Monte Carlo Methods - Random Walks

Zukisa Mkhize

8 September 2024

Contents

1 Introduction

1.1 Purpose of the Report

The purpose of this report is to utilize Monte Carlo methods to investigate various types of random walks , focusing mainly on their statistical properties and computational efficiencies.In this report we will explore the scaling behavior of the squared displacement $\langle R_N^2 \rangle$ as a function of the number of steps N for different random walk models , such as the simple random walk and self avoiding walks.
The report will delve into the efficiency of generating using two distinct approaches: rejection sampling, where self-intersecting walks are discarded, and myopic sampling, where the walker chooses its next step based on unvisited neighboring sites. The fraction of non-reversing walks retained during rejection sampling will be analyzed to estimate the number of self-avoiding walks for various step lengths, ultimately leading to an estimate of the connective constant

Finally, the distribution of walks and the impact of different sampling techniques, including self-normalizing importance sampling (Rosenbluth method), will be examined to assess their accuracy and consistency in comparison with traditional methods. This comprehensive investigation will provide insights into the characteristics of random walks and the effectiveness of Monte Carlo methods in simulating such stochastic processes.

1.2 Overview of Monte Carlo Methods

Monte Carlo methods are computational algorithms that utilize random sampling to obtain numerical results.The algorithms use randomness to solve problems that are deterministic in nature.Monte Carlo method are use for different problems such as optimization , numerical integration and generating draws from a probability distribution.These methods are implemented using computer simulation, and they can approximate solutions to problems that are too complex to analyze mathematically.[3]

In statistical physics , Monte Carlo methods are used to study systems with large number of interacting particle , simulating different phenomena by sampling states of a system according to a probability distribution.An example would be using the Metropolis algorithm to simulate the Ising model.

1.3 Importance of Random Walks

Random walks also known as drunkards are stochastic process with a widespread application across different fields of science such as finance statistical physics and engineering.In finance , random walks are used to model stock prices and other financial assets.Random walks are also use in network theory to analyze the structure of networks such as the internet , social networks or biological networks, helping use understand the spread of information or diseases.

1.4 Mathematical Recipe

Direct Sampling For a dataset of randomly generated independent elements of S distributed according to $\pi(x)$ the sample mean M_n and rms of this data is

$$M_N = \frac{1}{N}\Sigma_{i=1}^{N} f(x_i)$$

$$S_n^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(f(x_i) - m_n)^2$$

$$rms = \frac{S_n}{\sqrt{n}}$$

Importance Sampling For a dataset of randomly generated independent elements of S distributed according to $\pi(x)$ with weights $w(x)$ the sample mean is

$$w(x_i) = \frac{\pi(x)}{g(x)}$$

$$M_N = \frac{1}{N}\Sigma_{i=1}^{N} w(x_i) \cdot f(x_i)$$

$$S_n^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(w(x_i) \cdot (x_i) - m_n)^2$$

$$rms = \frac{S_n}{\sqrt{n}}$$

2 Simple Random walks (SRW)

2.1 Definition and Basic Concepts

A simple random is a stochastic process in which a walker moves step by step across a N-dimensional lattice.Each step the walker takes is chosen randomly with equal probability and the steps are independent of each other.In a two

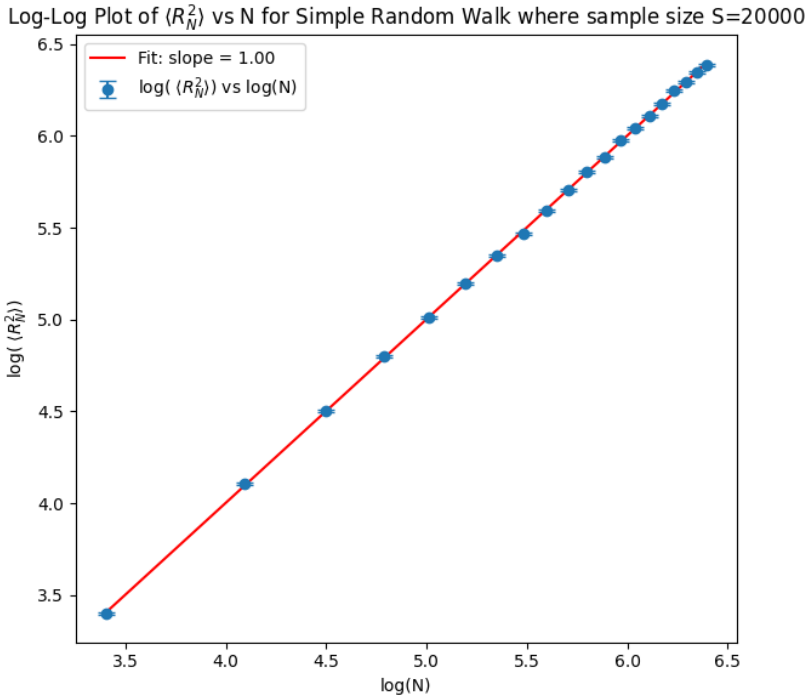
dimensional grid the walk has four possible directions to move .Each direction is equally likely with a probability of 1/4 .In this report we will investigate how the mean squared displacement R_N^2 scales with the number of steps.Simple random walks are characterized by their lack of memory and independent steps which lead to paths that are unbiased.Simple random walks can be used to model complex random processes such as diffusion or Brownian motion

2.2 Monte Carlo Simulation of SRW

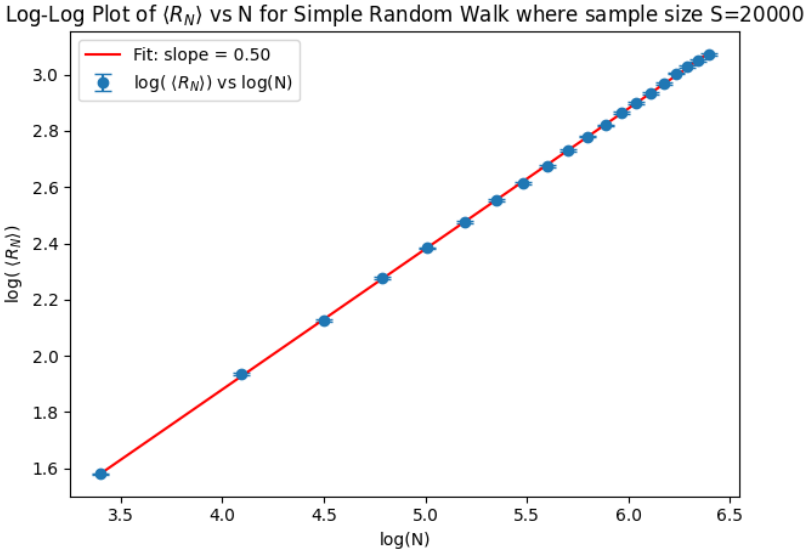
To implement the Monte Carlo method for simulating a simple random walk (SRW), I started by placing the walker at the origin, with coordinates x=0 and y=0. The walker randomly selected its next step from four possible directions—up, down, left, or right each with an equal probability of 1/4. This was done using a random number generator to determine the direction based on the range in which the random number fell. The walker took N steps, with N varying from 30 to 600 in increments of 30. For each N, I repeated the random walk 20,000 times to ensure accurate statistical results. After each trial, I calculated the squared displacement from the origin, and then averaged these squared displacements over all trials to obtain the mean squared displacement $\langle R_N^2 \rangle$ for each value of N. Finally, I plotted $\langle R_N^2 \rangle$ against N to analyze how it scales with the number of steps. of steps N.

2.3 Results and Analysis

2.3.1 Plots



(a) Plot showing Log-Log of N vs $\langle R_N^2 \rangle$.



(b) Plot showing Log-Log of N vs $\langle R_N \rangle$.

Figure 1: Comparison of Log-Log plots for N vs $\langle R_N^2 \rangle$ and N vs $\langle R_N \rangle$.

2.3.2 Analysis and Discussion

For a sample size of S= 20000 and a step count N ranging for 30 to 600 in increments of 30 we observe that the relationship between N vs $\langle R_N^2 \rangle$ is linear , with a slope of 1.00 .This result matches the theoretical prediction $\langle R_N^2 \rangle = N$ which is a fundamental property of simple random walks .This is because each step is independent and random , leading to a mean squared displacement that grows linearly with the number of steps .

Additionally, the log log plot of N vs $\langle R_N \rangle$ shows a slope of $1/2$, which is consistent with the theoretical prediction $\langle R_N \rangle = \sqrt{N}$.Both of results confirm that our simulation accurately captures the expected behavior of simple random walks.

3 Self-Avoiding Walks via Rejection Sampling (SAW-RS)

3.1 Introduction to Self-Avoiding Walks (SAW):

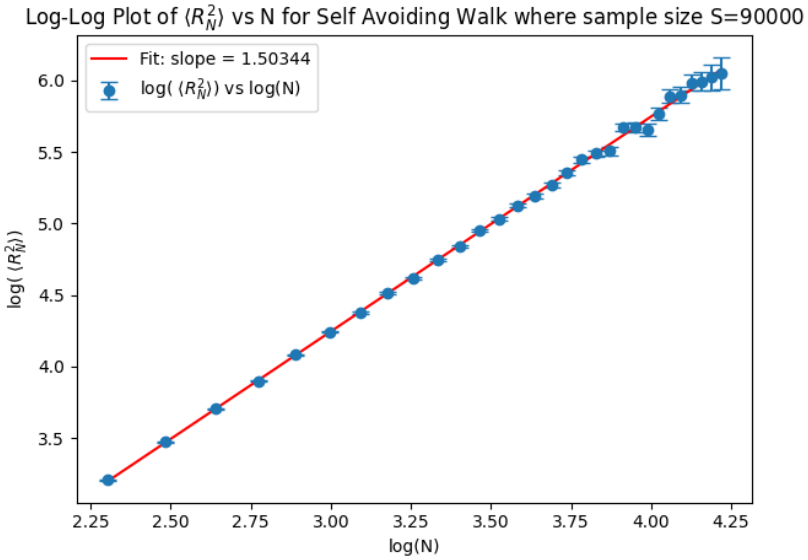
Self Avoiding walks are walks where the walker does not return to the same site it has already visited.Not much is known rigour-sly about SAW from a mathematical perspective , although multiple conjectures are believed to be true and are supported by numerical solution .In higher dimension , the SAW is believed to behave much like a ordinary random walk .These walks play a huge role in modelling topological and knot-theoretic behaviour of thread and loop like molecules such as proteins.SAWs were introduced to model real-life behavior of chain-like entities such as polymers whose physical volume prohibits multiple occupation of the same spatial point.Although properties of SAWs cannot be determined analytically, numerical solutions are used.[2]

3.2 Monte Carlo Simulation of SRW: Rejection Sampling

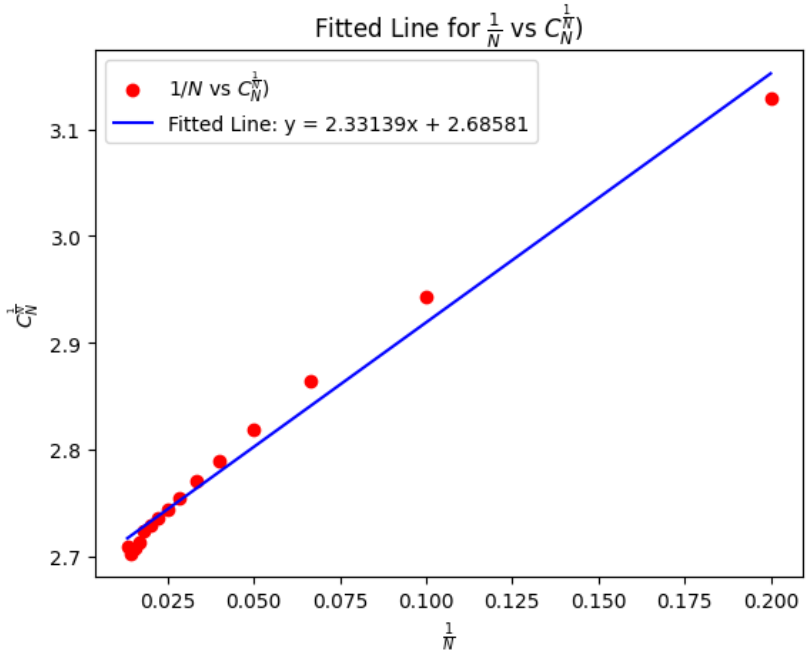
To generate self-avoiding walks (SAWs), we implement an algorithm that first generates non-reversing random walks (NRRWs) and then rejects any that self-intersect. While this method successfully produces SAWs, it is highly inefficient, especially for large N, due to the high rejection rate. As N increases, the probability of generating a valid SAW decreases significantly. For example, with N=20, only 1 in every 5 NRRWs is a valid SAW, and for N=39, the ratio drops to just 1 in every 48.

3.3 Results and Discussion

3.3.1 Plots



(a) Plot showing Log-Log of N vs $\langle R_N^2 \rangle$.



(b) Plot showing Log-Log of N vs $\langle R_N \rangle$.

Figure 2: Fitted Line for $C_N^{\frac{1}{N}}$ vs $\frac{1}{N}$.

3.3.2 Discussion

For a sample size of S = 90000 and step counts ranging from 10 to 70 in increments of 2 , the log-log plot of $\langle R_N^2 \rangle$ vs N yields a slope of 1.50344, closely matching the theoretical prediction $\langle R_N^2 \rangle = N^{1.5}$.this indicates that the mean squared displacement for self-avoiding walks grows faster than in simple random walks , where $\langle R_N^2 \rangle \propto N$ .The faster growth rate is because SAWs tend to be more extended and less compact to avoid self intersection the walker explores more space.Moreover the root mean square displacement increase more rapidly with N than in simple random walks.The percentage error between the theoretical value and the simulated result is 0.22933%, indicating that the simulation is highly accurate with the scaling law for SAWs.

However despite the accuracy , this method is very inefficient this is because of its high rejection rate as N increase the probability of the non-reversing walks self intersecting grows significantly as N grows.This results in the method being more biased towards short walks.For large values of N the RMS error becomes very large , this because many walks are rejected , the number of accepted valid walks decreases ,resulting in larger statistical fluctuation and greater uncertainty in our data. To obtain our Cn values we used this formula

$$\frac{C_N}{2d \cdot (2d-1)^{N-1}} = \frac{SAW}{TOTALWALKS}$$

The intercept of our for $C_N^{\frac{1}{N}}$ vs $\frac{1}{N}$ gives us our connective constant μ, which is the average number of sites available to the walker.The theoretical value of this constant for two dimensional walks approximately 2.63816 our value from the intercept is 2.68581 this value is very good estimation with a percentage error of 1.80612%

4 Self-Avoiding Walks via Myopic Sampling

4.1 Introduction to Myopic Sampling

Myopic sampling for a self avoiding random walks is a method where the walker chooses it steps based on the number of available steps it has at its current position.The goal of the walker is to avoided visited sites to ensure that the walk is not self intersecting .It is more efficient than rejection sampling because it avoids previously visited locations.However it is still not the most optimal way for simulating self avoiding walks , since it only considers immediate surroundings the walker can get stuck even thought there might have been an optimal sequence of steps that would have avoided such a situation.

4.2 Results and Discussion

4.2.1 Plot

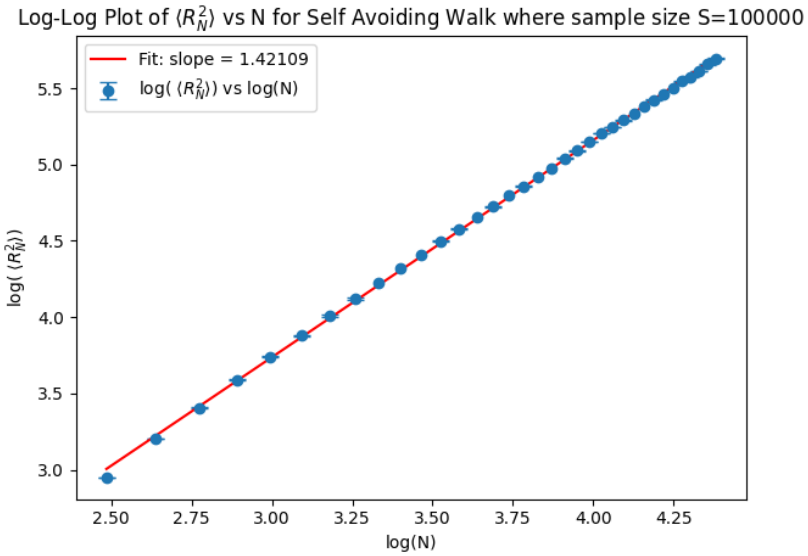


Figure 3: myopic sampling self avoiding walk

4.2.2 Discussion

For a sample size of S = 100000 and step counts ranging from 10 to 80 in increments of 2 , the log-log plot of $\langle R_N^2 \rangle$ vs N yields a slope of 1.42109, closely matching the theoretical prediction $\langle R_N^2 \rangle = N^{1.5}$ but with less accuracy than rejection sampling ..The percentage error between the theoretical value and the simulated result is 5.26067%.While rejection sampling produces a better estimate for scaling behaviour myopic sampling has a higher success rate in generating valid walks therefore produce more valid samples with less variance leading to smaller error bars.

4.3 Comparison to rejection sampling

While myopic sampling is more efficient than rejection sampling for generating self-avoiding random walks (SARWs), it is less accurate due to an inherent bias in the way it samples walks. Myopic sampling tends to favor more coiled or compact configurations, which skews the distribution of walks. This bias results in a smaller average squared displacement compared to rejection sampling, which samples SARWs uniformly across all configurations.

This difference in accuracy becomes evident when analyzing the data. In particular, the slope of the log-log plot of $\langle R_N^2 \rangle$ vs. N for myopic sampling de-

viates from the expected theoretical value, whereas the slope for rejection sampling aligns more closely with theory. This shows that while myopic sampling can quickly generate valid walks, its bias limits its effectiveness for accurately capturing the scaling behavior of SARWs, especially for longer walks.

5 Self-Normalizing Importance Sampling (Rosenbluth method)

5.1 Introduction to Importance Sampling

Importance sampling is a technique used in Monte Carlo method to improve the efficiency of estimating expectations of random variables, this method is used to sample from region that contribute significantly to the results but are are hard to sample from.Instead of sampling from probability distribution $f(x)$,importance sample use a more convenient sampling distribution $g(x)$,which allows us to sample data from regions of interest.This method does produce a bias but a correction factor, called the weight is introduce to adjust the bias .

5.2 Algorithm for Rosenbluth

The Rosenbluth method uses importance sampling, where the target distribution is $\pi(x) = \frac{1}{C_N}$, representing the uniform distribution over valid self-avoiding walks (SAWs). The sampling distribution $g(x)$ is proportional to the product of available directions at each step of the walk:

$$g(x) = C \cdot \frac{1}{k_1} \cdot \frac{1}{k_2} \cdot \frac{1}{k_3} \cdots \frac{1}{k_N}$$

where k_i represents the number of available directions at step i, and C is a normalization constant. The importance sampling weights for this method are given by:

$$w(x) = \frac{1}{C_N \cdot C} \cdot \prod_{i=1}^{N} k_i$$

These weights correct for the bias introduced by the sampling distribution, ensuring that walks with fewer available directions are properly accounted for in the final estimates. The Rosenbluth method thus increases the efficiency of generating SAWs by assigning appropriate weights, reducing the need for rejection sampling.

5.3 Results and Discussions

5.3.1 Plots

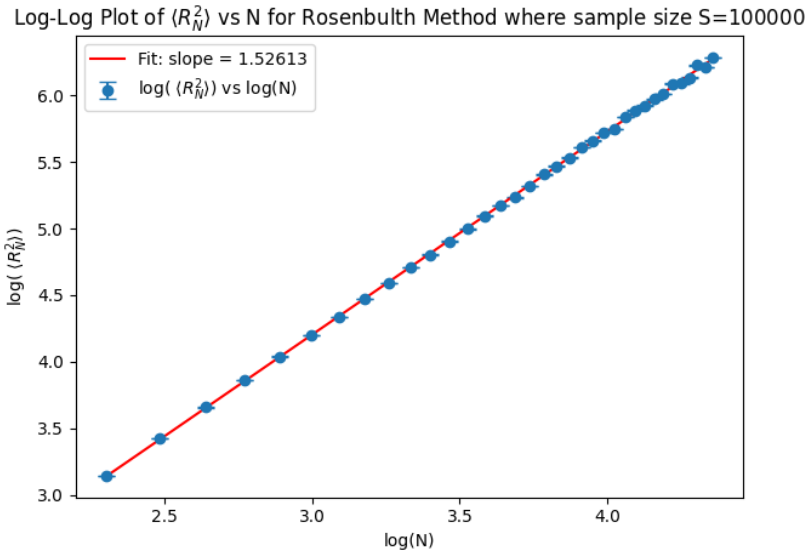


Figure 4: Importance sampling self avoiding walk

5.3.2 Discussion

For a sample size of S = 100000 and step counts ranging from 10 to 80 in increments of 2 , the log-log plot of $\langle R_N^2 \rangle$ vs N yields a slope of 1.52613, closely matching the theoretical prediction $\langle R_N^2 \rangle = N^{1.5}$ but with less accuracy than rejection sampling ..The percentage error between the theoretical value and the simulated result is 1.742%.

5.3.3 Comparison to the rejection method and myopic sampling

The Rosenbluth method is better at producing self-avoiding walks than rejection sampling and myopic sampling .Rejection method generates many self intersecting walks and discards them leading to inefficiency especially when generating large walks.Myopic sampling produces walks that are more biased toward being coiled up , this leads to an inaccurate distribution.The Rosenbluth methods corrects for this bias produced by myopic sampling and it is more efficient than rejection sampling making a much better algorithm for generating self-avoiding walks

6 Practical Applications for Random Walks

Simple Random Walks

- **Diffusion** Simple Random Walks are used to study diffusion in physics , where particles move randomly in a medium
- **Stock Price** Model stock price in financial markets this assumes that the prices changes are completely random and unpredictable over time[1].

Self-Avoiding Walks via Rejection Sampling

- **Polymer Physics** SAWs are used to model structures of long chain molecules .These polymers don't cross each other, SAWs allow use to study their configuration such as size and flexibility
- **Protein Folding** Self-avoiding walks are used to model how proteins fold in biological systems , this is because the process is restricted by the spatial constraints of the protein chain[1].

Self-Normalizing Importance Sampling

- **Monte Carlo Integration in Complex Systems** The self-normalizing importance sampling method is applied in Monte Carlo integration , in high-dimensional spaces where sampling is challenging.This method increase accuracy because it focuses on the important region.
- **Efficient Polymer Modeling** The Rosenbluth method is also used to model structures of long chain molecules but it generates SAWs efficiently.This reduces computational cost of rejection sampling[1].

Self-Avoiding via Myopic Sampling

- **Search Algorithms** Myopic sampling is used to traverse a graph without revisiting nodes

7 Conclusion

Although the rejection method provides a good estimate for the scaling properties of SAWs it is very inefficient die to it's high rejection rate ,especially for large N's .Myopic sampling is more efficiency in generating SAWs but introduces biases that affect the accuracy of the scaling properties.This method provides more walks than rejection method but the walks produces tend to be more coiled. The Rosenberg method is the most effective approach , because it balances both accuracy and efficiency.It generates walks better than the rejection and produce unbiased walks , overall it the the most reliable method.

References

[1] OpenAI. (n.d.). ChatGPT. https://openai.com/models/gpt/

[2] https://en.wikipedia.org/wiki/Self-avoiding_walk

[3] Wikipedia - https://en.wikipedia.org/wiki/MonteCarlomethod