

MATH 310

Probability & Statistics

Project Observations

MUHAMMAD MUNAWWAR ANWAR (MA04289)
KHUBAIB NAEEM KASBATI (KK04333)

GITHUB REPOSITORY : Probability Random Walks

June 19, 2020

Contents

Figures	3
Tasks	6
Task 1	6
Task 2	6
Task 3	6
Task 4	6
Task 5 & 6	6
Task 7	7
Task 8	7
Bibliography	15

List of Figures

1	Task 1: One Dimensional Discrete Random Walk	8
2	Tak 2: Distrubution of time required for two people who are intially 100 units away	9
3	Reflection from the Circumference	9
4	Task 3: Two Dimensional Discrete Random Walk, with main graph showing distribution of the points after simulation, subgraph showing distribution of x (Below x axis) and a subgraph showing distribution of y (Left of y axis)	10
5	Task 4: One Dimensional Discrete Random Walk , with main graph showing distribution of the points after simulation, subgraph showing distribution of x (Below x axis) and a subgraph showing distribution of y (Left of y axis)	11
6	Task 5 & 6: Two Dimensional Continuous Random Walk, with main graph showing distribution of the points after simulation, subgraph showing distribution of x (Below x axis) and a subgraph showing distribution of y (Left of y axis)	12
7	Task 7: Two Dimensional Random Walk with Discrete Step Size and Continous Angle, with main graph showing distribution of the points after simulation, subgraph showing distribution of x (Below x axis) and a subgraph showing distribution of y (Left of y axis)	13
8	Expected number of steps it would take for two nodes to meet.	14

Tasks

Task 1

In task 1, we ran the simulation 100,000 times, where each simulation consisted of 50 steps and the probability for the walk to move left, right or stay equally likely and the starting position was the point zero, the expected value of the distance travelled was 0.00992. The resulting distribution can be found in Figure 1. Furthermore, the distribution was symmetric about the mean value. However, if the chosen starting point is not zero or the probability to move left, right or stay is unequal then the mean of the distribution shifts or it becomes left or right skewed.

We know that this simulation is valid since a gaussian distribution is formed due to the Law of Large numbers and the expected movement at each turn is 0. $E(X) = \frac{1}{3} \times -1 + \frac{1}{3} \times 0 + \frac{1}{3} \times 1 = 0$ and hence $E(50X) = 50E(X) = 0$, Due to linearity of expectation and because each event is independent. This matches the result we achieved empirically.

Task 2

In task 2 we ran the simulation 10000 times, where in each simulation the starting distance x between two people was 100 and the probability for the both person to move left, right or stay was equally likely and same, the expected value of time required for both people to meet was 8468.63. The distribution can be seen in Figure 2. The distribution is left skewed. (If we change the probabilities it will change skewness and changing the initial distance shifts the center of the graph) The starting distance x between the two people appears to have a linear relation with number of time required to meet.

Our model is based on the assumption that there exists some outliers. And excluding them would not effect the expected value by much. We have added a timeout and if the simulation fails in that time then we ignore that simulation, this can lead to a correct average when the timeout is set accordingly but for large distances it can distort the real expected value by causing the simulations with near expected time to timeout.

Task 3

Suppose that a walker is at the position $P = (P_1, P_2)$ inside the test region at $(k - 1)^{th}$ step, but the next step would be to a position $Q = (Q_1, Q_2)$ that is either on the boundary or beyond the boundary. We divide the contemplated step from P to Q into two steps : (1) an actual step (the k^{th}) to the boundary point P' , where the segment \vec{PQ} intersects the boundary, and then (2) a step $(k + 1)^{st}$ of one unit to a point Q' inside the circle. Specifically, let P' be the point where the walker hits or would cross the boundary and α the angle of incidence of the walker's path with the tangent line to the circle at P' . The walker takes a step (k th step) to P' and then takes a step $((k+1)st)$ back into the circle on a path which forms an angle α with the tangent line at P' . This Reflection method can be seen in Figure 4. The distribution of these points can be seen in Figure 5

Task 4

Similar to task 1 when we ran the simulation 10000 times, where each simulation consisted of 50 steps and the step was modeled as a uniform random variable between 0 and 1, the expected distance in a particular step is 0.5 hence in 50 steps is 25. This matches with our empirical result of ≈ 25.03 Furthermore, the distribution was symmetric about the mean value. However, if the chosen starting point is not zero or the probability to move left, right or stay is unequal then the mean of the distribution shifts or it becomes left or right skewed. Shifting the start of the simulation by x units leads to the distribution being shifted by x units. The distribution can be seen in Figure 5.

Task 5 & 6

We have created a random walk model and ran a simulation 100,000 times with 10,000 steps each. With step size being a uniform random variable between 0 and 1, and angle as a uniform random variable between 0 and 2π . The results were that the distribution was very well spread around the center, with the x and y values having

a gaussian distrubution. The most common points were near the center of the circle. This checks out since the expected movement in x,y direction is 0 per step. Hence throughout the origin, on average it should be near the origin. The model has been carried forward from Task 3.

$$\int_0^{2\pi} \int_0^1 r \cos(\theta) = 0$$

$$\int_0^{2\pi} \int_0^1 r \sin(\theta) = 0$$

Task 7

We have created a random walk model and ran a simulation 100,000 times with 10,000 steps each. With step size being a discrete random variable which takes the values $\{0, 0.5, 1\}$, and angle as a uniform random variable between 0 and 2π . The results in task 7 were analagous to those in task 5 and 6. The distrubution was very well spread around the center, with the x and y values having a gaussian distrubution. Most of the points were near the center of the origin. The model has been carried forward from Task 3.

Task 8

In task 8, when we ran the simulations 14000 times the expect value of time that it takes for two nodes to meet was 111,504. The initial location of the model was chosen by using polar coordinates and modelling radius as uniform random variable between 0 and 100 and the angle as a uniform random variable between 0 and 2π . In order to determine the appropriate number of simulations required we will use 95% confidence interval and precision level at 2000.

Level of Precision : The maximum degree the true population mean can deviate from the sample mean estimation,subject to a given confidence interval.

The confidence interval is calculated as follows :

$$CI = (\bar{x} - z \times \frac{s}{\sqrt{n}}, \bar{x} + z \times \frac{s}{\sqrt{n}})$$

where \bar{x} is the sample mean, s is the standard deviation of the sample mean , n is the sample and z is the statistic associated with a certain confidence interval. Since we have chosen a 95% confidence interval, the z statistic equals to 1.96 approximately.

After some algebaric maniuplations , we can write the above equation as follows :

$$P(\frac{-\phi}{s\sqrt{n}} < \frac{\bar{x} - \mu}{s\sqrt{n}} < \frac{\phi}{s\sqrt{n}})$$

where \bar{x} is the sample mean, s is the standard deviation of the sample mean , n is the sample size, μ population mean,and ϕ is the level of precision.

According to the Central Limit Theorem, the middle term in the above inequality is normally distrubuted with mean zero and standard deviation one. Consequently, we can solve the lower and upper bounds of the equation for the only unknown variable , n .

We know that s , the sample standard deviation , 117,551.58 , $z = 1.96$ and $\phi = 2000$

$$\sqrt{n} = \frac{1.96 \times 117,551.58}{2000} , \text{ so } n = 13,271 \text{ simulation}$$

Since our sample distribution was not a normal distribution, the distribution of the sample variance was unknown. This means that our answer we got for n , number of simulations is not the theoretical minimum. Although, a larger sample size means that the distribution of sample variance approaches to the normal distribution.

Figures

Figure 1: Task 1: One Dimensional Discrete Random Walk

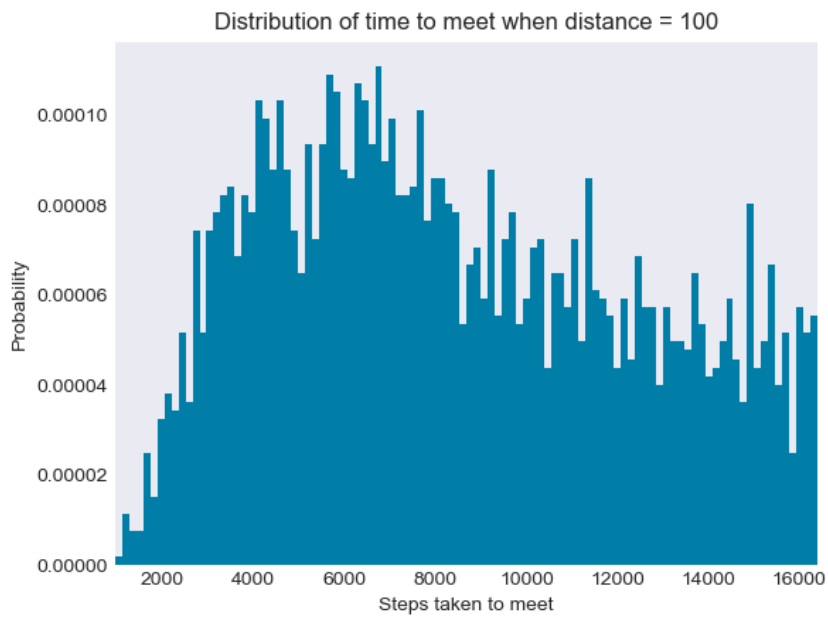


Figure 2: Tak 2: Distrubution of time required for two people who are initially 100 units away

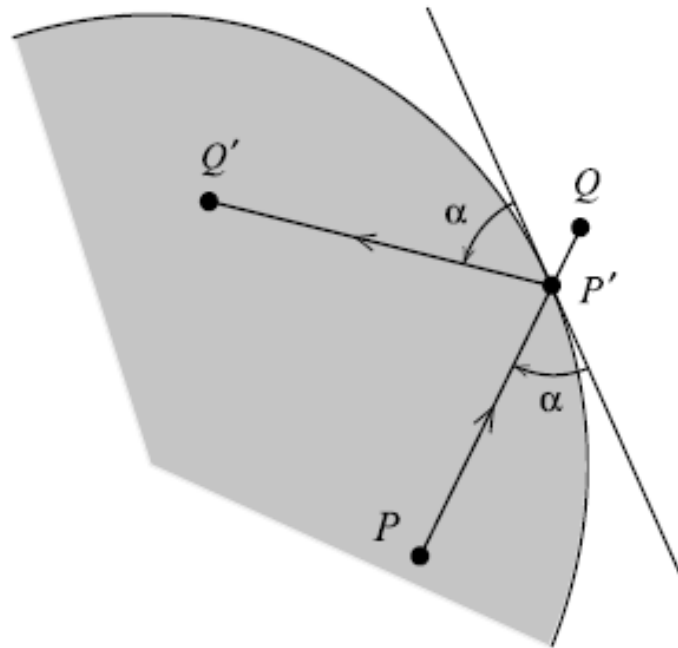


Figure 3: Reflection from the Circumference

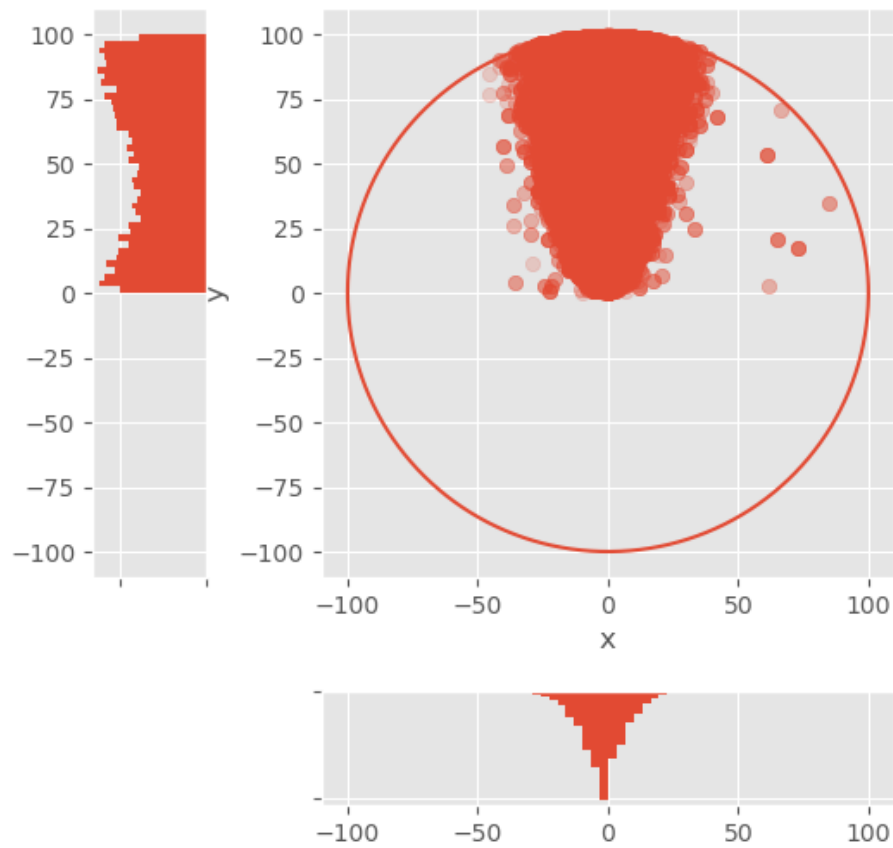


Figure 4: Task 3: Two Dimensional Discrete Random Walk, with main graph showing distribution of the points after simulation, subgraph showing distribution of x (Below x axis) and a subgraph showing distribution of y (Left of y axis)

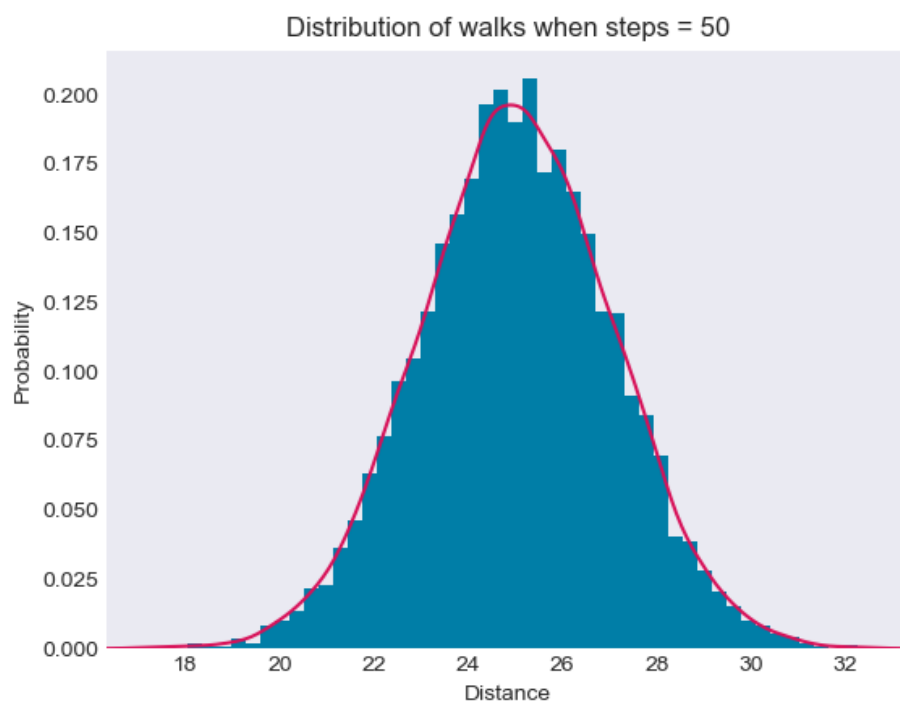


Figure 5: Task 4: One Dimensional Discrete Random Walk , with main graph showing distribution of the points after simulation, subgraph showing distribution of x (Below x axis) and a subgraph showing distribution of y (Left of y axis)

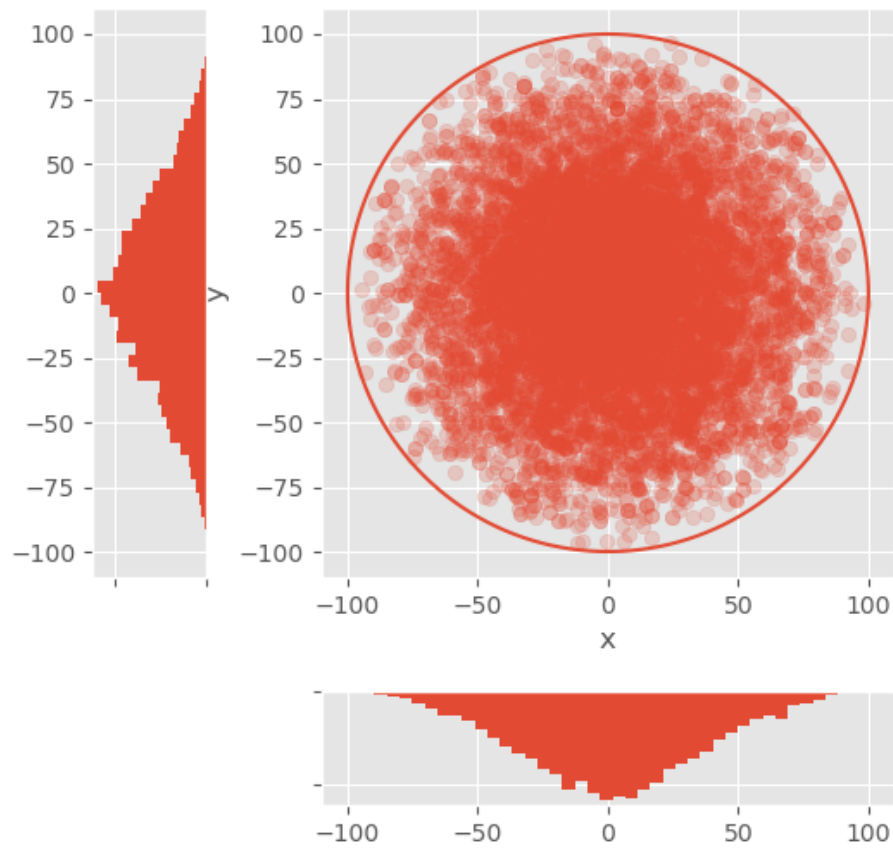


Figure 6: Task 5 & 6: Two Dimensional Continuous Random Walk, with main graph showing distribution of the points after simulation, subgraph showing distribution of x (Below x axis) and a subgraph showing distribution of y (Left of y axis)

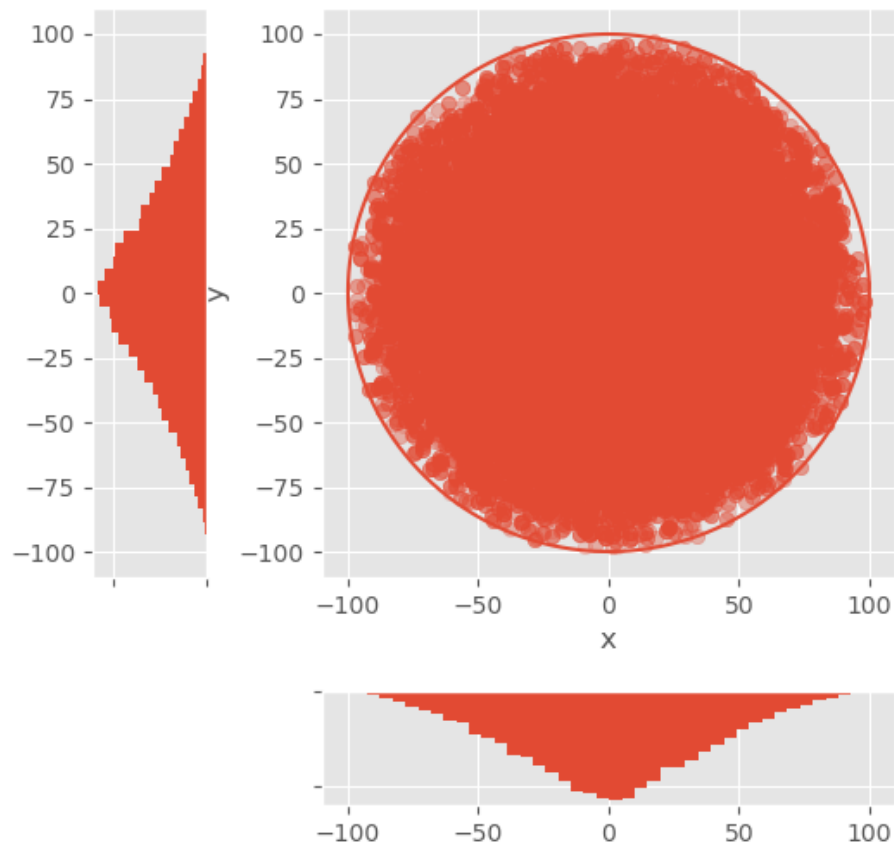


Figure 7: Task 7: Two Dimensional Random Walk with Discrete Step Size and Continuous Angle, with main graph showing distribution of the points after simulation, subgraph showing distribution of x (Below x axis) and a subgraph showing distribution of y (Left of y axis)

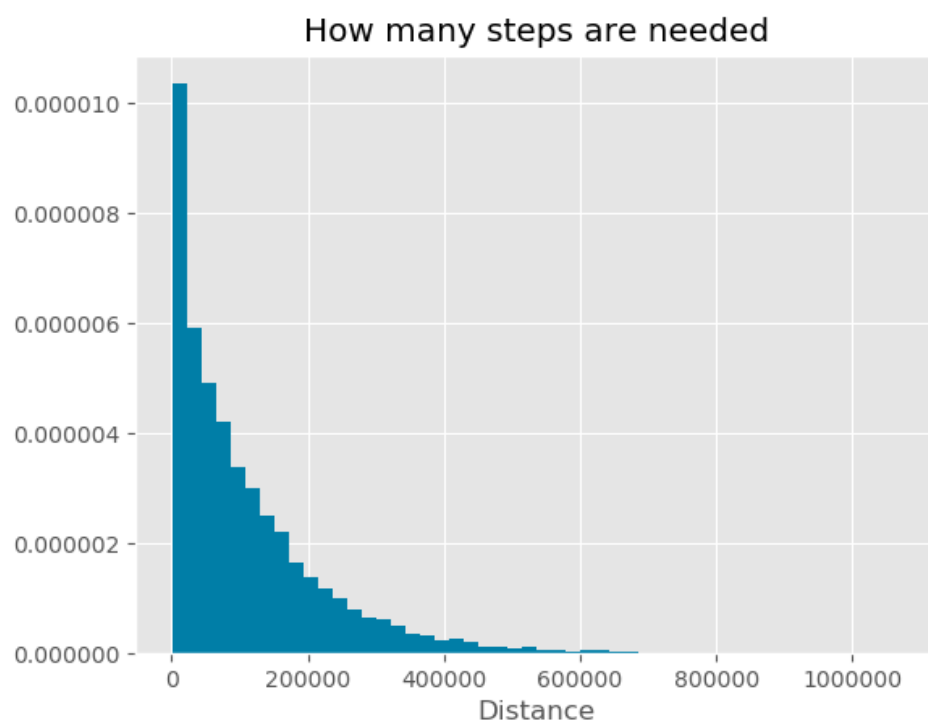


Figure 8: Expected number of steps it would take for two nodes to meet.

Bibliography

- [1] http://www.cplusplus.com/reference/random/discrete_distribution/discrete_distribution/
- [2] <https://stackoverflow.com/questions/43839974/create-c-discrete-distribution-from-list>
- [3] http://www.cplusplus.com/reference/random/uniform_real_distribution/
- [4] <https://docs.scipy.org/doc/numpy-1.15.0/reference/generated/numpy.random.choice.html>
- [5] <https://jakevdp.github.io/PythonDataScienceHandbook/04.08-multiple-subplots.html>
- [6] <https://numpy.org/doc/stable/reference/random/generated/numpy.random.uniform.html>
- [7] Sanaz Sadegh (2020). 2D random walk confined in a circular domain with reflective boundaries (<https://www.mathworks.com/matlabcentral/fileexchange/60673-2d-random-walk-confined-in-a-circular-domain-with-reflective-boundaries>),MATLAB Central File Exchange. Retrieved June 19, 2020.
- [8] <https://math.stackexchange.com/questions/1013230/how-to-find-coordinates-of-reflected-point>
- [9] Mathematical Computing: An Introduction to Programming Using Maple By David Betounes, Mylan Redfern
- [10] OPTIMAL NUMBER OF TRIALS FOR MONTE CARLO SIMULATION BY MARCO LIU, CQF (https://www.valuationresearch.com/wp-content/uploads/kb/SpecialReport_MonteCarloSimulationTrials.pdf)
- [11] <https://stackoverflow.com/questions/33203645/how-to-plot-a-histogram-using-matplotlib-in-python-with-a-list-of-data>