PHSX815: Computational Methods in Physical Sciences
Project 1
Project Paper
Title: Random Walk in 1D

Vishakha Lingadahally

May 16, 2021

Introduction

In Science, we are often interested in a set of outcomes rather than a single event. Stochastic simulations allow us to generate and examine a series of simulations of a system in which the steps are governed by random choice. A simple algorithm for flipping of coin can be a useful tool to 'choose' if a step in a phenomenon will occur or not. Consider a drunkard trying to walk along the pavement in the middle of the night. In their drunk state of mind, they can only move sideways-either to the left or the right from their current position. Whether they take a step to the left or the right can be dictated by flipping a coin and checking if the outcome is 'Heads' or 'Tails'. The philosophy behind the seemingly simple thought experiment has a wide range of application in physics, chemistry as well as biology; to describe the diffusion of gas molecules, simulate bacterial motion and learn the pattern of genetic drift.

In this project, we simulate the random walk experiment in 1D and compute the likelihood for our hypothesis of the drunkard stepping to either left or right with equal probabilities. The likelihood of the occurrence is computed using the likelihood for a binomial probability distribution. The likelihood of a binomial distribution is given by the following expression:

$$L(p|n) = \frac{n!}{k!(n-k)!} p^{(k)} q^{(n-k)}$$

where in our context, n is the total number of steps taken, k is the number to steps taken to the right, p is the probability of stepping to the right and q is that stepping to the left.

Understanding the code

Flipping the coin

The first section of the code consists of importing the necessary modules for computation and plotting. We then begin by flipping the coins and understanding their output. To see why this is relevant in random walk, think of a gas molecule that has equal probabilities of moving either to the left of the right. To dictate whether the molecule moves to the left or the right by flipping a coin. This will obviously change every time that we run the code cell. To convert this to a 'heads' and 'tails' readout, we can assume that this is a totally fair coin. This means that the probability of getting 'heads' to get P(H) is the same as flipping a 'tails' P(T) such that P(H)+P(T)=1 . This means that for a fair coin, P(H)=P(T)=0.5 . To convert our coin flips to 'heads' or 'tails', we simply have to test if the flip is above or below 0.5 . If it is below 0.5, we say that the coin was flipped 'heads', otherwise, it is 'tails'.

Testing if the coin-toss is fair

Now, in order to check if our coin-toss is fair, we toss it multiple times, say 1000 times. Let's flip a coin one thousand times and compute the probability of getting 'heads'. The simulated probability is very close to our expected P(H)=0.5, but not exactly. This is the nature of stochastic simulations-it's based on repetitions. If we were to continue to flip a coin more number of times, our simulated P(H) would get closer and closer to 0.5 . This is why doing many repetitions of stochastic simulations is necessary to generate reliable statistics.

The random walk

We start at position zero and flip a coin at each time step. If it is less than 0.5, we take a step left. Otherwise, we take a step to the right. At each time point, we keep track of our position and then plot our trajectory.
In order to compute the the number of steps taken to the right and that taken to the left, we solve the simultaneous equations n_l + n_r = 100 and n_l - n_r = final position, where n_l is the number of steps taken to the left and n_r is the number of steps taken to the right. We plug these values in the likelihood for our binomial distribution, which can be stated as "The likelihood of the probability of stepping to the right, given that the drunkard takes n_l steps to the left and 100 - n_l steps to the right.
As our steps are based on the generation of random numbers, this trajectory will change every time the code is run. The power of stochastic simulation comes from doing them many times over. The next step would be to run the steps, say, one thousand times and plot all of the traces. We plot the mean at each time point as a blue line.

Understanding the output

The user sees the following results as the output of the code:

- The results of the of three consecutive coin flips that our code simulates in the beginning is displayed. Then, these probability outputs are converted to 'heads' and 'tails'.
- Then, the predicted and simulated probabilities are then displayed. The predicted probability, naturally, is 0.5 and the simulated probability is the ratio of the number of heads to that of total number of tosses.
- The number of steps taken to the left and the number of steps taken to the right are computed for our hypothesis.
- The likelihood for the binomial distribution with equal probabilities of stepping to either to the left or the right is computed.

- The plots of the tracks of the random walk for both the hypotheses are generated, along with the random walk simulated multiple times (n=1000).

Output interpretation

Let us take a look at the importance of repeating the experiments that are stochastic in nature. The following plots are those of 10, 100 and 1000 trials of random walk respectively (each with 100 steps).

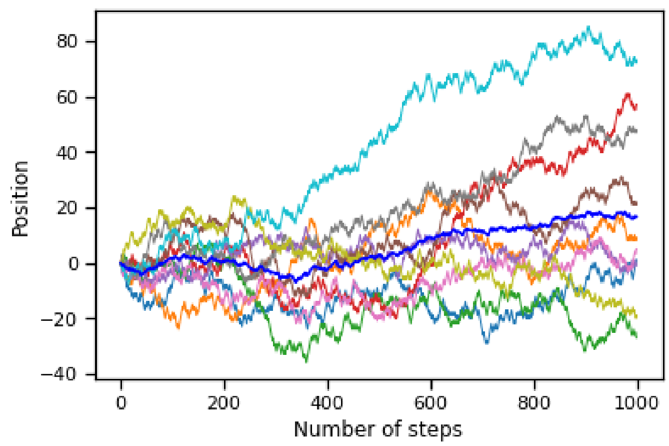

Figure 1: Set of 10 tracks for 100 steps

At n=10, we see that the mean computed at each step (the blue track) isn't exactly a straight line positioned at y=0, as we would expect for a typical random walk.
For n=100:

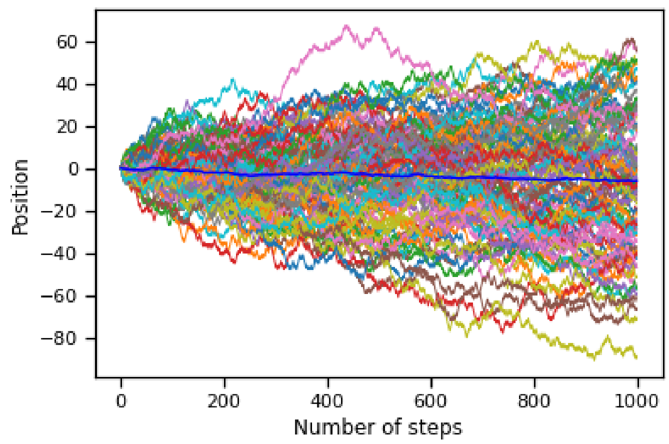

Figure 2: Set of 100 tracks for 100 steps

At n=100, we see that the data points or tracks get thicker as we move towards Position=0, from either top or bottom. The blue 'curve' which represents the mean computed at each point resembles a straight line better, at y=0.

For n=1000:

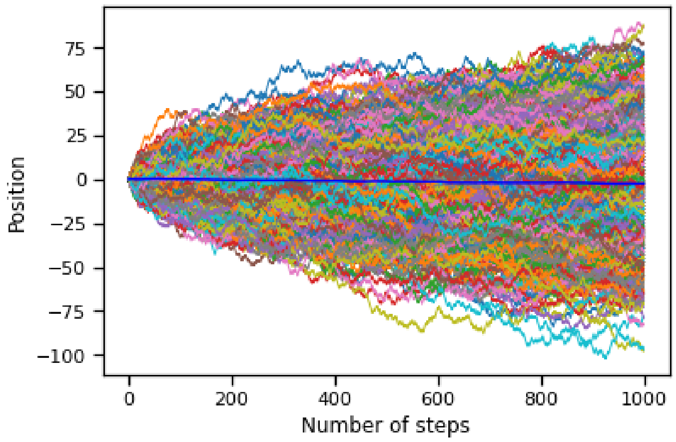

Figure 3:

At n=1000, the mean is essentially almost a straight line at y=0.
The comparison of the position and the straightness of the blue curve representing the mean at each point makes it clear how important repetition of trials are.
Taking a look at the histogram of the probability distribution of the directions of the steps, at a given step provides an insight into the stochastic nature of the random walk. Here's a histogram for the distribution at Step=0.

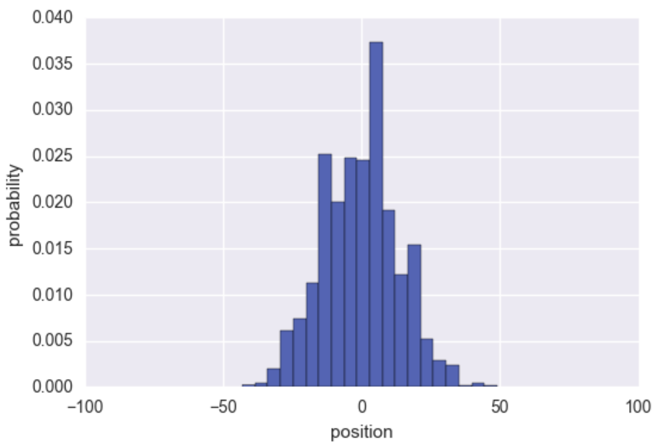

Figure 4: Caption

It resembles a Gaussian distribution. One can also see the Central Limit Theorem in action by varying the number of simulations and observing how the distribution resembles Gaussian distribution with increasing number of simulations.

Let us take a look at random walk at slightly unequal probabilities of stepping to the left or to the right. This can be achieved by changing the parameter 'step_prob' in the code. Say, if step_prob=0.4, i.e., the probability of stepping to the left was lower than stepping to the right, this is what our set of tracks would look like:

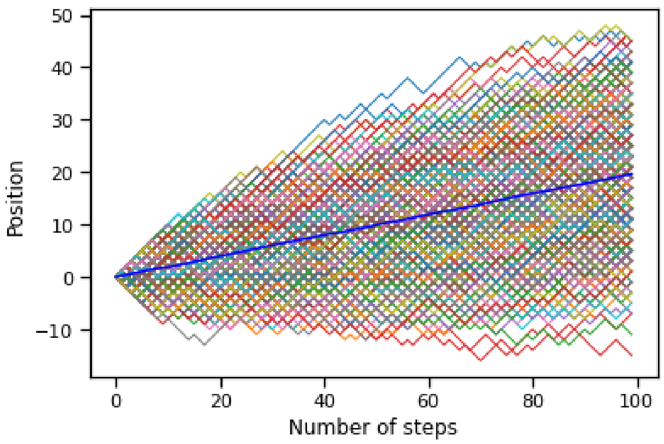

Figure 5: Set of tracks for step_prob=0.4

Alternatively, for step_prob=0.6, this would be our set of tracks:

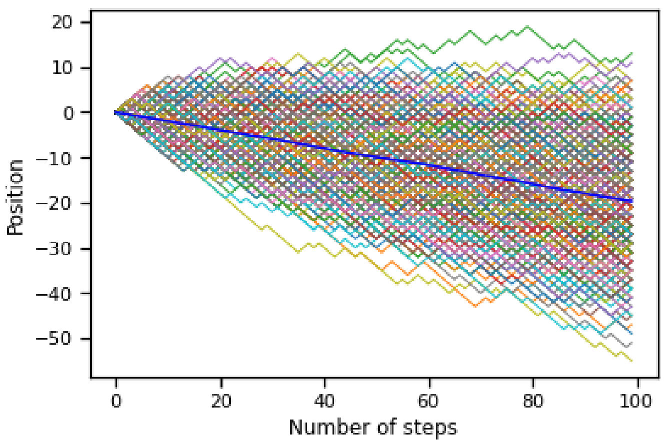

Figure 6: Set of tracks for step_prob=0.6

With varying probabilities of stepping to either of the sides, one can explore how the likelihood varies for repeated experiments, which I plan to do in Project 2.

Summary

- The random walk for a hypothesis of equal probabilities of stepping on either side has been simulated in 1D. The behavior of the distribution with increased number of trials indicate the importance of repeated experiments in stochastic scenarios.
- By taking a look at alternate hypotheses, one can compare the likelihoods for different values of the parameter of interest to determine the most favored outcome in the stochastic process.

References:

- github.io
- stackoverflow.com
- medium.com
- towardsdatascience.com