CS215 Assignment-1
Problem 3

Josyula Venkata Aditya[1] and Kartik Sreekumar Nair[2]

[1]210050075
[2]210050083

October 1, 2022

Problem 3

Simulate $N := 10^4$ independent random walkers (as discussed in class) along the real line, each walker starting at the origin, and each walker taking 10^3 steps each of length 10^{-3} .

Approach

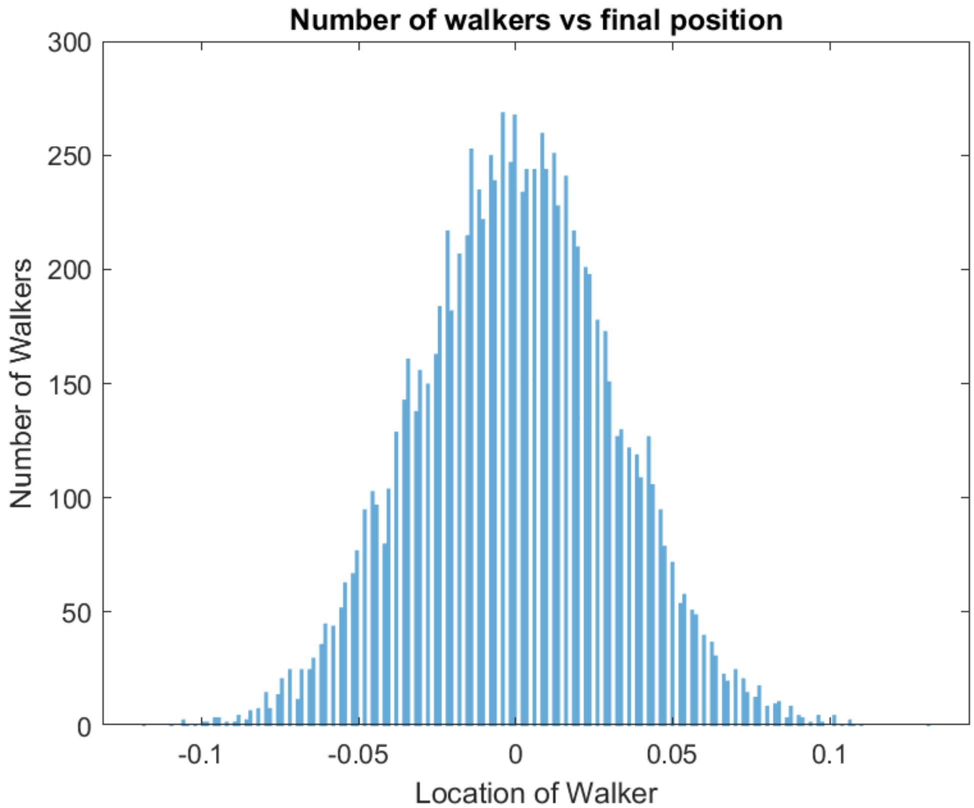
Each random walker has a 0.5 chance of taking a step in the +ve direction and a 0.5 chance of a step in the negative direction. To model this we used the `randi()` function to generate steps of $\{+10^{-3} \text{ or } -10^{-3}\}$ with equal probability.

We generated all the steps for the 10^4 walkers and 10^3 steps of each of them together as a matrix.

To get the final positions of each walker we just had to sum the matrix along the second dimension.

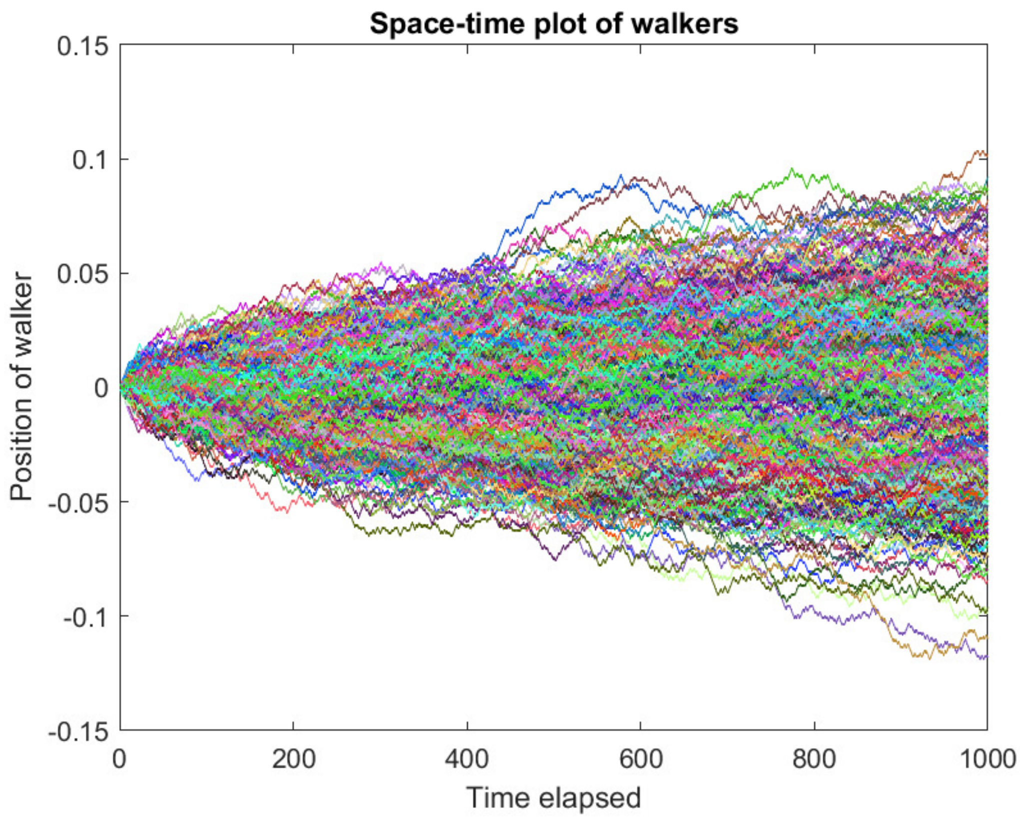
Similarly, to get the position of the walker at each time step we used the cumulative sum (`cumsum()`) function along the second dimension (here we only took the first 10^3 random walkers as required by the question).

- Plot a histogram of the final locations of all the random walkers. We used the `histogram(·)` function to plot the histogram from the array of the final positions of the random walkers.



Central limit theorem - The random walkers have the same distribution for their final locations and the paths taken by one walker is independent of the other walkers; it becomes easier to believe from this graph that the distribution as $N \to \infty$, will actually be Gaussian.

- For the first 10^3 walkers, plot space-time curves that show the path taken by each walker (as depicted in the class slides). On the graph, draw each path in a different randomly-chosen color for better clarity.



- Consider a random variable X. Consider a dataset that comprises N independent draws (e.g., modeled by $X_1, \cdots, X_N$) from the distribution of X.

  Use the law of large numbers to show that the random variable

$$\hat{M} := \frac{(X_1 + \cdots + X_N)}{N} \tag{1}$$

  converges to the true mean

$$M := E[X] \tag{2}$$

  as $N \to \infty$. Prove that the expected value of the random variable

$$\hat{V} := \sum_{i=1}^{N} \frac{(X_i - \hat{M})^2}{N} \tag{3}$$

  tends to the true variance

$$V := Var(X) \tag{4}$$

  as $N \to \infty$.

  The result for $\hat{M}$ follows directly from the law of large numbers. Proof of the law is the following: Let

$$X_l = \frac{X_1 + X_2 \cdots X_N}{N} \tag{5}$$

is a random variable denoting the average of X_is where X_is follow the same distribution as X. We know that $E[X+l] = E[X]$ from the linearity property of expectation. Also, since the trials are independent,

$$Var(X_l) = \frac{Var(X_1) + Var(X_2) \cdots Var(X_N)}{N^2} = \frac{\cdot Var(X)}{N} \tag{6}$$

From Chebyshev's inequality, we know that

$$P(|X_l - E[X]| \geq \epsilon) \leq \frac{Var(X_l)}{\epsilon^2} = \frac{Var(X)}{N\epsilon^2} \tag{7}$$

Assuming that $Var(X)$ is finite, we can always find a sufficiently large N for a given ϵ such that

$$P(|X_l - E[X]| < \epsilon) \geq 1 - \frac{Var(X)}{N\epsilon^2} \tag{8}$$

In other words, $X_l \to E[X]$ as $N \to \infty$ (choose arbitrarily small epsilon). Hence $\hat{M}$ converges to M. Now, we proceed to show it for $\hat{V}$.

$$\begin{aligned}
\hat{V} &= \sum_{i=1}^{N} \frac{(X_i - \hat{M})^2}{N} && (9)\\
&= \sum_{i=1}^{N} \frac{1}{N}\left(X_i^2 - 2X_i\hat{M} + \hat{M}^2\right) && (10)\\
&= \sum_{i=1}^{N} \frac{X_i^2}{N} - 2\hat{M}\sum_{i=1}^{N}\frac{X_i}{N} + \hat{M}^2 && (11)\\
&= \sum_{i=1}^{N} \frac{X_i^2}{N} - \hat{M}^2 && (12)\\
&\longrightarrow E\left[X^2\right] - M^2 && (13)\\
&= Var(X) && (14)
\end{aligned}$$

Where we used the law of large numbers on the first term $\sum_{i=1}^{N} \frac{X_i^2}{N}$, as X_i^2 is also a random variable, to show that it converges to $E[X^2]$ in 13. And $\hat{M} \to M \implies \hat{M}^2 \to M^2$

- Report the empirically-computed mean $\hat{M}$ and the empirically-computed variance $\hat{V}$ of the final locations of the random walkers.

  Empirical mean: $1.386 \cdot 10^{-4}$
  Empirical variance: $1 \cdot 10^{-3}$

- What should the values of the true mean and the true variance be for the random variable that models the final location of the random walker, as function of the step length and the number of steps ?

  True mean is 0 since there is an equal probability to step either left or right.
  Variance $= 2Dt = \frac{t(\Delta z)^2}{\Delta t} = n \cdot (\Delta z)^2 = 10^3 \cdot (10^{-3})^2 = 10^{-3}$

  Alternatively, variance in number of steps$(V_s) = npq = 0.25 \cdot 10^3 \implies$
  variance in the locations $= V_s \cdot (2\Delta z)^2 = 10^{-3}$

- Report the error between the empirically-computed mean and the true mean. Report the error between the empirically-computed variance and the true variance.

  Error in mean(absolute): $1.386 \cdot 10^{-4}$
  Error in variance(absolute): $4.8 \cdot 10^{-6}$
  Error in variance(relative): 0.0048

  Refer `problem3.mlx` for code and data.