Report of Random Walk

H24116049 莊秉宸

● One-Dimension

In one-dimensional random walking, I walk through 1000 walks of n steps, where n takes the values $\{10^2, 10^3, 10^4, 10^5, 10^6\}$. Also, I record the sections in which the particle is located during a single walk. Followings are the results.

- The particle stays at section 1 for **559556396** time steps.
- The particle stays at section 2 for **550397928** time steps.

In this simulation, we can observe that the time spent by the particle in section-1 is **approximately equal to that** in section-2.

Then, we plot the distribution of the L1-norm and L2-norm of the final position of the particle.

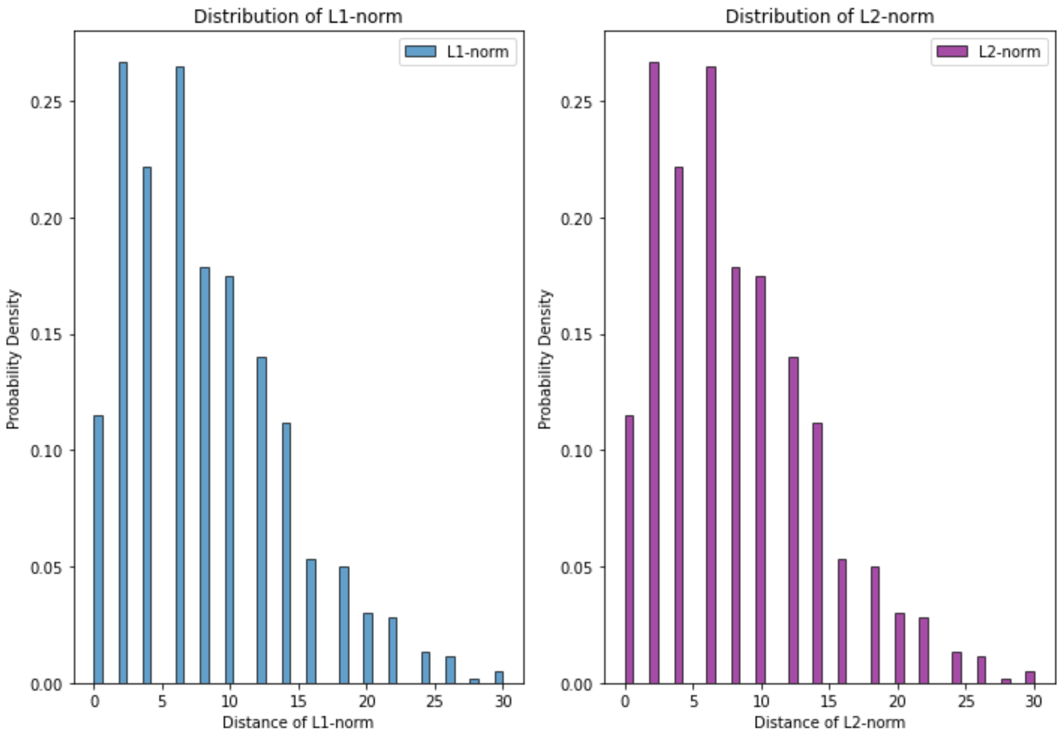

Figure 1: The L1 and L2-norm distribution of final positions after **100** steps

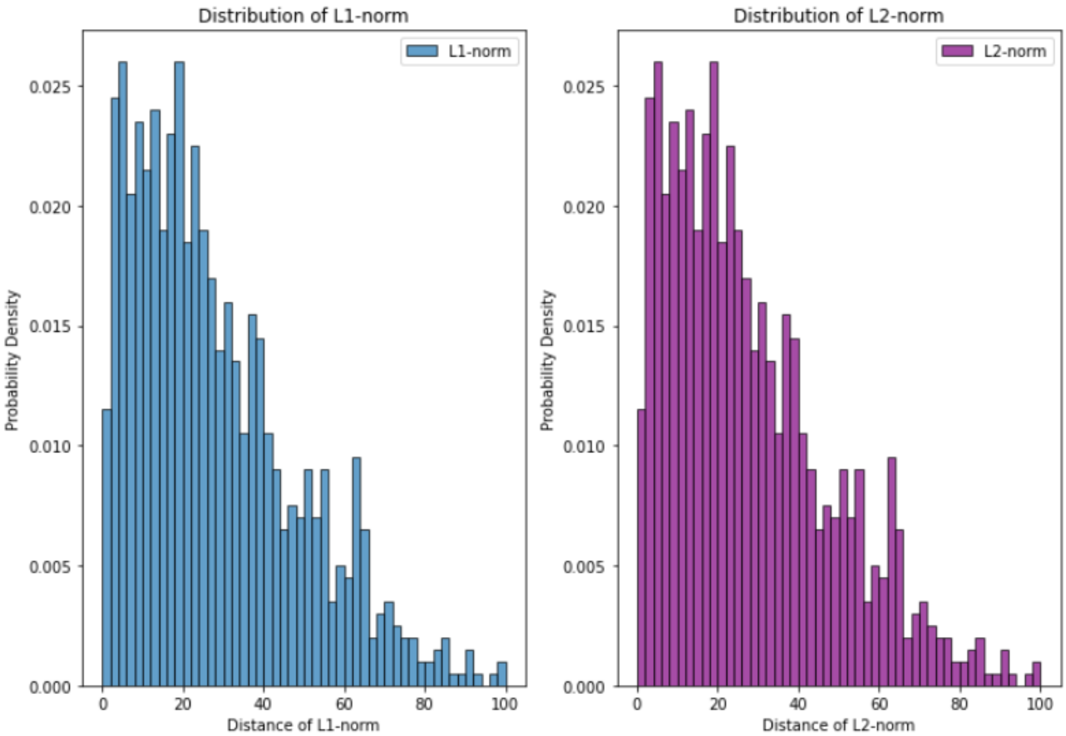

Figure 2: The L1 and L2-norm distribution of final positions after **1000** steps.

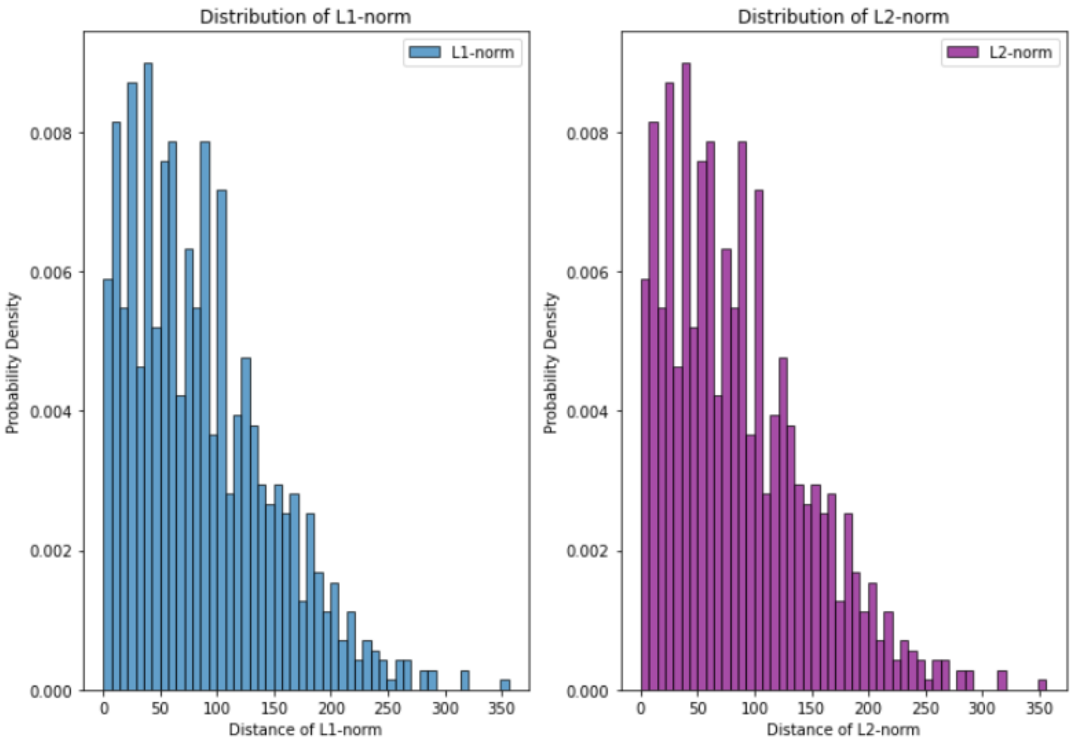

Figure 3: The L1 and L2-norm distribution of final positions after **10000** steps.

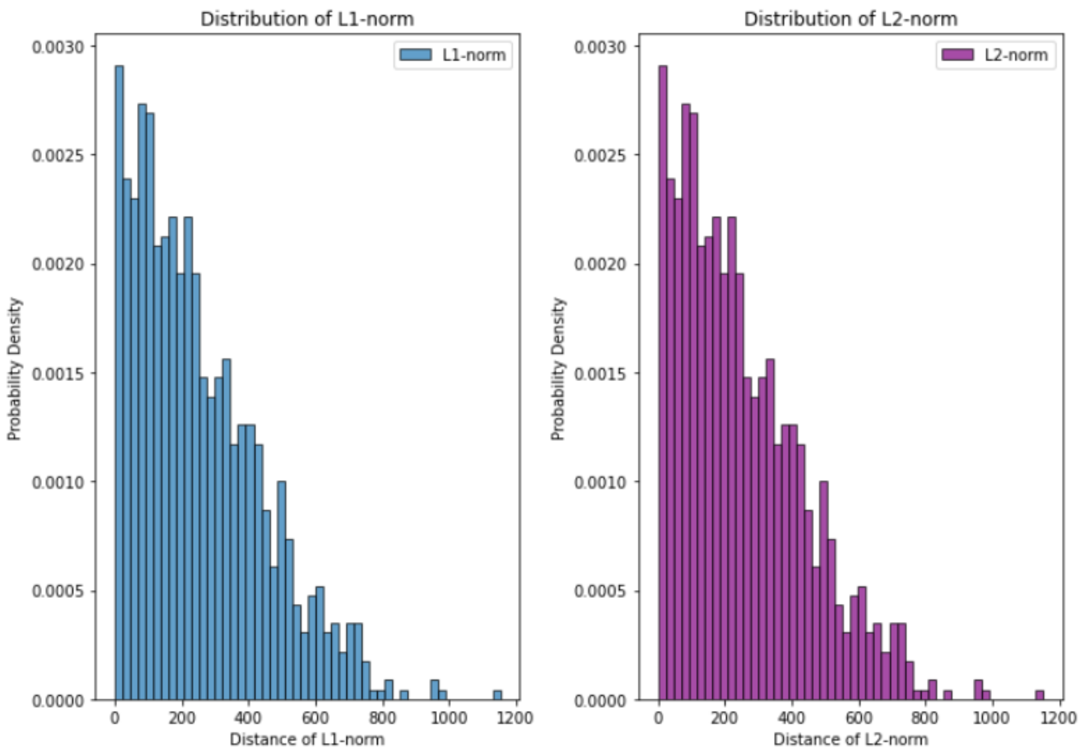

Figure 4: The L1 and L2-norm distribution of final positions after **100000** steps.

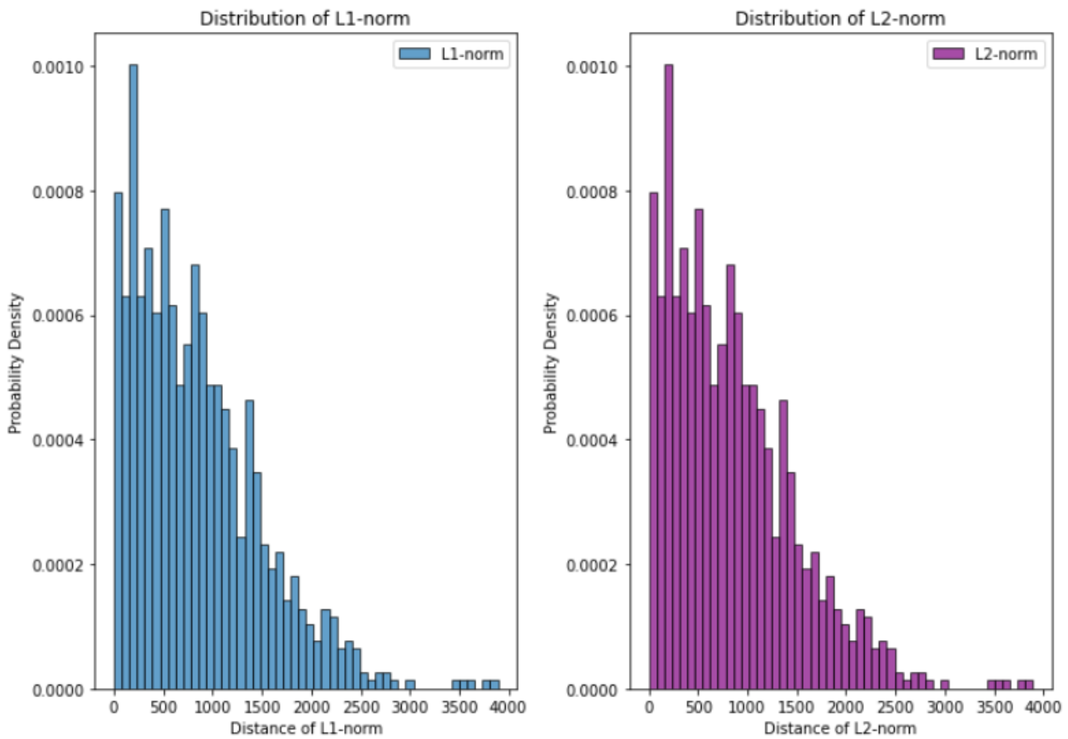

Figure 5: The L1 and L2-norm distribution of final positions after **1000000** steps.

Given the recorded time steps and the count of particle returns to the origin: **1145676**, we estimate and plot the distribution of the number of time steps required to return to the origin, and estimate the expected value of steps required for a particle to go back to origin: $\mathbf{E(time\ steps\ to\ the\ origin)}$

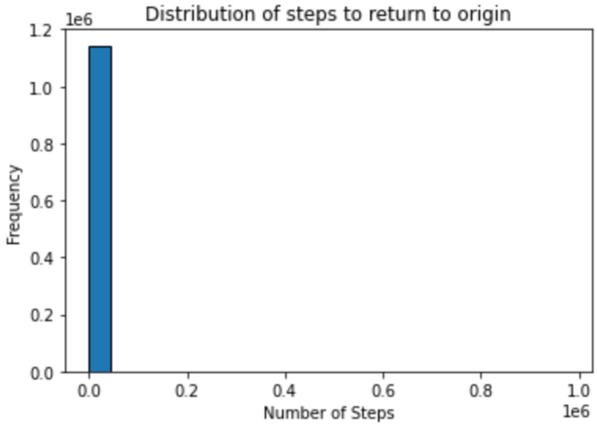

Figure 6: Distribution of steps returning to the origin.

And the expectation of time steps required to the origin: **467.750649**

Obviously, there are some outliers in the data in order to make the expected value less observation. So, we analyze the data to identify and remove outliers, we calculate the first and third percentile, and determine the IQR range. Subsequently, we calculate the lower-bound $(Q1 - 1.5 * IQR)$, and the upper-bound $(Q1 + 1.5 * IQR)$, Values falling outside these bounds will be seen as the outliers.
With our new data, we plot a new distribution without the outlier:

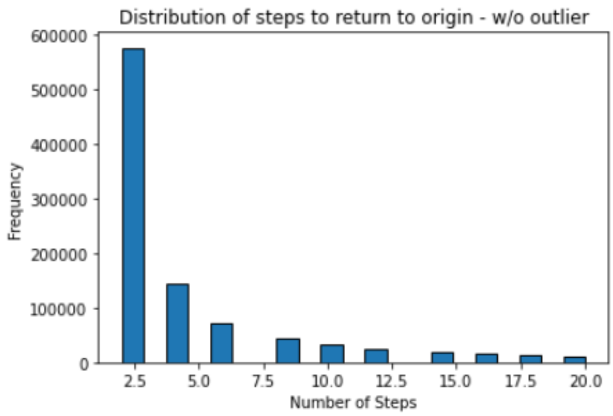

Figure 7: Distribution of steps returning to the origin w/o outliers.

Also, the new expected value of data w/o outliers is **4.277625**

Two-Dimension

Similar to the random walk in one dimension, we observe each section in which the particle is located during a single walk:

- The particle stays at section 1 for **290673626** time steps.
- The particle stays at section 2 for **279823949** time steps.
- The particle stays at section 3 for **262423063** time steps.
- The particle stays at section 4 for **274885306** time steps.

We can easily observe that the 4 values are closed approximately the same.

Next, we plot the distribution of the L1-norm and L2-norm of the final position of the particle.

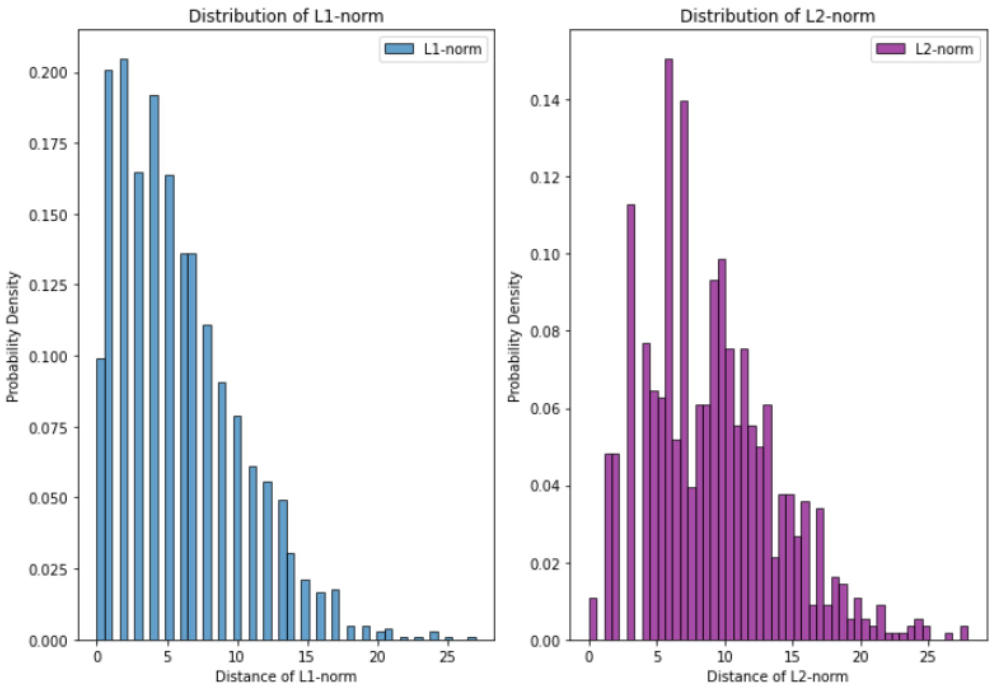

Figure 8: The L1 and L2-norm distribution of final positions after **100** steps

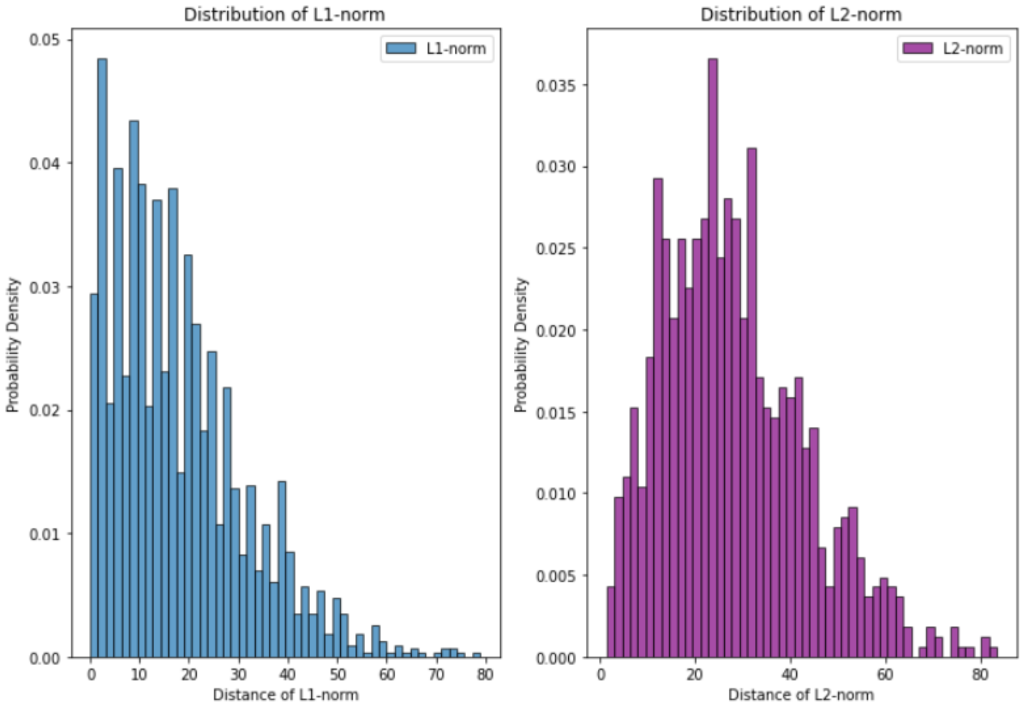

Figure 9: The L1 and L2-norm distribution of final positions after **1000** steps

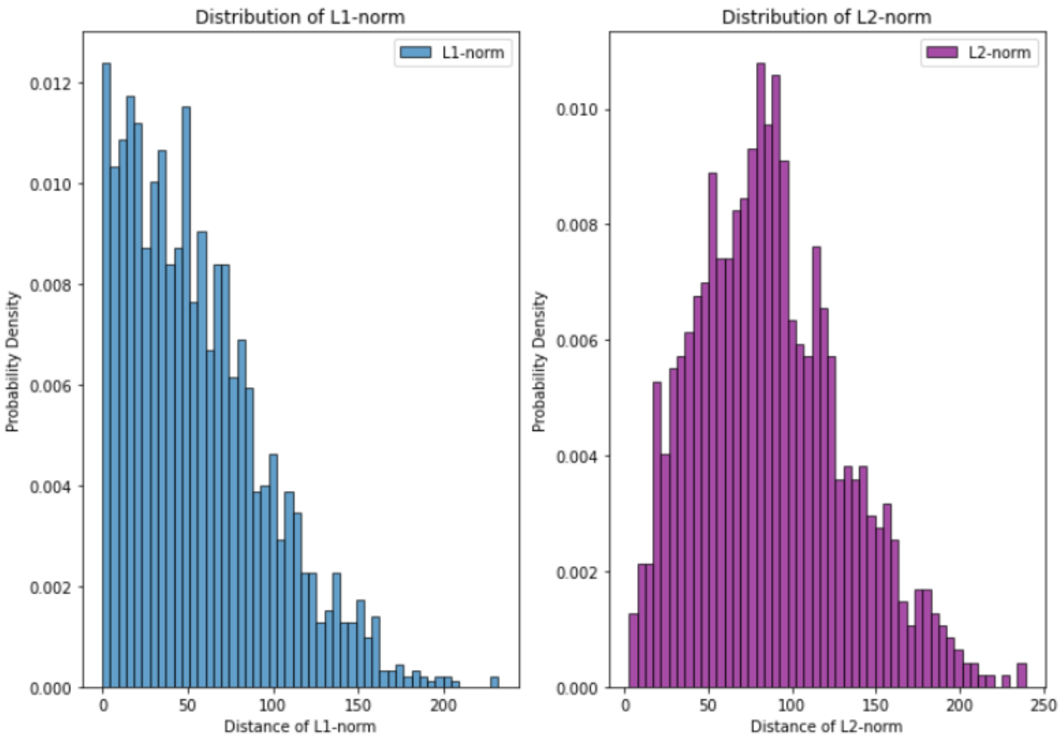

Figure 10: The L1 and L2-norm distribution of final positions after **10000** steps

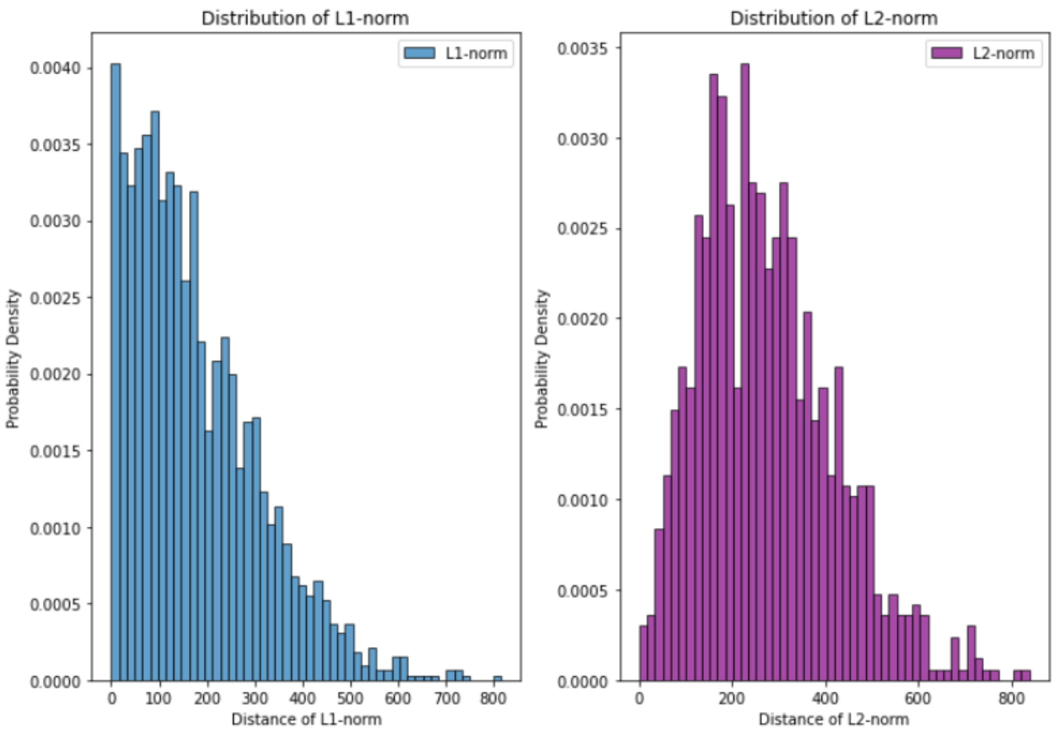

Figure 11: The L1 and L2-norm distribution of final positions after **100000** steps

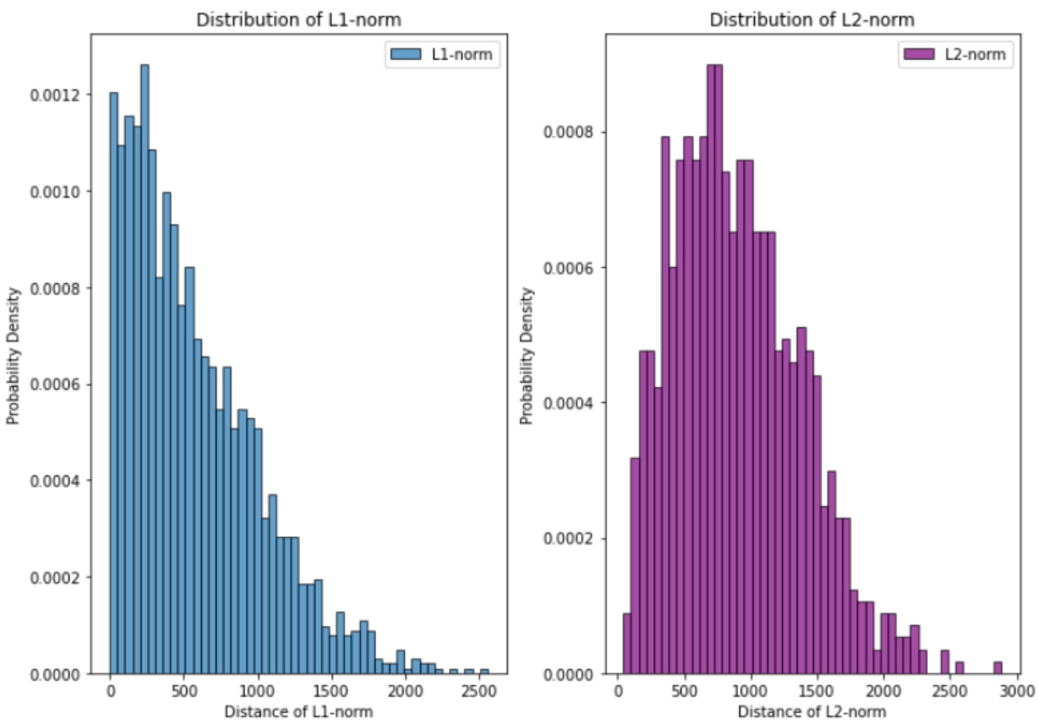

Figure 12: The L1 and L2-norm distribution of final positions after **1000000** steps

Next, we plot the distribution of steps taken by the particle to return to the origin, providing a visual representation of the data and estimate the expected value.

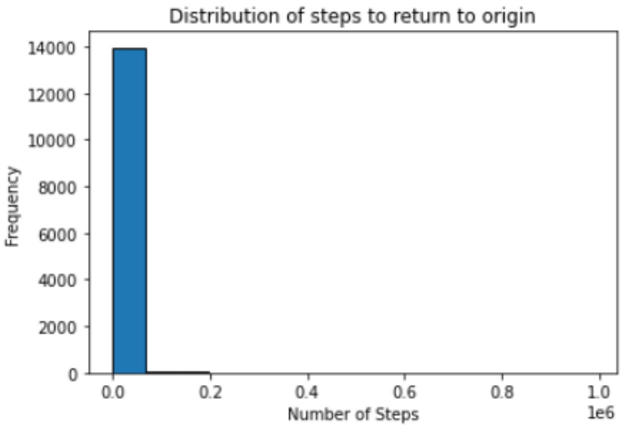

Figure 13: Distribution of steps returning to the origin.

And the expectation of time steps required to the origin: **5146.101685**

Similarly, we analyze the data to identify and remove outliers, then plot the new distribution.

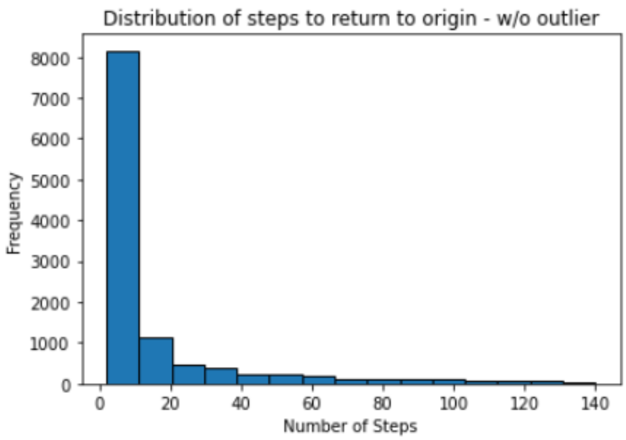

Figure 14: Distribution of steps returning to the origin w/o outliers.

Also, the new expected value of data w/o outliers is **15.160691**

- **Three-Dimension**

As above, we observe each section in which the particle is located during a single walk:

- The particle stays at section 1 for **126080914** time steps.
- The particle stays at section 2 for **139898862** time steps.
- The particle stays at section 3 for **138005071** time steps.
- The particle stays at section 4 for **149253768** time steps.
- The particle stays at section 5 for **131577785** time steps.
- The particle stays at section 6 for **143214108** time steps.
- The particle stays at section 7 for **136444821** time steps.
- The particle stays at section 8 for **140687832** time steps.

We can easily observe that the 8 values are closed approximately the same.

Then we plot the distribution of the L1-norm and L2-norm of the final position of the particle.

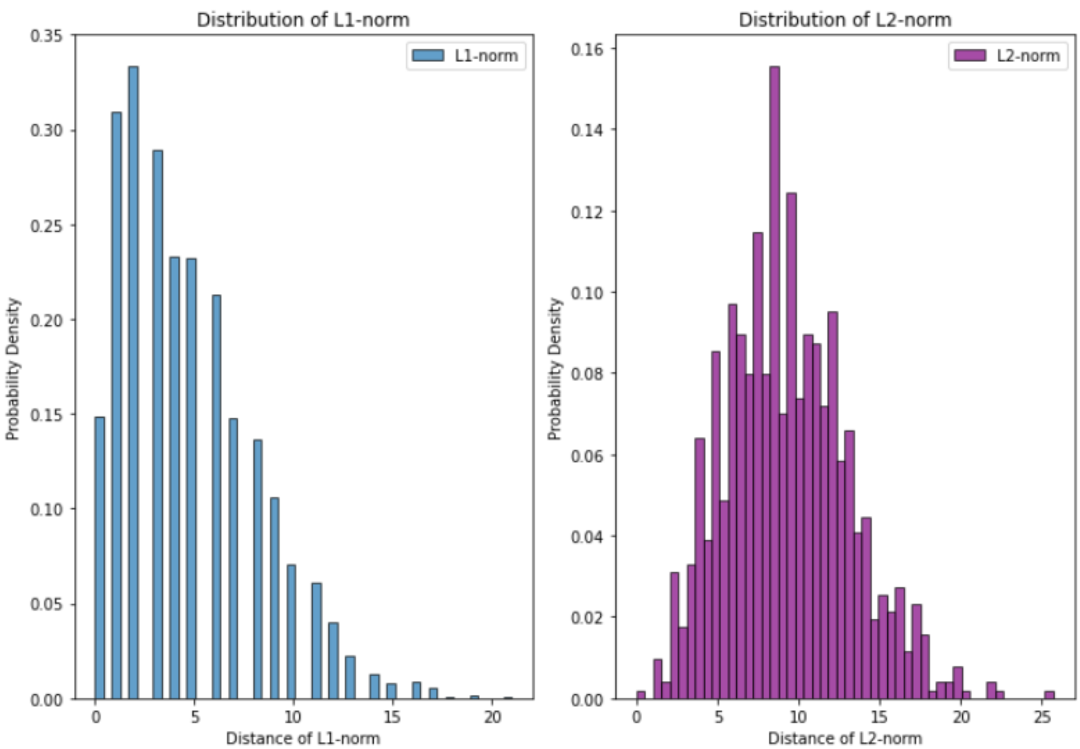

Figure 15: The L1 and L2-norm distribution of final positions after **100** steps

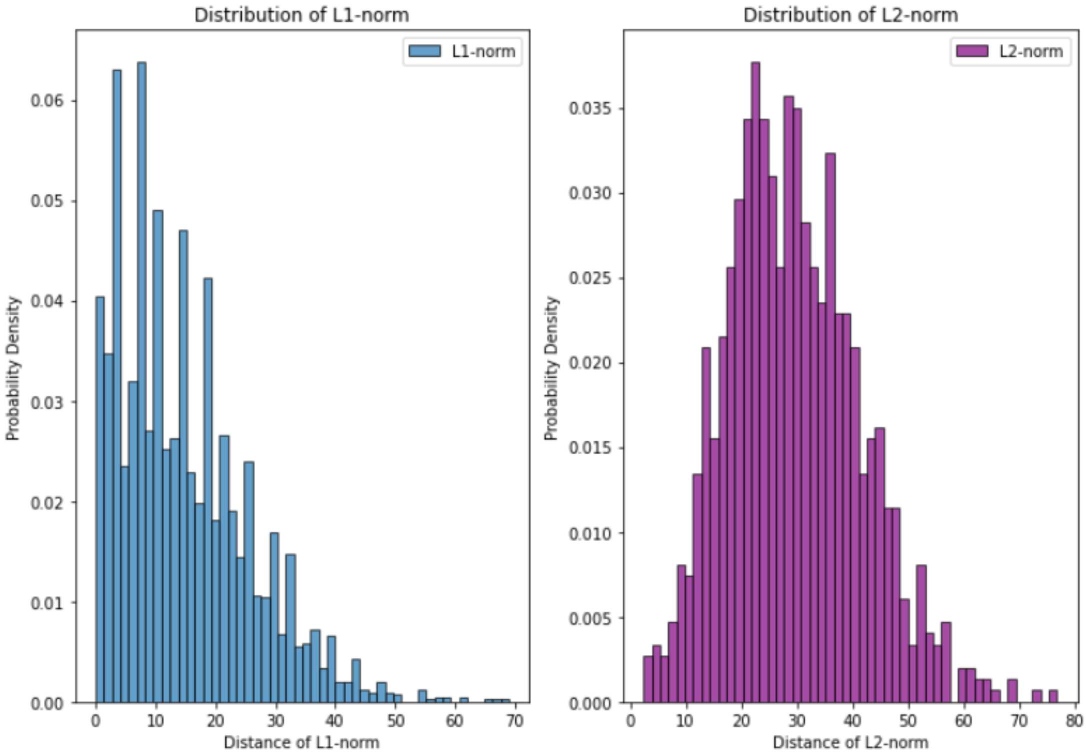

Figure 16: The L1 and L2-norm distribution of final positions after **1000** steps

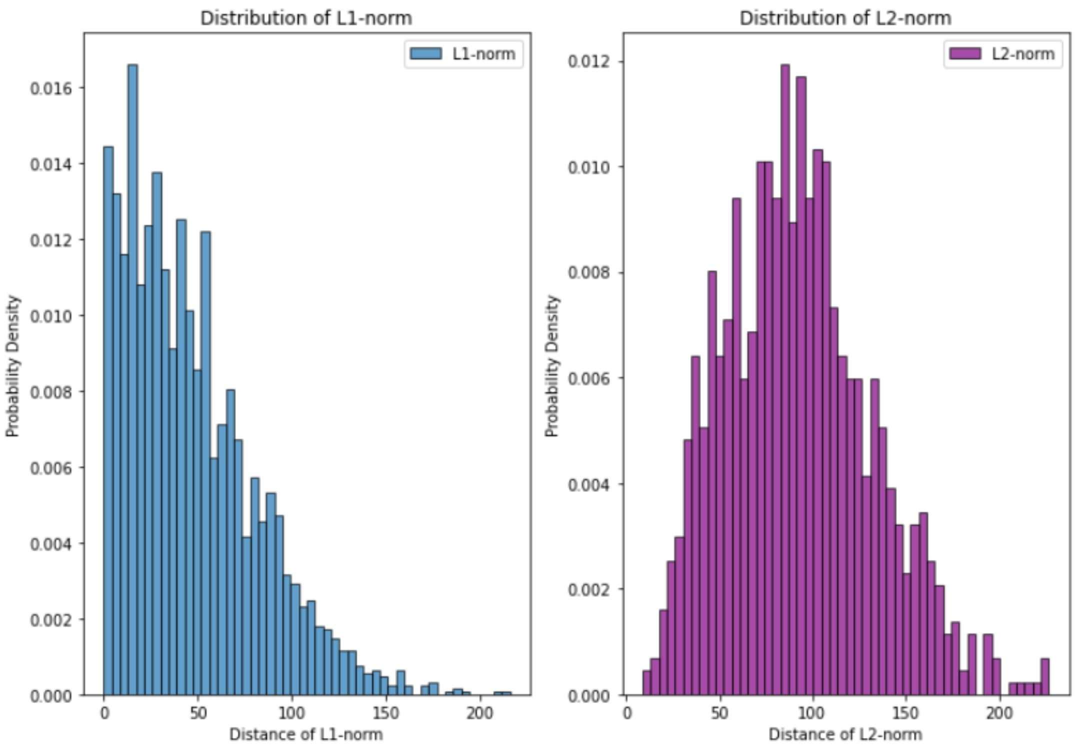

Figure 17: The L1 and L2-norm distribution of final positions after **10000** steps

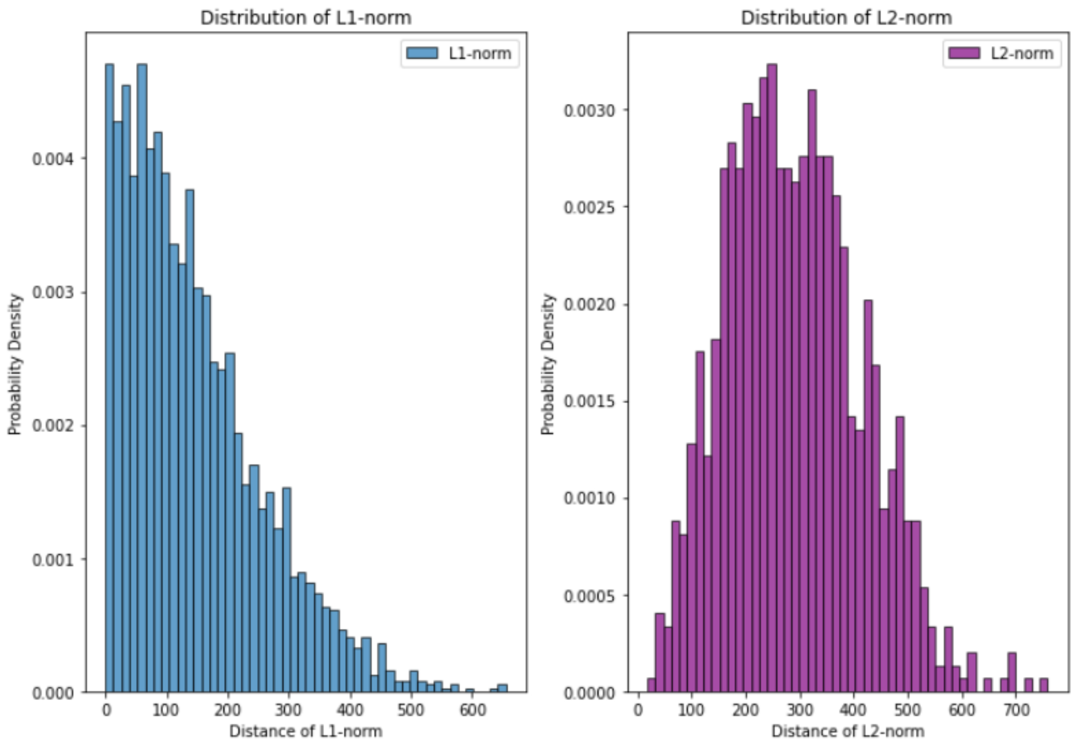

Figure 18: The L1 and L2-norm distribution of final positions after **100000** steps

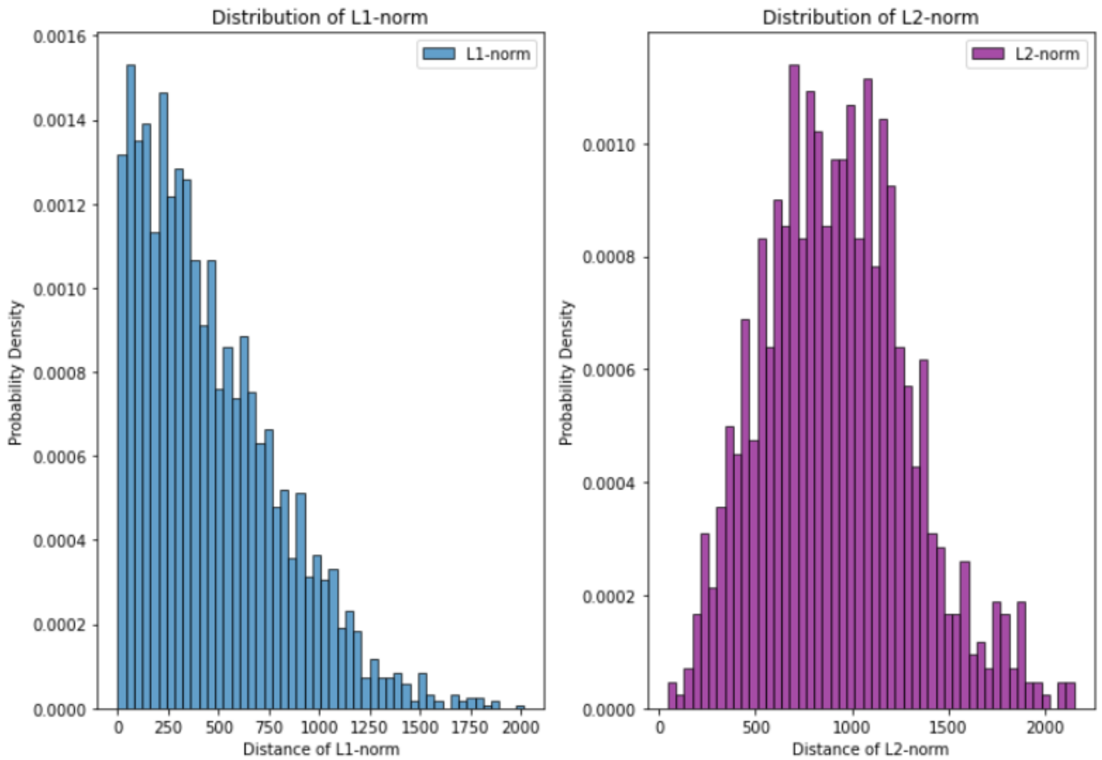

Figure 19: The L1 and L2-norm distribution of final positions after **1000000** steps

Given that the particle returned to the origin 2748 times, we proceed to plot the distribution of the steps taken by the particle to return to the origin and estimate the expected value:

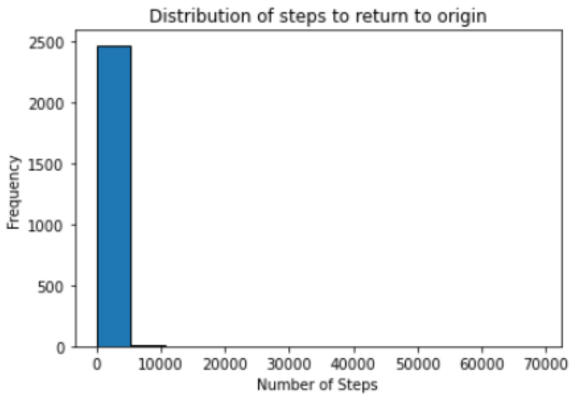

Figure 20: Distribution of steps returning to the origin.

And the expectation of time steps required to the origin: **109.204197**

Also, we identify and remove the outlier data point and plot the distribution again:

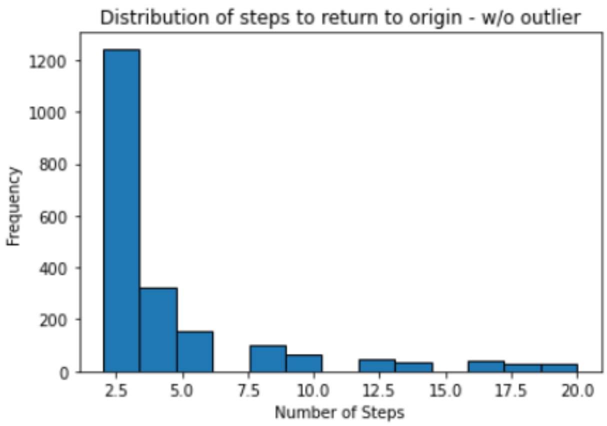

Figure 21: Distribution of steps returning to the origin w/o outliers.

The new expectation of time step required to the origin is **4.309408**

Interestingly, we observed a notable drop in the expected value from 15.16 to 4.31 when comparing with the two-dimensional case. The possible reason is that the occurrence of returning to the origin is much rarer and more elusive in higher dimensions, the only chance is that the initial steps select the complementary pairs of steps (e.g. (1,0,0) and (-1,0,0)), resulting the decrease of the expected value.

Four-Dimension

Similarly, we observe each section in which the particle is located during a single walk:

- The particle stays at section 1 for **67074185** time steps.
- The particle stays at section 2 for **63768474** time steps.
- The particle stays at section 3 for **67785508** time steps.
- The particle stays at section 4 for **69703077** time steps.
- The particle stays at section 5 for **67747172** time steps.
- The particle stays at section 6 for **59429419** time steps.
- The particle stays at section 7 for **65423178** time steps.
- The particle stays at section 8 for **80241151** time steps.
- The particle stays at section 9 for **75195405** time steps.
- The particle stays at section 10 for **62975351** time steps.
- The particle stays at section 11 for **70373811** time steps.
- The particle stays at section 12 for **70367588** time steps.
- The particle stays at section 13 for **72828688** time steps.
- The particle stays at section 14 for **65855557** time steps.
- The particle stays at section 15 for **65530671** time steps.
- The particle stays at section 16 for **77676655** time steps.

We can easily observe that while there are some differences between sections, the values of the 8 sections are close to each other.

Next we plot the distribution of the L1-norm and L2-norm of the final position of the particle.

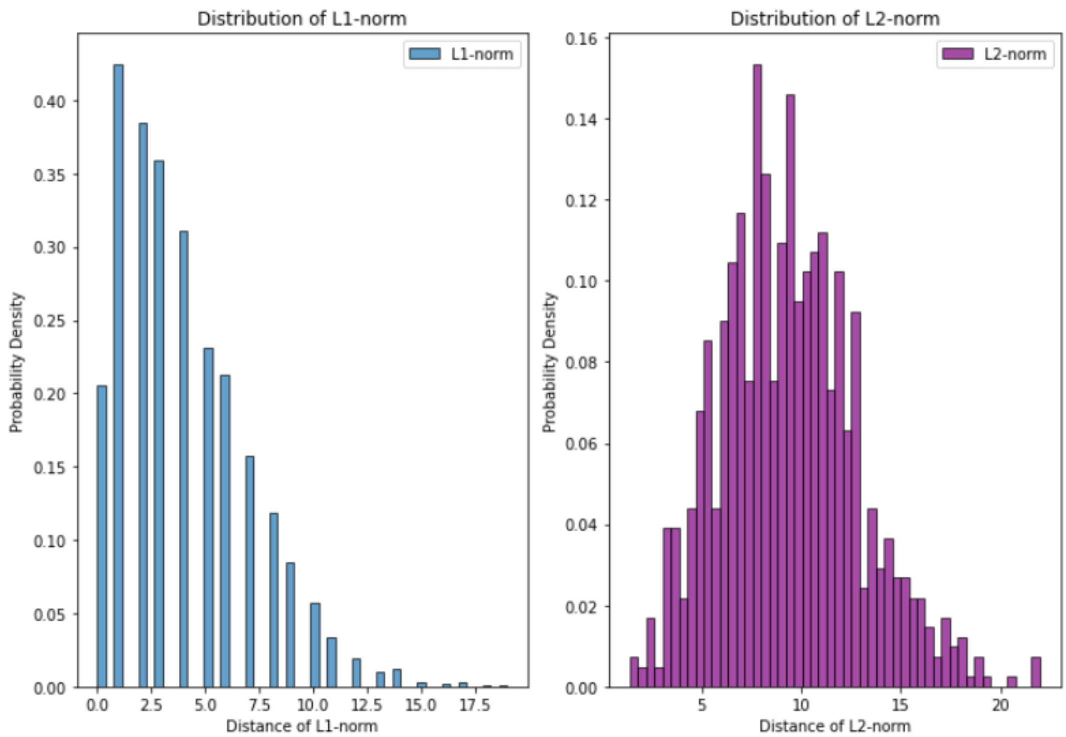

Figure 22: The L1 and L2-norm distribution of final positions after **100** steps

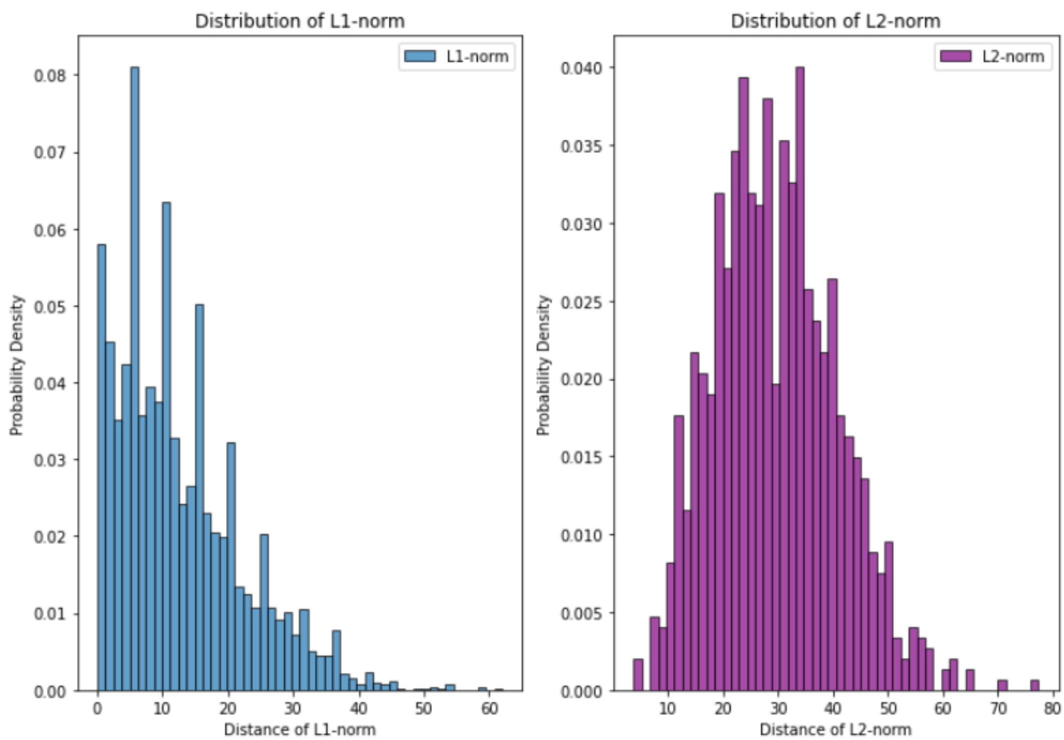

Figure 23: The L1 and L2-norm distribution of final positions after **1000** steps

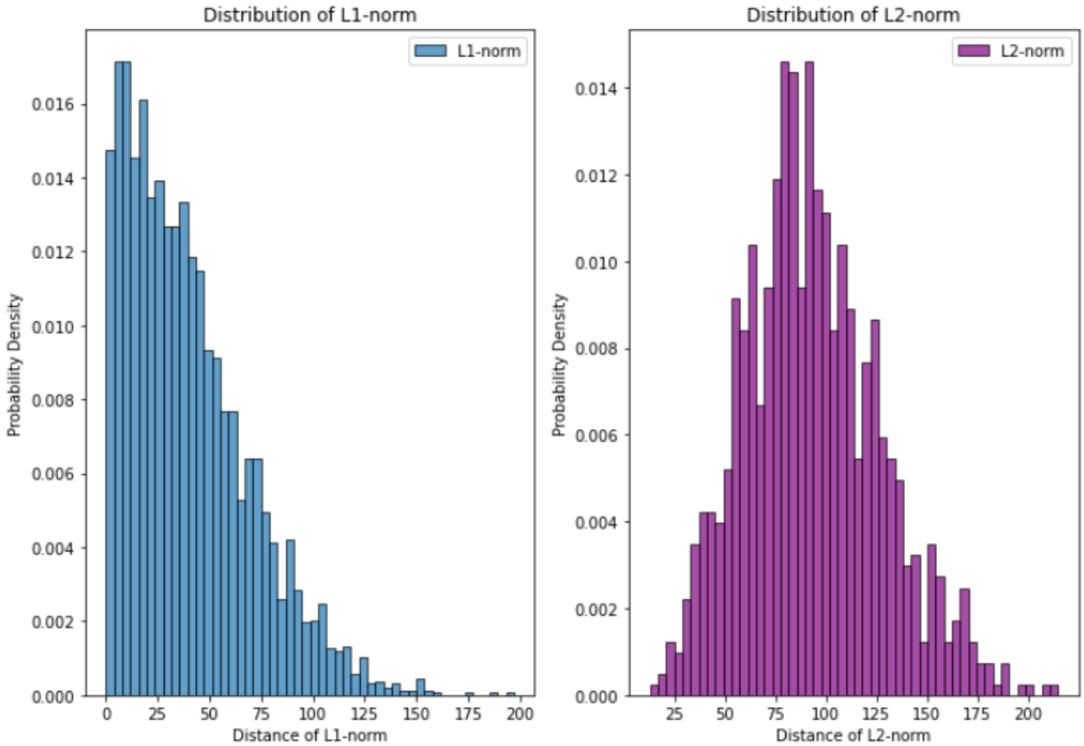

Figure 24: The L1 and L2-norm distribution of final positions after **10000** steps

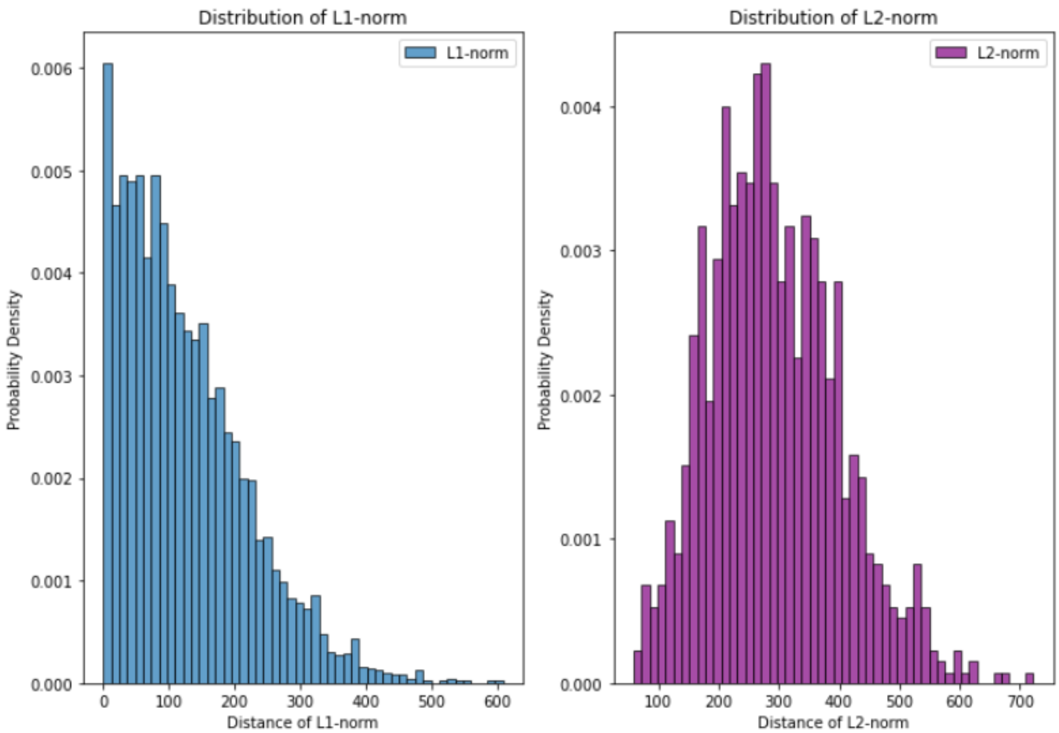

Figure 25: The L1 and L2-norm distribution of final positions after **100000** steps

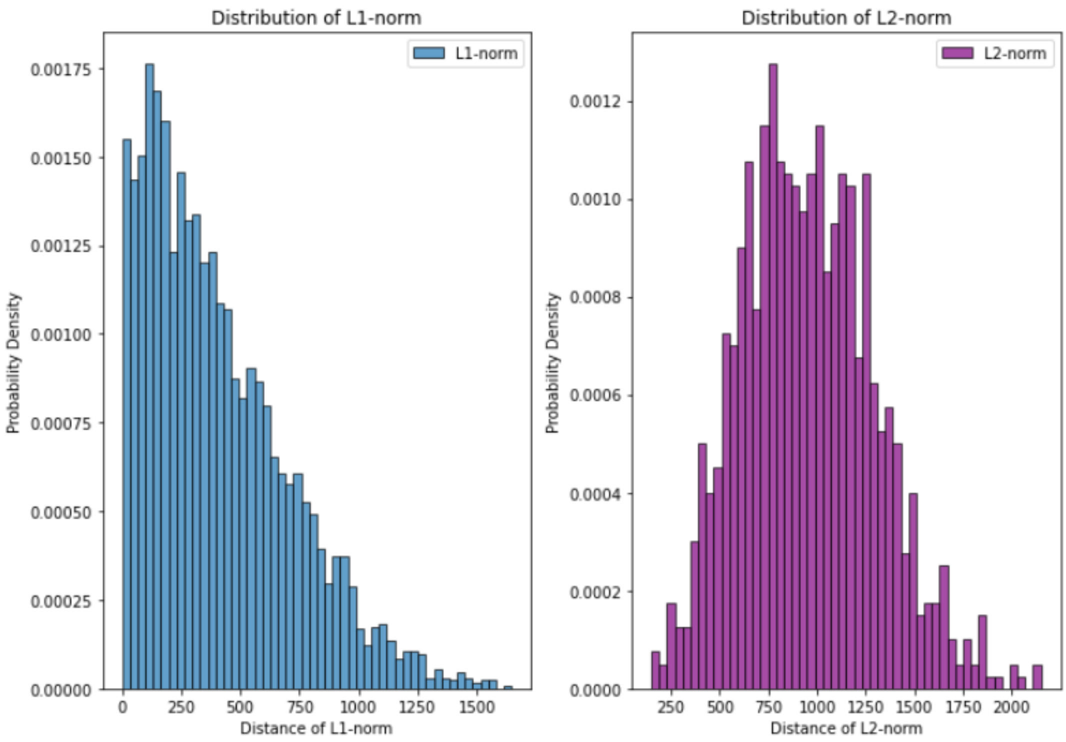

Figure 26: The L1 and L2-norm distribution of final positions after **1000000** steps

Given that the particle returned to the origin 1183 times, we proceed to plot the distribution of the steps taken by the particle to return to the origin and estimate the expected value:

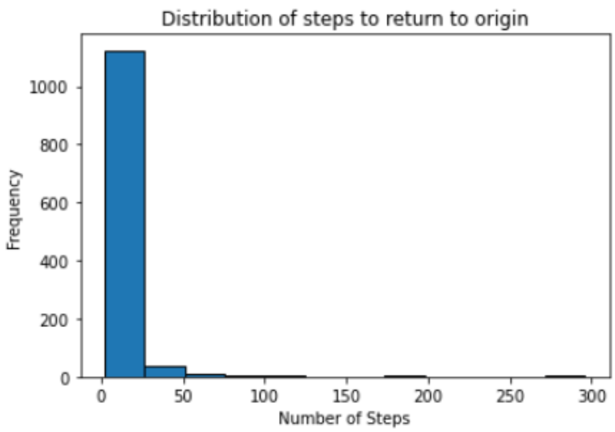

Figure 27: Distribution of steps returning to the origin.

The expectation of time steps required to the origin: **7.486052**

Again, we identify and remove the outlier data point and plot the distribution again:

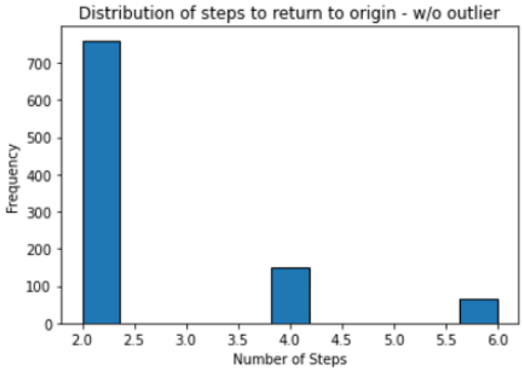

Figure 28: Distribution of steps returning to the origin w/o outliers.

The expectation of time steps required to the origin: **2.575820**

In four dimensions, the circumstances become more evident. The expected value is closely to 2, indicating that most of the particles returning to origin rely on the initial two steps, so we can conclude that in this case, the data point of 2 becomes an **"outlier"**, making it more challenging to observe the true expected value accurately.

Observing the above condition, we try to removed data point "2" from the steps-to-the-origin dataset, replot the distribution of time steps, and recalculate the expect value for dimension 1st and 4th to observe how the circumstances affect the results. (**Note**: Because I re-execute the random walking code blocks, the distribution and the expectations show **slightly different** from the values obtained in the previous analysis.)

- ## One Dimension

Following the analysis, I recorded the number of steps it takes for the particle to return to the origin and removed the instances where this occurred in 2 steps.

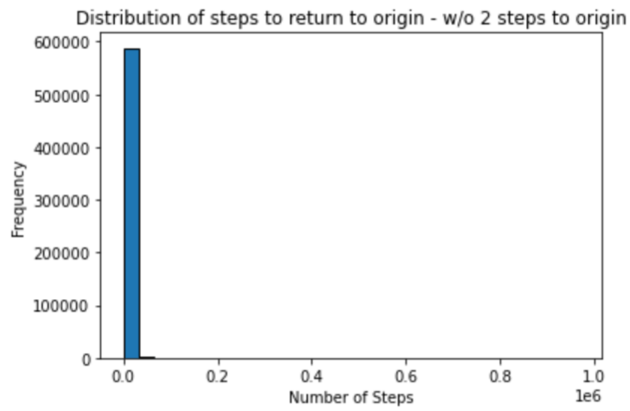

Figure 29: Distribution of steps returning to the origin w/o 2-steps

And the new expectation of time steps required to the origin: **948.931785.**

Note that the expect values of raw dataset and outlier-removed dataset are **473.917058** and **4.280610,** respectively. The expectation increases drastically. For each round in the one-dimension space, there is a probability 1/2 that the particle will return to origin in 2 steps, which take a large proportion of the return-to-origin events, significantly constraining the expect value. by removing them, we can better observe the true expectations of one-dimension random walking.

- ## Four Dimension

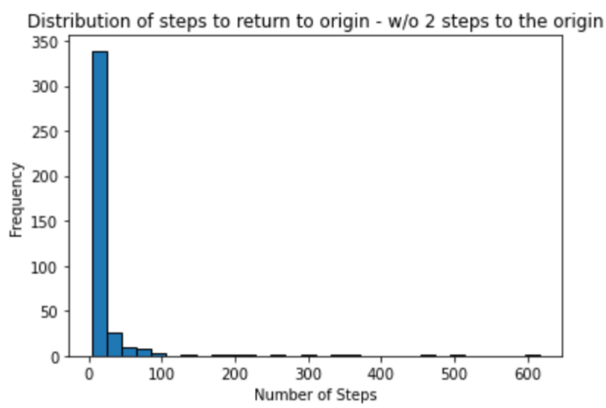

Figure 30: Distribution of steps returning to the origin w/o 2-steps

the new expectation of time steps required to the origin: **21.596977**

Comparing to the expectation of original and processed data, which are **8.782912** and **2.575820**, respectively, the new expect value increases but **not significantly**. This is because, in higher dimension, the particle has more directions to move randomly, making it more difficult to return to the origin after an amount of steps.

● Martingale in One-dimension

Since the random walking process is fair, the expected position after each step is 0, which relates to the statistical notion of **martingale**. To observe how it works, we record the position n_-, n_+ as the negative and position sections the article located, define $\mathbf{m \stackrel{\text{def}}{=} \frac{1}{2}n_0 + \max(n_-, n_+)}$. We then plot the distribution of m/n using 1000 samples for different length n, specifically: $n = \{10^2, 10^3, 10^4, 10^5, 10^6\}$.

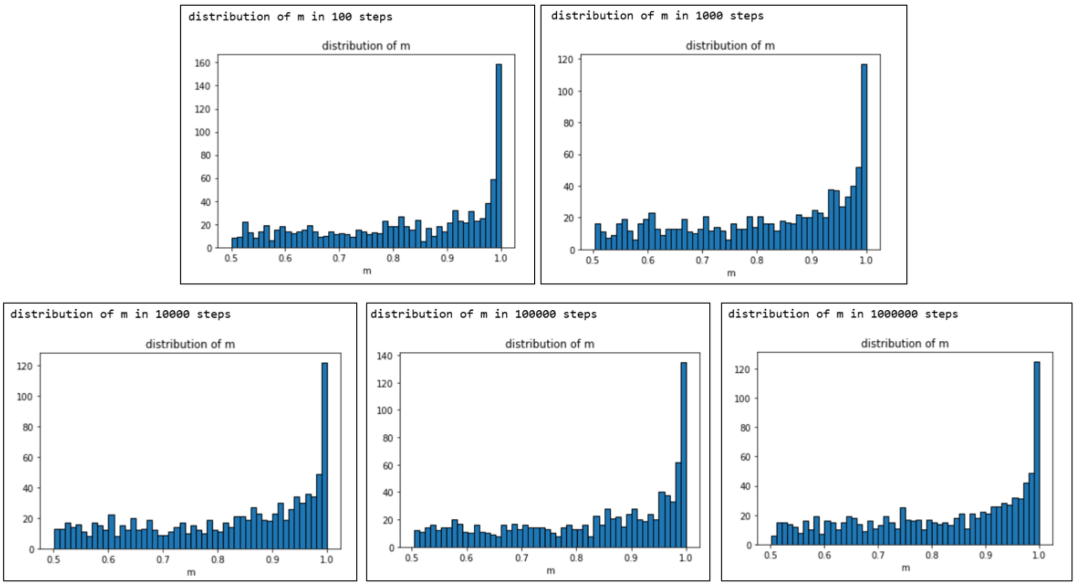

Figure 31: Distribution of m/n

We can observe that the m/n distribution has a lower bound at 0.5 and a peak at 1.0, and as n increases, the distribution curve become smoother. It is interesting to note that when we observe n_0, n_- and n_+ respectively, which the particle spent most of the time in specific section in some case (e.g. $[n_0, n_-, n_+] = [4, 95, 1]$). This implies that m/n will be close to 1.0. Additionally, if we dig into the final positions of each particle, most of the positions are close or to 0, indicating that the probability of choosing positive and negative step is fair, which demonstrates that the random walking process follows a notion of martingale.