sum_{i=1}^N f(X_i) \right) \\ &= \frac{1}{N^2} \sigma^2 \left(\sum_{i=1}^N f(X_i) \right) \end{aligned}$$

Since X_i are uniform iid r.v., the variance of the sum is equal to the sum of the variance.

$$\begin{aligned} &= \frac{1}{N^2} \sum_{i=1}^N \sigma^2(f(X_i)) \\ &= \frac{1}{N^2} N \sigma_{f(X_i)}^2 \\ \sigma_{\hat{\mathcal{I}}}^2 &= \frac{\sigma_{f(X_i)}^2}{N} \\ \sigma_{\hat{\mathcal{I}}} &= \frac{\sigma_{f(X_i)}}{\sqrt{N}} \\ \sigma_{\hat{\mathcal{I}}} &\propto \frac{1}{\sqrt{N}} \quad \square \end{aligned}$$

This implies that, in naïve Monte Carlo integration, reducing the error by half requires increasing the number of random points by a factor of 4.

We will explore variable reduction methods in the last section of this project. For the rest of this project however, we will only utilize naïve Monte Carlo integration that we implement using pseudo-random numbers.

2.4 An example of naïve Monte Carlo Integration

In order to better motivate Monte Carlo integration we will provide a simple visualization.

Example 1. Estimating $\int_2^5 100 - 8x^2 + x^3 dx$ with 5 random points.

We begin by generating 6 random real numbers between 2 and 5. We will evaluate these points in $f(x) = 100 - 8x^2 + x^3$ and plot it.

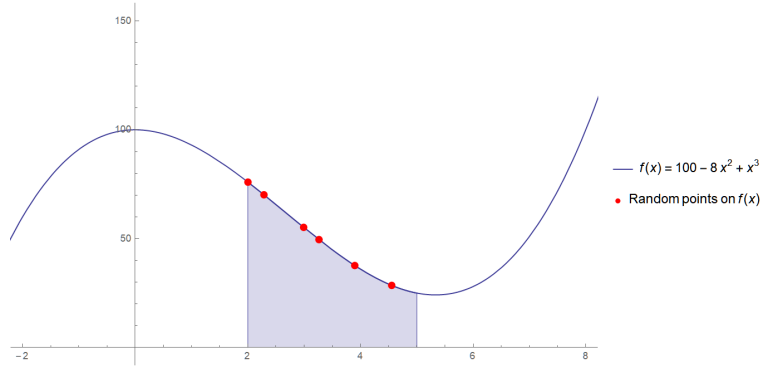


Figure 1: The plot of $f(x)$ and 6 random points plotted on it.

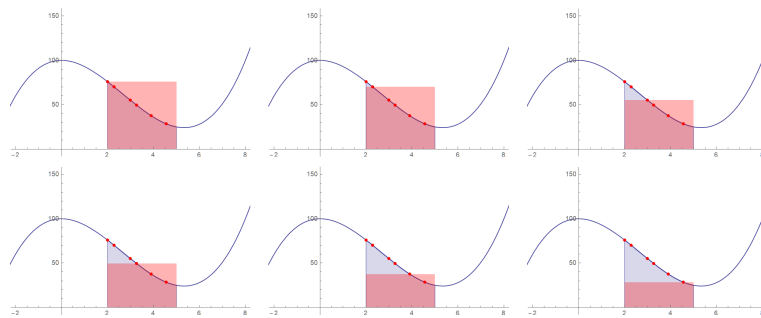


Figure 2: Rectangles of width $(b - a)$ with height corresponding to the random points.

Recall eq. 2.2, the multiplication of the $(b - a)$ term can be visualized as rectangles of width $(b - a)$ as above. It is now clear to see that taking the mean of these rectangle areas will give us an approximation to the integral that improves with the addition of more points.

3 Estimating Pi

In this section we will discuss the application of Monte Carlo techniques for estimating the value of pi.

3.1 Elementary Method

In order to estimate the value of pi using Monte Carlo simulations we will estimate the area of a unit circle inscribed in a unit square.

The experiment is conducted by taking random sample points in the region of the unit square. For an unbiased estimator of area of the circle it is assumed that the random sample points are uniformly distributed.

The bounding box area is $A_{box} = r^2 = 1$ and the area of the unit circle inscribed in the square is $A_{circle} = \pi r^2 = \pi$. With N_{box} total sample points and N_{circle} is the total sample points lying inside the unit circle.

Since the sample points are uniformly distributed within the bounding box, the ratio of the circle to the area of the bounding box is approximately equal to the ratio of the number of sample points falling in the circle to the number of points falling in the bounding box.

$$\frac{A_{pi}}{A_{box}} = \frac{\pi r^2}{r^2} = \pi \approx \frac{N_{pi}}{N_{box}} \quad (3.1)$$

An example of a single experiment is shown below.

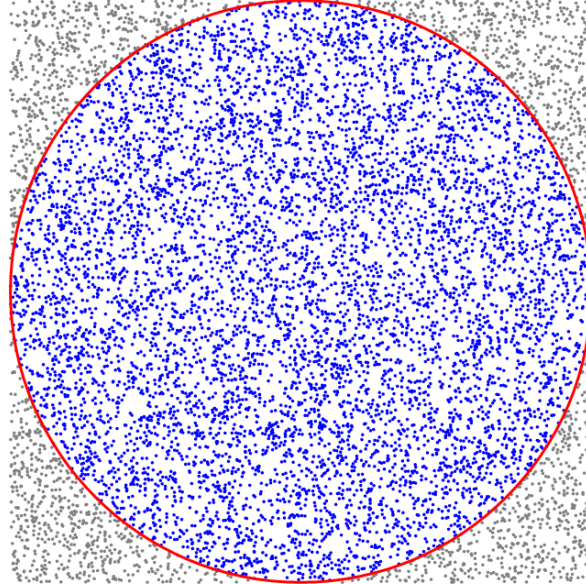


Figure 3: Experiment with 10,000 points, $N_{pi}/N_{box} = 3.1512$

We now repeat the experiment 10^3 times with 10^6 points. Taking the mean of these 10^3 experiments gives us the following approximation. $\pi \approx 3.14154 \pm 0.00165$, with $-1.696 \cdot 10^{-3} \%$ error.

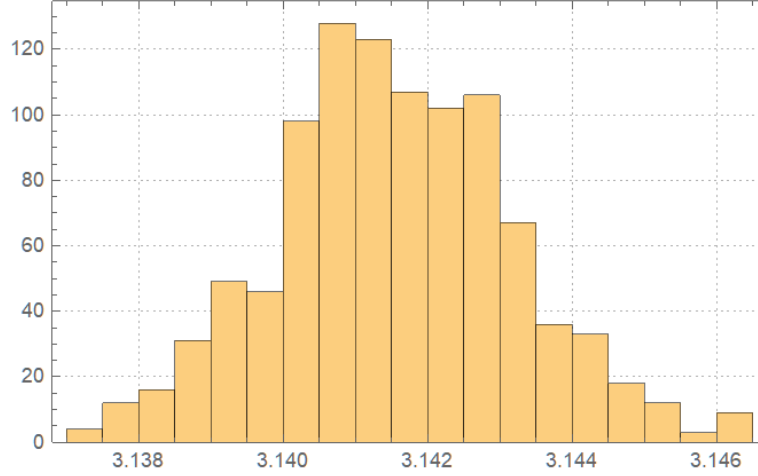


Figure 4: Histogram of experiment with 10^6 points, repeated 10^3 times.

3.2 Buffon's Needle

"A large plane area is ruled with equidistant parallel lines, the distance between two consecutive lines is 'a'. A thin needle of length $l < a$ is tossed randomly onto the plane. What is the probability that the needle will intersect one of the lines?"

This forms the basis of the Buffon's Needle problem. Which can be solved using elementary integral calculus [5]. with the help of this analytical solution, the experiment can be used to approximate π using Monte Carlo simulations.

Theorem 3.1. *The probability of a needle intersecting a line is given by $p = \frac{2l}{a\pi}$.*

For a given needle of length l we model the dropping of the needle on the ruled plane with parallel lines a units apart as follows.

Let x be the distance from the center of the needle to the nearest line and θ be the acute angle between the needle and the lines. Let x be a uniform random variable over the interval

The uniform probability density function of x between 0 and $\frac{a}{2}$ is

$$\begin{cases} \frac{2}{a} & 0 \leq x \leq \frac{a}{2} \\ 0 & \text{otherwise} \end{cases}$$

and θ a uniform random variable over the interval $(0, \frac{\pi}{2})$ with the probability density function:

$$\begin{cases} \frac{2}{\pi} & 0 \leq x \leq \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

The two random variables, x and θ , are independent Therefore the joint probability density function of (x, θ)

$$\begin{cases} \frac{4}{a\pi} & 0 \leq x \leq \frac{a}{2} \text{ and } 0 \leq \theta \leq \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

for the short needle case ($l \leq a$) A needle intersects a line if

$$\left(x \leq \frac{l}{2} \sin \theta \right)$$

The probability that the needle will intersect a line is

$$p = \int_0^{\frac{\pi}{2}} \int_0^{\frac{l}{2} \sin \theta} f(x, \theta) dx d\theta = \int_0^{\frac{\pi}{2}} \int_0^{\frac{l}{2} \sin \theta} \frac{4}{a\pi} dx d\theta = \frac{2l}{a\pi}$$

$$p = \frac{2l}{a\pi} \quad \square$$

If experiment has n needles out of which m needles intersect lines, then the value of p can be estimated as

$$\hat{p} = \frac{m}{n}$$

π can therefore be estimated by

$$\hat{\pi} = \frac{2nl}{ma} \quad (3.2)$$

An example of a single experiment with 1000 needles is shown below. We then repeat the experiment 10^3 times with 10^6 needles. Taking the mean of these 10^3 experiments gives us the following approximation. $\pi \approx 3.14018 \pm 0.04510$, with $-4.481 \cdot 10^{-2} \%$ error.

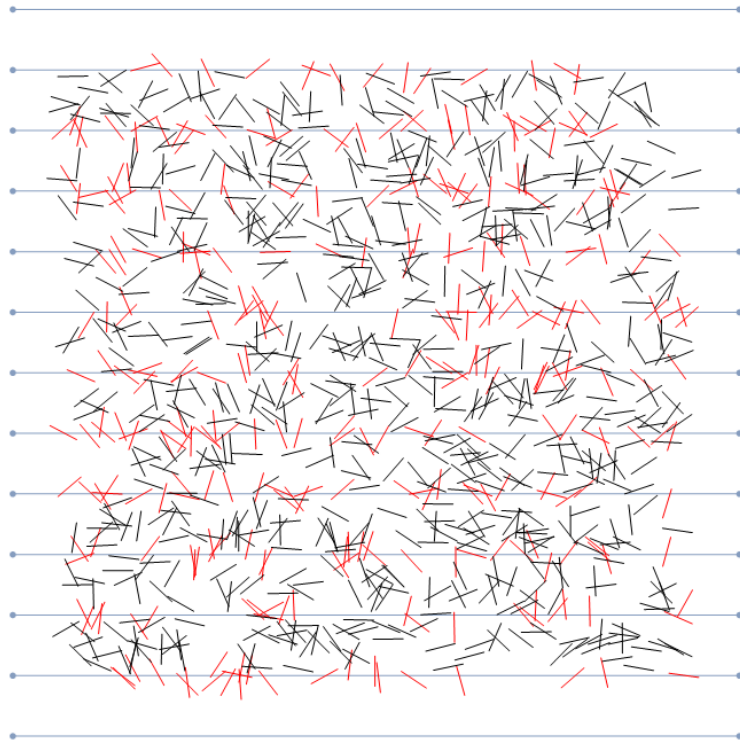


Figure 5: Experiment with 1000 needles, $\hat{\pi} = 3.14465$

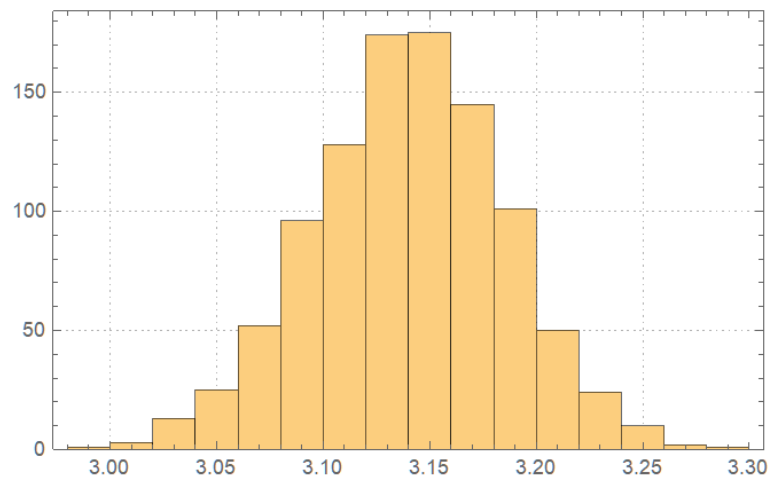


Figure 6: Histogram of Buffon's needle experiment with 10^4 needles, repeated 10^3 times.

3.3 Buffon's Noodle

Buffon's noodle poses the question, "What is the expected number of crossings if we don't drop a rigid needle but rather a flexible noodle?". In this section we will estimate π using Buffon's noodle problem.

For straight needle N there are n line crossing for n intersection of needles with the infinite grid and thus the expectation of line crossing as $E[N] = \frac{2a}{\pi d}$. where a is the length of needle. Using the corollary of this result, that $E[N]$ is a function of the length of the needle only, not its shape [6].

Theorem 3.2. *Let N be a noodle of length ' t ' thrown at random onto an infinite grid of parallel lines with common distance d between them. Then the expected number of line-crossings $E[N]$ is given by $E[N] = \frac{2t}{\pi d}$.*

We take into consideration the fact that, a noodle due to its shape could cross a line in more than one spot. In general if a needle intersects n times in m attempts, then a combination of a such needles is expected to be intersected na times in m attempts.

If this combination is formed by infinitesimally small needles, then it will become equivalent to a noodle.

Consider a sequence of polygonal lines L_1, L_2, \dots which approaches the curve of the noodle N uniformly. Let $L_{i1}, L_{i2}, \dots, L_{im}$ be the segments of the line L_i , such that $L_i = L_{i1} + L_{i2} + \dots + L_{im}$. Let t_{ij} be the length of L_{ij} such that $t_{ij} < d$ for all i and j .

The expectation of line crossing the grid be $E[L_{ij}]$ hence $E[L_{ij}] = 2t_{ij}/\pi d$. Now letting t_i be the length of L_i , we know that t_i approaches t as i tends to infinity. Hence, if $E[L_i] = E[L_j] + \dots + E[L_n]$ for all i , then

$$E[L_i] = \sum_{j=1}^n \frac{2t_j}{\pi d} = \frac{2}{\pi d} \sum_{j=1}^n t_j = \frac{2t}{\pi d} \quad (3.3)$$

hence in the limit

$$E[N] = \frac{2t}{\pi d} \quad \square$$

3.3.1 A note on the simulation of noodles

Simulating a "noodle" isn't a very straightforward task computationally, since "noodles" aren't very well defined. Here, we choose to simulate it by considering a line of length l divided into a random number of segments (between 5 and 10), and then randomizing the angles between the segments. One could also choose to make smooth noodles, by using bezier curves instead.

An example of a single experiment with 1000 noodles is shown below. We now repeat the experiment 250 times with 10^4 noodles. Taking the mean of these 250 experiments gives us the following approximation. $\pi \approx 3.18615 \pm 0.04602$, with a +1.418 % error.

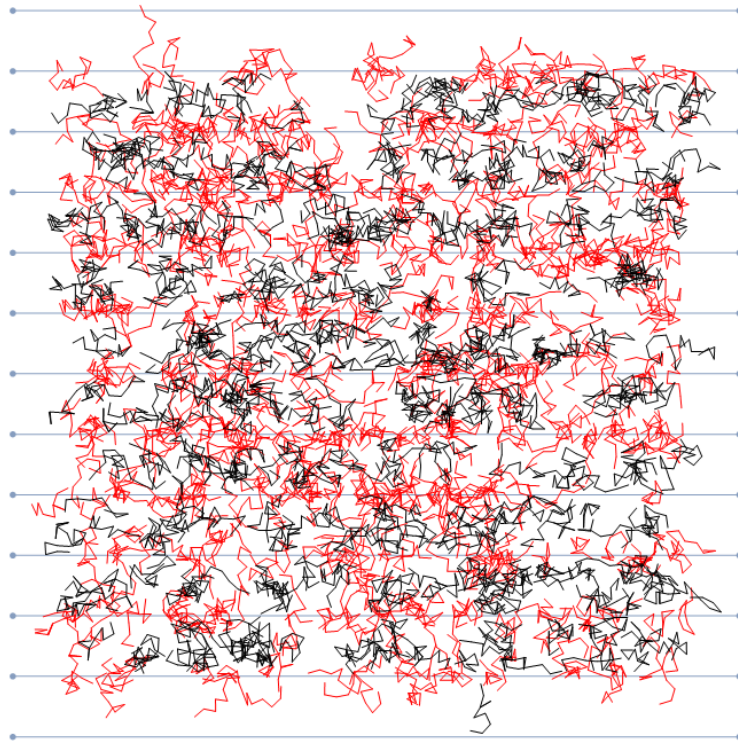


Figure 7: Experiment with 1000 noodles, $\hat{\pi} = 3.36984$

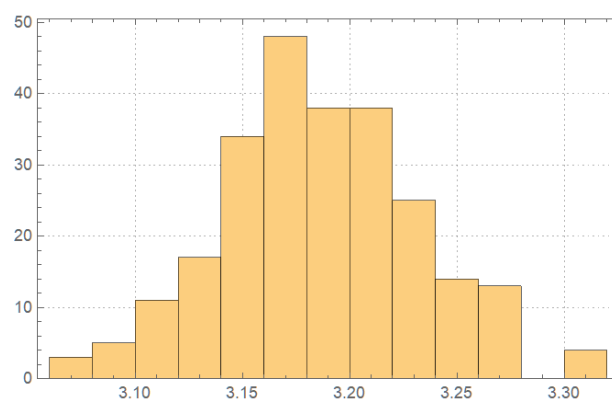


Figure 8: Histogram of Buffon's noodle experiment with 10^4 needles, repeated 250 times.

4 Estimating Euler's number

If $U_1, U_2 \dots U_n$ are some independent and uniformly distributed random variables over $(0,1)$. Let $S_n = \sum_{i=1}^n U_i$ and if N is the minimum value of n for which $S_n > 1$. The expectation of N is in fact equal to e .

This problem originated from as a exercise from a famous Russian textbook by Gnedenko [7]. For this section we will be drawing from two papers written by K. G. Russel on this topic [8, 9].

Theorem 4.1. *For iid uniform r.v. $U_1, U_2 \dots U_n$ over $(0,1)$, with $S_n = \sum_{i=1}^n U_i$ and if N is the minimum value of n for which $S_n > 1$, then $E[N] = e$.*

Note that our goal is to show that

$$E[N] = \sum_{n=2}^{\infty} nP(N = n) = e \quad (4.1)$$

If we can show that $P(N = n) = \frac{n-1}{n!}$, then we are done. Hence we continue as follows,

$$\begin{aligned} P(N = n) &= P(S_n > 1 \cap S_{n-1} < 1) \\ &= P(S_{n-1} < 1) - P(S_n < 1) \end{aligned}$$

We have further simplified our problem, now all that remains is to show that $P(S_n < 1) = \frac{1}{n!}$. This would complete the proof. We use the CDF of the Irwin-Hall distribution [10].

The Irwin-Hall distribution is defined as, the sum of n independent and identically distributed, uniform distributions on the unit interval. The CDF of Irwin-Hall distribution is given as ³,

$$F(x) = \frac{1}{n!} \sum_{k=0}^{\lfloor x \rfloor} (-1)^k \binom{n}{k} (x - k)^n$$

For $x=1$

$$F(1) = P(S_n < 1) = \frac{1}{n!} \sum_{k=0}^1 (-1)^k \binom{n}{k} (1 - k)^n = \frac{1}{n!} \quad \square$$

The astute reader may ask what is $E[N]$ if we consider $S_n > a$ for any real number a ? This general form is answered in the paper by Russell [9].

We now continue with the simulation to predict e by choosing 10^6 random uniform numbers between $(0,1)$ and counting the average number of times it takes for the sum to cross 1.

We now repeat the experiment 10^4 times with 10^6 numbers per experiment. Taking the mean of these 10^4 experiments gives us the following approximation. $e \approx 2.71827 \pm 0.00087$, with $-3.405 \cdot 10^{-4} \%$ error.

³This can be derived by integrating the pdf, which is in turn derived by induction by taking the convolution of $f_n * f$ to prove the $n + 1^{th}$ case.

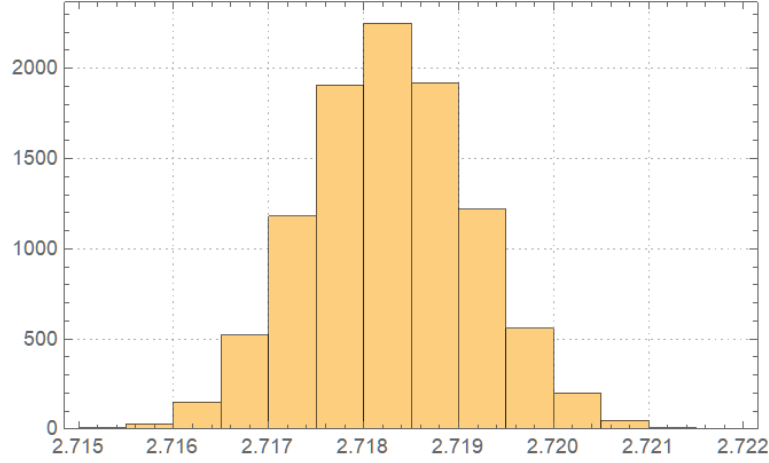


Figure 9: Histogram of e experiment with 10^6 numbers, repeated 10^4 times.

5 Estimating Euler-Mascheroni constant and Stieltjes constants

We begin this section by defining the Euler-Mascheroni constant and Stieltjes constants.

The Euler Mascheroni Constant (γ) is defined as the limiting difference between the harmonic series and the natural logarithm, i.e.

$$\begin{aligned}\gamma &= \lim_{n \rightarrow \infty} (H_n - \log n) \\ &= \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{1}{k} - \log n \right)\end{aligned}\tag{5.1}$$

Stieltjes constants can be thought of as a generalization of the Euler-Mascheroni constant, defined as,

$$\gamma_n = \lim_{m \rightarrow \infty} \left(\sum_{k=1}^m \frac{(\log k)^n}{k} - \frac{(\log m)^{n+1}}{n+1} \right)\tag{5.2}$$

It is clear to see that $\gamma_0 = \gamma$. It is interesting to note that Stieltjes constants were originally discovered in the Laurent series expansion (about 1) of the Riemann Zeta function [11].

$$\zeta(z) = \frac{1}{z-1} + \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \gamma_n (z-1)^n$$

We can in fact further generalize Stieltjes constants by considering the Laurent series of the Hurwitz-Zeta function instead of Riemann zeta. However, we will not further examine this.

5.1 Euler-Mascheroni constant

We will estimate the Euler-Mascheroni constant by first approximating the Gumbel distribution [12].

The Gumbel distribution accepts two parameters, α and β . It is defined by the following probability density function,

$$\begin{aligned} P(x) &= \frac{1}{\beta} e^{\frac{x-\alpha}{\beta} - e^{\frac{x-\alpha}{\beta}}} \\ &= \frac{1}{\beta} \exp \left(\frac{x-\alpha}{\beta} - \exp \left(\frac{x-\alpha}{\beta} \right) \right) \end{aligned}$$

and the following cumulative distributive function,

$$F(x) = 1 - \exp \left(- \exp \left(\frac{x-\alpha}{\beta} \right) \right)$$

It can be seen using integral representation of γ that the mean of the Gumbel distribution is,

$$\mu = \alpha - \gamma\beta$$

Our goal is therefore to simulate a Gumbel distribution with parameters $\alpha = 0$ and $\beta = -1$. The mean of this would therefore be γ .

We simulate the Gumbel distribution by using its quantile function (i.e. inverse CDF).

Calculating the inverse function for $F(x)$ with $\alpha = 0, \gamma = -1$ we get.

$$Q(p) = \log(-\log(p)) \tag{5.3}$$

Where p takes values between 0 and 1. Taking n random points p between 0 and 1, and taking the mean of their respective $Q(p)$ we approximate γ .

Taking 10^6 points and repeating the experiment 10^4 times gives us the following approximation $\gamma \approx 0.57721 \pm 0.00128$. With $-4.824 \cdot 10^{-4}\%$ error.

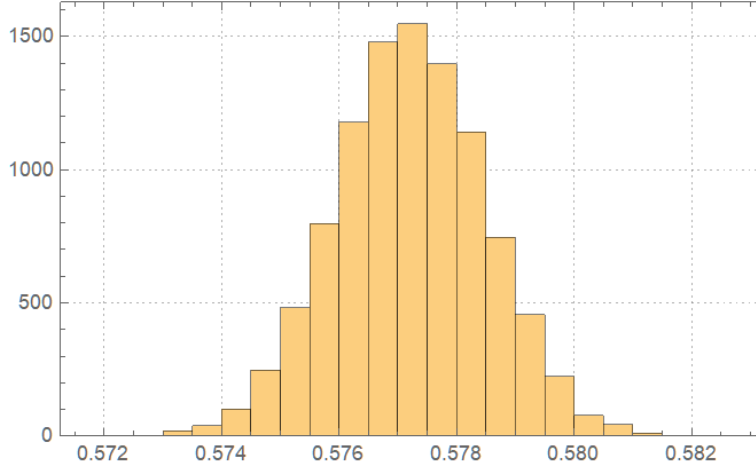


Figure 10: Histogram of experiment with 10^6 points, repeated 10^4 times.

5.2 Stieltjes constants

Calculating Stieltjes constants is not a trivial task, there exist many infinite series representations of Stieltjes constants which converge very slowly [11] (see eq.12). Extensive work has also been done in implementing fast algorithms to compute Stieltjes constants with high accuracy [13].

We will approximate a few Stieltjes constants by estimating the following integral definition.

$$\gamma_n = \frac{(-1)^n n!}{2\pi} \int_0^{2\pi} e^{nix} \zeta(e^{ix} + 1) dx \quad (5.4)$$

The above integral is derived by using Cauchy's differentiation formula on

$$(-1)^n \gamma_n = \frac{d^n}{dz^n} \left(\zeta(z) - \frac{1}{z-1} \right) \Big|_{z=1}$$

which is in turn obtained by comparing the Laurent series of the Zeta function with a traditional Taylor series.

The reader should note that calculating Stieltjes constants with Monte Carlo integration is not at all practical due to the low accuracy. For practical approximations other forms of numerical integration techniques are used [13, 14].

We will estimate eq. 5.4 using the naïve Monte Carlo integration described in the first section. The following table shows the results we achieved.

γ_n	No. of points chosen	No. of times repeated	Mean and Standard Deviation	Real values	Percentage Error
1	10^5	10^4	-0.0728663 ± 0.00258	-0.0728159	$-2.00069 \cdot 10^{-2}$
2	10^5	10^4	-0.00969426 ± 0.00524	-0.0096904	$+4.02199 \cdot 10^{-2}$
3	10^5	$1.5 \cdot 10^4$	$+0.00204329 \pm 0.01551$	+0.0020538	$-5.13216 \cdot 10^{-1}$
4	10^5	$3 \cdot 10^4$	$+0.00233049 \pm 0.06136$	+0.0023253	$+2.20283 \cdot 10^{-1}$

Table 1: Computed values of the first 4 Stieltjes constants (not including 0)

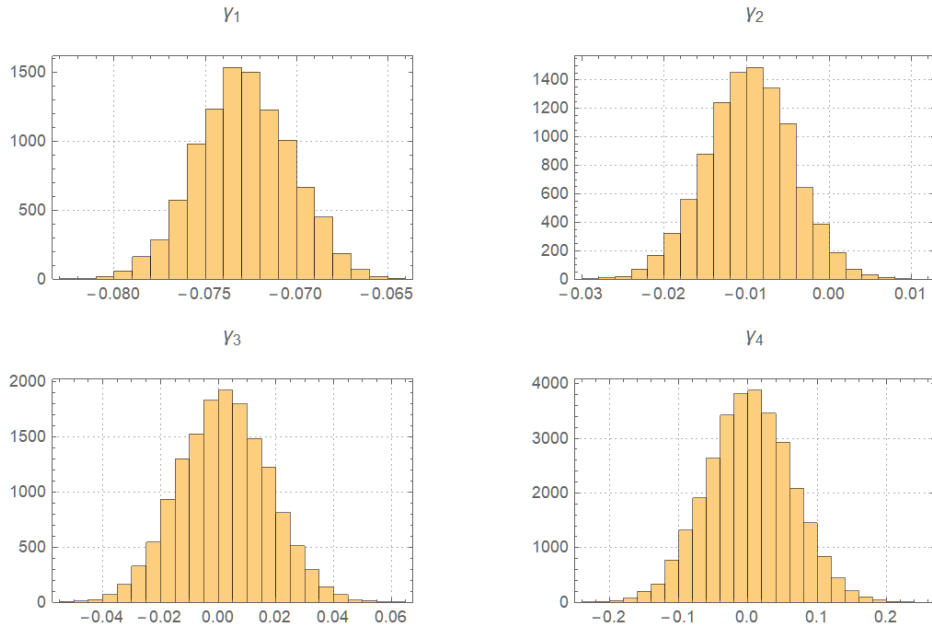


Figure 11: Histogram of the 4 experiments.

6 Estimating Phi

We conclude the estimation of mathematical constants on a lighter note.

Since the golden ratio (φ) is an algebraic number, estimating φ is very trivial. Since we know that $\varphi = \frac{1+\sqrt{5}}{2}$, we can simply choose an integral which evaluates to the desired result. We will choose $\int_4^5 \frac{3}{2} + \frac{1}{4\sqrt{x}} dx$ which is evidently equal to φ .

We estimate the integral with the method described in Section 1.

We repeat the experiment 10^4 times with 10^6 points per experiment. Taking the mean of these 10^4 experiments gives us the following approximation, $\varphi \approx 1.61803 \pm 3.79948 \cdot 10^{-6}$, with $6.544 \cdot 10^{-7} \%$ error. In fact the approximation is correct up to 7 decimal places!

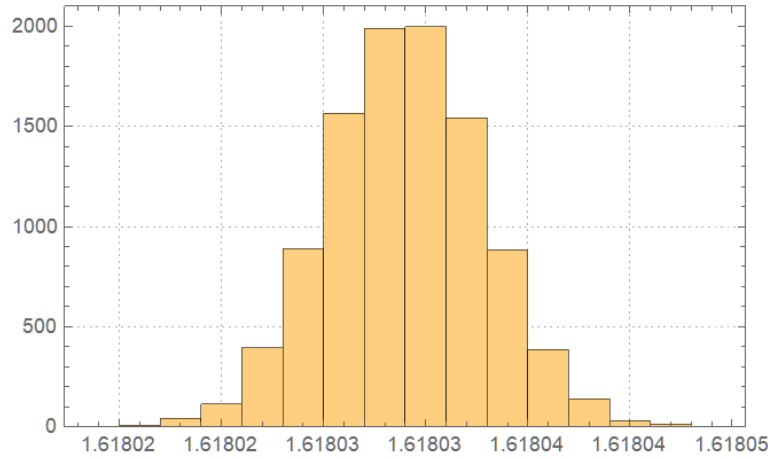


Figure 12: Histogram of experiment with 10^6 points, repeated 10^4 times.

7 Methods of integral variance reduction

Throughout the project so far we have only considered what is known as "naïve" Monte Carlo integration. We will now justify its nomenclature by listing out a few modified methods of Monte Carlo integration that are so designed to reduce variability [1, 15].

To demonstrate the effectiveness of these method we will be estimating the following integral,

$$\int_0^1 \int_0^1 \frac{\sin x \sin y}{xy} dx dy \quad (7.1)$$

It is a well known fact that $\int_0^z \frac{\sin x}{x} dx$ has no elementary closed form, in fact it is just represented by $Si(z)$ for simplicity. It is then clear to see that eq. 7.1 is equal to $Si(1)^2$.

7.1 Antithetic variates

Antithetic variates is a technique for variance reduction, in which we use the antithetic pair $(-x_1, \dots, -x_n)$ of pre-obtained samples (x_1, \dots, x_n) in the Monte Carlo simulation. The advantage of this technique is that, it reduces the number of samples to be taken to generate N integration nodes, and thus, reduces the variance of the sample

We will give a short proof of the above statement.

Theorem 7.1. *Random variable X and Y , on the same probability space are antithetic if they have same distribution and their covariance is negative.*

Take two variable X and Y . the variance of the average of x, y is

$$\text{Var} \left[\frac{1}{2}(x + y) \right] = \frac{1}{4}[\text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)]$$

since we know that,

$$\text{Cov}(X, Y) = \rho \sqrt{\text{Var}X} \sqrt{\text{Var}Y}$$

where ρ is the correlation between X and Y . It can be clearly seen that $\text{Var}(X+Y)$ is minimised when $\rho = -1$ i.e if X and Y are perfectly negatively correlated.

□

Now we will compare antithetic variables vs naïve Monte Carlo with eq. 7.1.

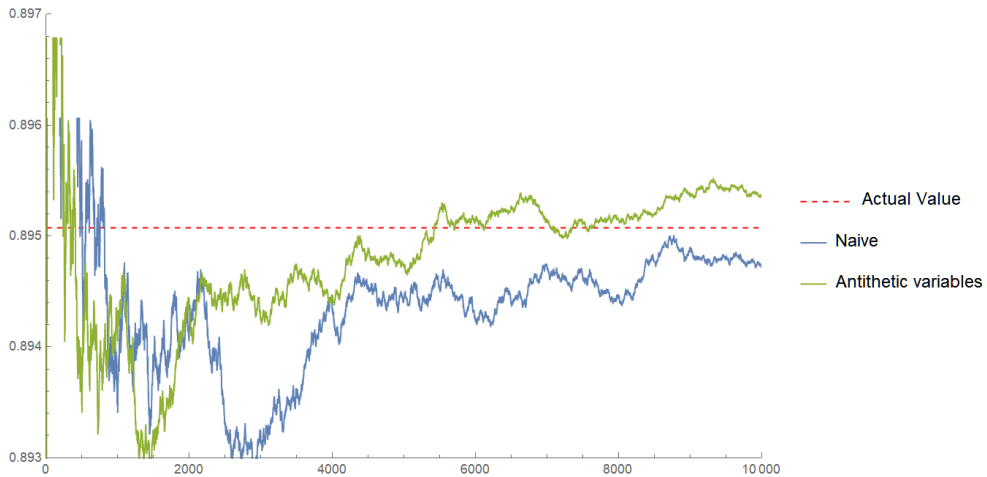


Figure 13: Comparison of estimating e.q. 7.1 with naïve Monte Carlo vs Monte Carlo with antithetic variables (for points ranging from 1 to 10,000).

7.2 Control variates

In this variance reduction technique, we use a known integral with a value approximately equal to the unknown integral to reduce the error in the estimate of unknown integral.

Theorem 7.2. $I[f] \approx \frac{1}{N} \sum_{i=1}^N (f(X_i) - g(X_i)) + \theta$

Suppose we want to find $\mu = E[f(x)]$. We find some $\theta = E[g(x)]$ where $g(x) \approx f(x)$, hence $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N f(X_i)$ and $\hat{\theta} = \frac{1}{N} \sum_{i=1}^N g(X_i)$ for a r.v. X . We can estimate μ by the following estimator,

$$\hat{\mu}_{\text{diff}} = \frac{1}{N} \sum_{i=1}^N (f(x_i) - g(x_i)) + \theta$$

This estimator is motivated due to the linearity of integrals. The expected value of $\hat{\mu} = \mu$, since $E[\hat{\theta}] = \theta$.

$$\text{Var}(\hat{\mu}_{\text{diff}}) = \frac{1}{N} \text{Var}(f(X) - g(X))$$

If $f(X) \approx g(X)$, then,

$$\text{Var}(f(X) - g(X)) < \text{Var}(f(X))$$

Hence, we use $(\hat{\mu}_{\text{diff}})$ to reduce the variance. Here $g(X)$ is known as the control variate. □

We will now compare Monte Carlo integration in estimating eq. 7.1 using both control variates as well as antithetic variables.

To generate a known integral $g(x)$ which is close to $f(x)$ we will employ the use of Taylor series. Finding the Taylor series of $f(x)$ about 0, we estimate the integrand as follows,

$$\frac{\sin x \sin y}{xy} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} x^n y^m \frac{\cos \frac{m\pi}{2} \cos \frac{n\pi}{2}}{(m+1)!(n+1)!}$$

Using $n=m=0$ to $n=m=2$ as a approximation we get,

$$\frac{\sin x \sin y}{xy} \approx 1 - \frac{x^2}{6} - \frac{y^2}{6} + \frac{x^2 y^2}{36}$$

We therefore use this approximation as $g(x)$,

$$g(x) = \int_0^1 \int_0^1 1 - \frac{x^2}{6} - \frac{y^2}{6} + \frac{x^2 y^2}{36} dx dy = \frac{289}{324} \quad (7.2)$$

Finding a Taylor series is obviously not practical for most integrands, but even very rough approximations will significantly reduce variability.



Figure 14: Comparison of estimating e.q. 7.1 with control variate Monte Carlo vs Monte Carlo with antithetic variables and control variate (for points ranging from 1 to 10,000).

To put this the effectiveness of the control variate method in perspective we shall see it in comparison to the naïve method.

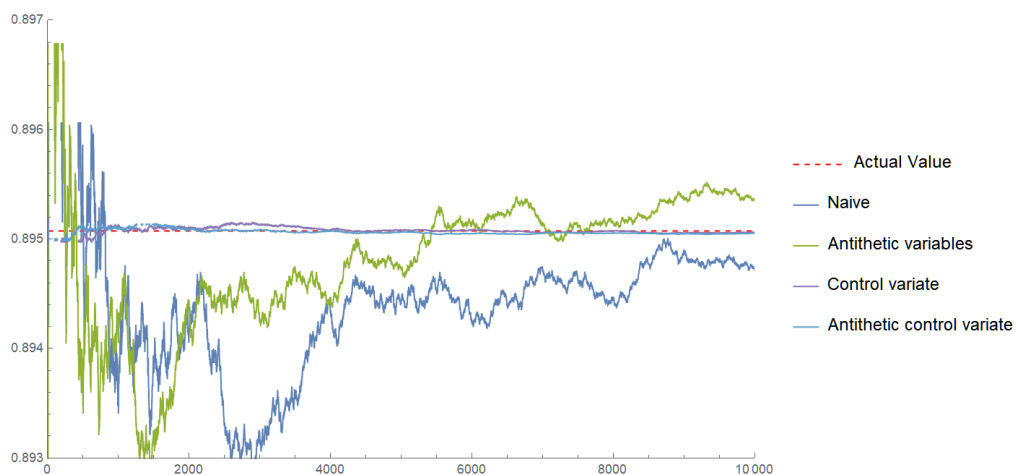


Figure 15: Comparison of the control variate with naïve Monte Carlo.

The reduction in variance is now much more striking.

8 Quasi Monte-Carlo integration

Quasi-Monte Carlo integration, simply put, is Monte Carlo integration using quasi-random generated numbers to form the integration nodes (as opposed to pseudo-random in naïve integration) [16].

The reader should note that in this section, we will only provide a cursory overview of Quasi Monte-Carlo integration and its results. Detailed proofs will not be given as they are far beyond the scope of this project. A vast amount of the theory surrounding this topic has been carried out by H. Niederreiter in a series of papers [17, 18] and books [19] which we recommend for further reading.

8.1 Quasi-random sequences (low-discrepancy sequences)

Quasi-random numbers are numbers generated from quasi-random sequences, which are in turn defined as sequences with low discrepancy [20]. We will first define what is meant by discrepancy.

Discrepancy can be intuitively thought of as a measure of uniformity. We will now see the formal definition for discrepancy for a one-dimensional sequence [17].

Let x_1, x_2, \dots, x_N be N numbers in $I = [0, 1]$, let $E \subset I$. Define a function $A(E; N)$ such that, it counts the number of $n, 1 \leq n \leq N$, for $x_n \in E$. The discrepancy D_n of the N numbers in I is then given by,

$$D_n = \sup_J \left| \frac{A(J; N)}{N} - |J| \right| \quad (8.1)$$

Where, J is any sub-interval of I . The star discrepancy (D_N^*) is a more commonly used definition which is as follows.

$$D_N^* = \sup_{0 \leq t \leq 1} \left| \frac{A([0, t]; N)}{N} - t \right| \quad (8.2)$$

These above definitions are easily generalized to $[0, 1]^s$ by calculating $|J|$ with the respective metric.

D_n and D_n^* are related by the following inequality [17]

$$D_n^* \leq D_n \leq 2^s D_n^* \quad (8.3)$$

8.2 Error in Quasi Monte-Carlo integration

Niederreiter has shown [17] that for any dimension $s \geq 1$, there exists an infinite sequence of points in I^s such that, for some constant c_s only dependent on dimension,

$$D_N = O\left(c_s \frac{\log N^s}{N}\right)$$

The famous Koksma–Hlawka inequality states that [17],

$$\left| \frac{1}{N} \sum_{n=1}^N f(x_n) - \int_{[0,1]^s} f(x) dx \right| \leq V(f) D_N^*$$

Where, $V(f)$ is the bounded total variation of the function ⁴.

The Koksma-Hlawka inequality along with eq. 8.3 is used to show that the error of Quasi Monte-Carlo integration is given by $O\left(c_s \frac{\log N^s}{N}\right)$ as opposed to $O\left(\frac{1}{\sqrt{N}}\right)$ when using random numbers.

It is then clear to see that for a sufficiently small dimension s and c_s the error by Quasi Monte-Carlo integration is much lower than naïve Monte-Carlo integration.

8.3 Examples of Quasi-Random sequences

We have so far restrained from giving any examples of methods used to generate quasi-random sequences. In this section we will describe a few important methods ⁵.

8.3.1 Van der Corput and Halton sequence

We begin by describing Van der Corput sequences [19], which are conceptually the easiest to understand.

For any integer $n \geq 0$ as follows, we define the following function, known as the radical-inverse function,

$$\phi_b(n) = \sum_{j=1}^m a_j b^{-j} = \frac{a_1}{b} + \frac{a_2}{b^2} \dots \quad (8.4)$$

for a_j ranging 0 to $b - 1$, $m, b \in \mathbb{Z}$ and $b \geq 2$. The reader should note the similarity of this function with the base b representation of a number. $\phi_b(n)$ is clearly contained in $[0, 1]$.

⁴Note that this is not variance. Discussing total variation rigorously is well beyond the scope of this project, it can be roughly thought of as a infinitesimal absolute value.

⁵We will not be proving the values of discrepancies of these examples.

The Van der Corput sequence is defined for any range of n 's as,

$$x_n = \phi_2(n) \quad (8.5)$$

With star discrepancy equal to,

$$D_N^* = O\left(\frac{\log N}{N}\right) \quad (8.6)$$

Below we enumerate the first 5 values of the Van der Corput sequence.

n	n_2	$\phi_2(n)$
0	0	0
1	1	$0.1_2 = 1/2_{10}$
2	10	$0.01_2 = 1/4_{10}$
3	11	$0.11_2 = 3/4_{10}$
4	100	$0.001_2 = 1/8_{10}$

Table 2: First 5 values of Van der Corput sequence

However, the limitation of Van der Corput should now be clear to see. Van der Corput sequences only generates single dimensional numbers.

Halton sequences [21] are a very natural extension of Van der Corput sequences for higher dimensions s , which we define as,

$$x_n^s = (\phi_{p_1}(n), \phi_{p_2}(n), \dots, \phi_{p_s}(n)) \quad (8.7)$$

Where, $p_1, p_2 \dots p_s$ must be co-prime integers. It is common to simply use the n^{th} prime number at each stage. The reason for this specific selection of bases is to reduce the dimension dependent constant in the discrepancy.

The star discrepancy of Halton sequences is, as expected,

$$D_N^* = O\left(\frac{\log N^s}{N}\right) \quad (8.8)$$

It has been theorized that Van der Corput and Halton sequences have the lowest discrepancy of any quasi-random sequence. This is however not yet been proved. Nevertheless, Halton sequences are very rarely used for numeric integration due to its tendency of becoming overly predictable in higher dimensions.

An early fix for this problem was given by Faure [22] who proposed to use bases in the Halton sequence as the first prime number greater than dimension s , and further to re-order each element using certain permutations based on Pascal matrices.

Below we have displayed a 2-dimensional Halton sequence with 1500 points using base 2 and 3. To illustrate the predictable nature for high dimensions we will also display a 2-dimensional graph using the 10th and 11th prime respectively.

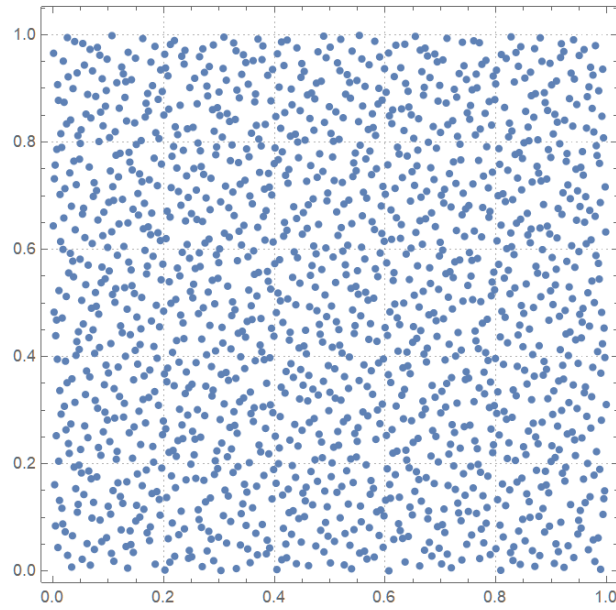


Figure 16: 2-D Halton sequence with 1500 points using 2^{nd} and 3^{rd} prime as bases.

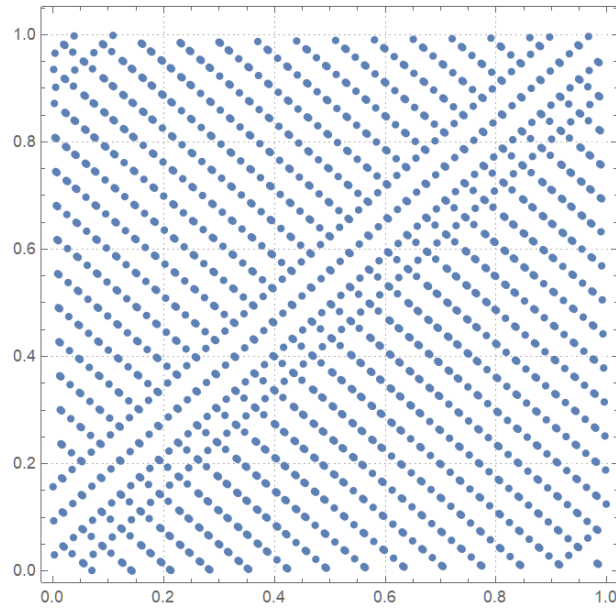


Figure 17: 2-D Halton sequence with 1500 points using 10^{th} and 11^{th} prime as bases. This helps visualize the problem of overly predictive points that occur in higher dimensions.

8.3.2 A cursory view of (t,s) sequences

Before expounding on a few other types of quasi-random sequences (namely, Sobol and Niederreiter) we will first list out a few important definitions in a summarized manner. Throughout this section we will extensively refer to a paper by J. Dick and H. Niederreiter [23].

Recall the definition of the radical-inverse function (see e.q. 8.4). For some $\mathbf{x} \in [0, 1]^s$ consider,

$$[\mathbf{x}]_{b,m} = (\phi_b(x^1), \dots, \phi_b(x^s))$$

Here m is the upper bound of summation. Note the difference between the Halton sequence, here we use a fixed base b .

For integers $b \geq 2, s \geq 1$ and $t \geq 0$. Consider a sequence of points x_0, x_1, \dots , in $[0, 1]^s$. This sequence of points is called a **(t, s) sequence** in base b . If for all integers $k \geq 0$ and $m > t$, the set $\mathcal{P}_{k,m}$ consisting of the points $[x_n]_{b,m}$ with $kb^m \leq n < (k+1)b^m$ has the following property- *for every half-open sub-interval E of $[0, 1]^s$ of the following form*

$$E = \prod_{i=1}^s [a_i b^{-c_i}, (a_i + 1) b^{-c_i}) \quad (8.9)$$

for some integers $c_i > 0$ and $0 \leq a_i < b^{c_i}$ for $1 \leq i \leq s$ and volume b^{t-m} , there are exactly b^t points of $\mathcal{P}_{k,m}$ contained in E .

For the purposes of this project we are not providing detailed motivation for the concept of (t,s) sequences, we state it in a simplified definition since both Niederreiter and Sobol sequences are in fact (t,s) sequences.

8.3.3 Niederreiter sequence

Niederreiter sequences are considered to be the best quasi-random sequence for numerical integration. Here we shall provide a very brief overview of Niederreiter sequences [23, 18] without going into details of the construction, with an example at the end.

Consider for some fixed dimension $s \geq 1$, consider a list of pairwise co-prime non-constant polynomials belonging to any finite field of order q ,⁶ with $e_i = \deg(p_i) \geq 1$ for $1 \leq i \leq s$, i.e., $p_1, p_2, \dots, p_s \in \mathbb{F}_q$.

The generating matrices for the Niederreiter sequence are constructed using the constants of the Laurent series expansions over these polynomials and objects known as (t,m,s) nets⁷.

⁶With the ordinary restriction that q is a power of some prime.

⁷We have refrained from defining (t,m,s) nets as we are not providing the full proof. They are closely related to (t,s) sequences.

For more detailed expositions refer to the papers by Niederreiter [23, 24, 18]. The matter of note is that the Niederreiter sequences in fact form (t,s) sequences with the exact value of t given as follows,

$$t = \sum_{i=1}^s (e_i - 1)$$

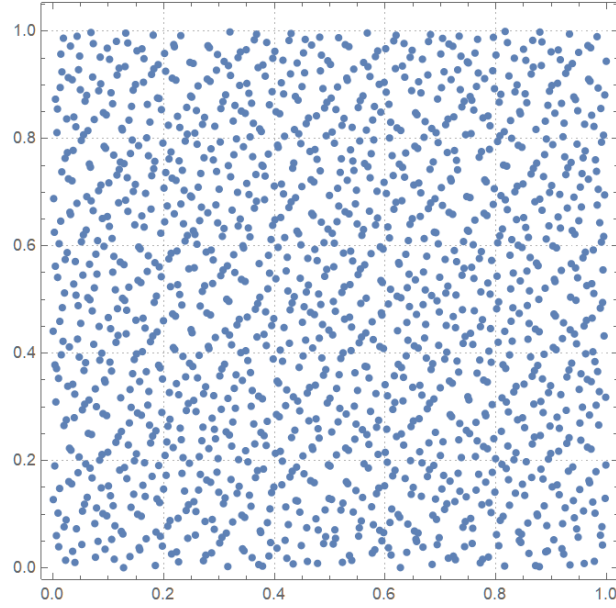


Figure 18: An example of a 2-Dimensional Niederreiter sequence with 1500 points.

8.3.4 Sobol sequence

Sobol sequences can be thought of as a special case of the Generalized Niederreiter sequence [23, 25]. It also uses polynomials to generate the quasi-random sequence.

Sobol sequences can be thought of as a (t,s) sequences in base 2 that uses polynomials as such, $p_1(x) = x \in \mathbb{F}_2[x]$ and p_2, \dots, p_s being primitive polynomials of successive order arranged in non-decreasing degrees to construct the generating matrices.

Note the difference, Sobol uses primitive polynomials in base 2 while Niederreiter uses any pairwise co-prime polynomials in any base.

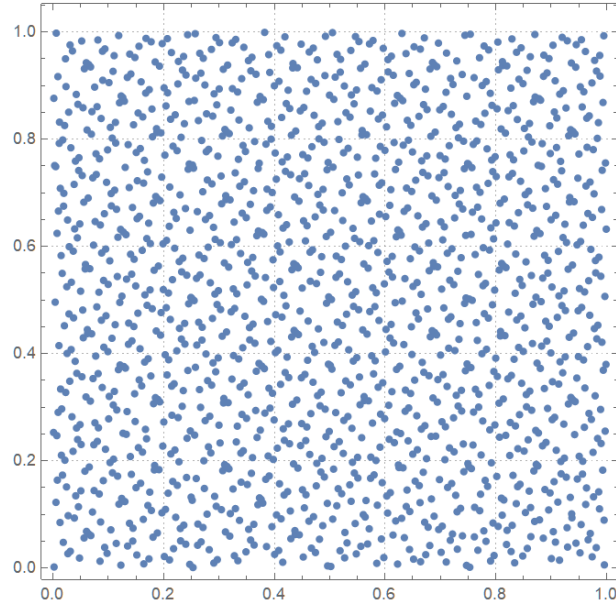


Figure 19: An example of a 2-Dimensional Sobol sequence with 1500 points.

8.4 Computing Apéry's constant

We will end the project by computing a final example using Quasi Monte Carlo integration.

Apéry's constant is $\zeta(3) = \sum_{n=0}^{\infty} \frac{1}{3^n}$. A simple integral representation of Apéry's constant is given as follows,

$$\zeta(3) = \int_0^1 \int_0^1 \int_0^1 \frac{1}{1 - xyz} dx dy dz$$

For comparisons sake we will estimate the integral with Quasi Monte-Carlo and naïve Monte Carlo to provide a direct comparison. This is shown in the graph below.

Repeating the experiment using the Niederreiter sequence with 10^6 points repeated 10^2 times we obtain the following results.

$\zeta(3) \approx 1.202055 \pm 3.095 \cdot 10^{-5}$, with a percentage error of $-1.514 \cdot 10^{-4}\%$.

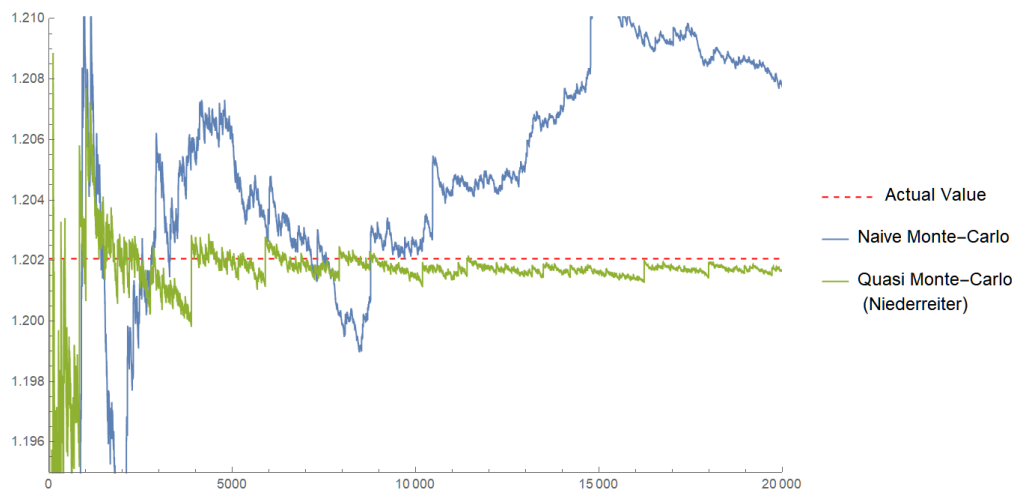


Figure 20: Quasi Monte Carlo vs naïve Monte Carlo estimates for Apéry's constant (no. of points ranging from 1 to 20,000).

References

- [1] Russel E. Caflisch. Monte carlo and quasi-monte carlo methods. *Acta Numerica*, 7:1–49, 1998.
- [2] Stefan Weinzierl. Introduction to monte carlo methods. arXiv:hep-ph/0006269, 2000.
- [3] Eric W. Weisstein. Strong law of large numbers From MathWorld—A Wolfram Web Resource. <https://mathworld.wolfram.com/StrongLawofLargeNumbers.html>. Last visited on July 15, 2021.
- [4] Eric W. Weisstein. Monte carlo integration From MathWorld—A Wolfram Web Resource. <https://mathworld.wolfram.com/MonteCarloIntegration.html>. Last visited on July 15, 2021.
- [5] Maximilian J. Wang and Jin Wang. Removing the inherent paradox of the buffon's needle monte carlo simulation using fixed-point iteration method. In *Proceedings of the Winter Simulation Conference 2014*, pages 3674–3683, 2014.
- [6] J. F. Ramaley. Buffon's noodle problem. *The American Mathematical Monthly*, 76(8):916–918, 1969.
- [7] B. V. Gnedenko. *The Theory of Probability*, page 194. Mir Publishers, 1978. Ex. 22.
- [8] K. G. Russell. Estimating the value of e by simulation. *The American Statistician*, 45(1):66–68, 1991.

- [9] K. G. Russell. On the number of uniform random variables which must be added to exceed a given level. *Journal of Applied Probability*, 20(1):172–177, 1983.
- [10] Philip Hall. The Distribution Of Means For Samples Of Size N Drawn From A Population In Which The Variate Takes Values Between 0 And 1, All Such Values Being Equally Probable. *Biometrika*, 19(3-4):240–244, 12 1927.
- [11] Eric W. Weisstein. Stieltjes constants From MathWorld—A Wolfram Web Resource. <https://mathworld.wolfram.com/StieltjesConstants.html>. Last visited on July 15, 2021.
- [12] Eric W. Weisstein. Gumbel distribution From MathWorld—A Wolfram Web Resource. <https://mathworld.wolfram.com/GumbelDistribution.html>. Last visited on July 15, 2021.
- [13] Fredrik Johansson and Iaroslav Blagouchine. Computing stieltjes constants using complex integration, 2018.
- [14] Rick Kreminski. Newton-cotes integration for approximating stieltjes (generalized euler) constants. *Mathematics of Computation*, 72(243):1379–1397, 2003.
- [15] R. C. H. Cheng. The use of antithetic variates in computer simulations. *The Journal of the Operational Research Society*, 33(3):229–237, 1982.
- [16] William J. Morokoff and Russel E. Caflisch. Quasi-monte carlo integration. *Journal of Computational Physics*, 122(2):218–230, 1995.
- [17] Harald Niederreiter. Quasi-monte carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society*, 84:957–1041, 1978.
- [18] Harald Niederreiter. Low-discrepancy and low-dispersion sequences. *Journal of Number Theory*, 30(1):51–70, 1988.
- [19] L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences*. A Wiley-Interscience publication. Wiley, 1974.
- [20] William J. Morokoff and Russel E. Caflisch. Quasi-random sequences and their discrepancies. *SIAM Journal on Scientific Computing*, 15(6):1251–1279, 1994.
- [21] J. H. Halton. Algorithm 247: Radical-inverse quasi-random point sequence. *Commun. ACM*, 7(12):701–702, December 1964.

- [22] H. Faure. Using permutations to reduce discrepancy. *Journal of Computational and Applied Mathematics*, 31(1):97–103, 1990.
- [23] Josef Dick and Harald Niederreiter. On the exact t-value of niederreiter and sobol' sequences. *Journal of Complexity*, 24(5):572–581, 2008.
- [24] Harald Niederreiter. Constructions of (t,m,s)-nets and (t,s)-sequences. *Finite Fields and Their Applications*, 11(3):578–600, 2005. Ten Year Anniversary Edition!
- [25] I.M Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967.