Computational Biology Assignment 2

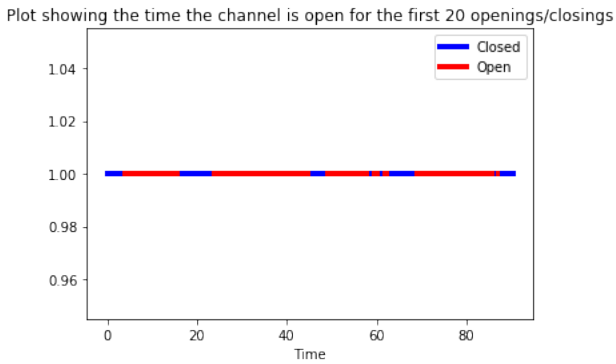
1604827
CS904 Computational Biology

February 19, 2020

Task 1

A

For this task I have used $T = 100000$, in which approximately 7500 events take place in that time.



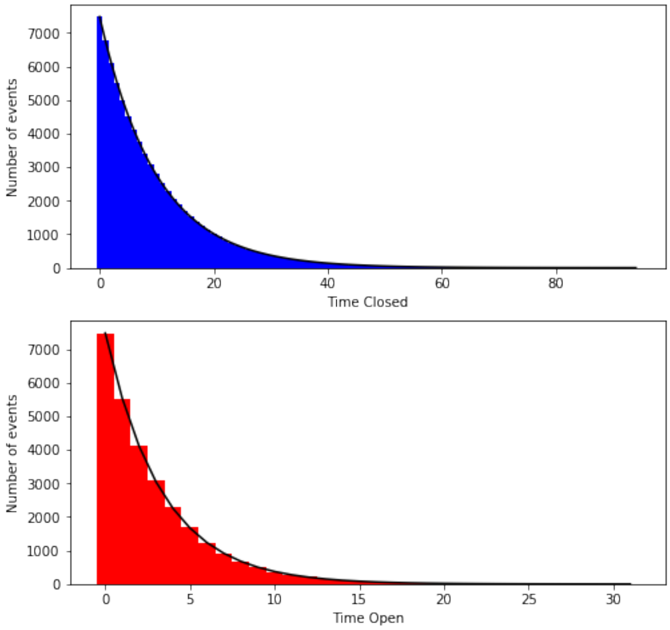
B

The results of running with $T = 100000$ produces an estimate of average open time of 3.34148 with the actual value being 3.33333 and the estimate of average closed time being 10.024446, this shows that the simulation is working as expected.

C

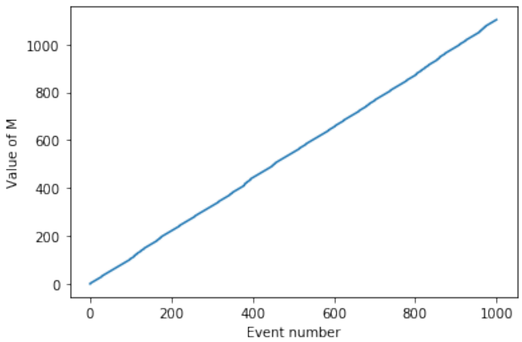
The black lines in the plots represents the theoretically expected distributions and the histogram represents the cumulative number of events that lasted at least that length of time

Figure 1: Top: Histogram of time closed, Bottom: Histogram of time open

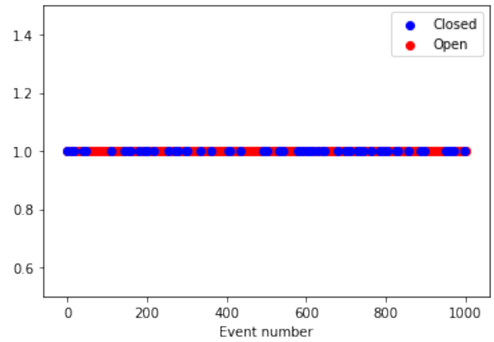


Task 2

Producing simulations until M reaches 1000. This figure shows the increase in the value of M for each time an event occurs.



This plot shows for which of the events the channel is open or closed. (As this shows every event the scatter points are very close, I chose to display the results like this as visualising all events like in task 1 it would be indistinguishable between being open constantly and the true nature of the graph, hence the closed points are disproportionate).

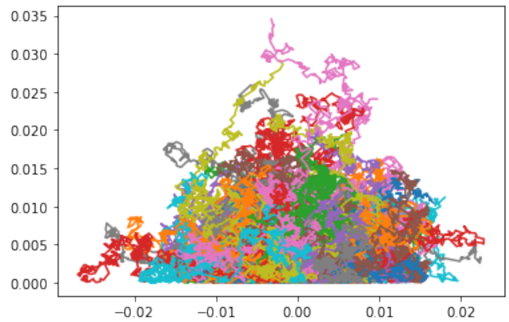


Task 3

The diagrams in A and B shows the path of each of the calcium ions. Where each of the particles move in a random direction by a distance proportional to the time between each event.

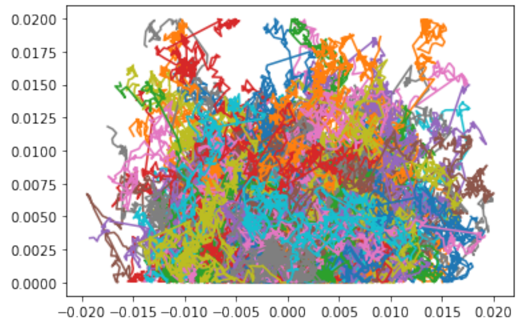
A

This shows the diffusion of the ions given a membrane on the y axis.



B

This shows the diffusion of ions given a membrane in the shape of a box. The box as can be seen, is sized between -0.02 and 0.02 for the x co-ordinate and between 0 and 0.02 for the y co-ordinate.



C

Obviously ions that have only entered a small amount of time before don't have the time to move very far away from the origin, hence after a small amount of time the msd is very small and the msd increases with time.

Figure 2: MSD for part A

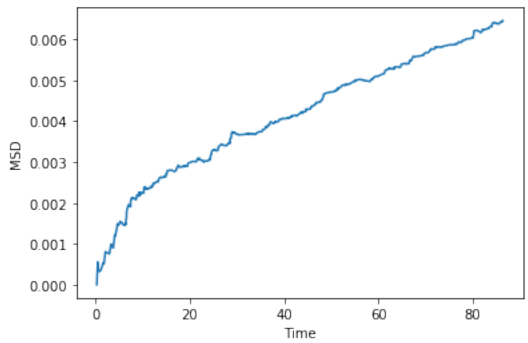
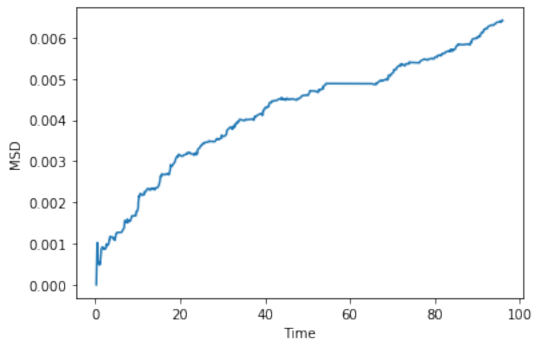


Figure 3: MSD for part B

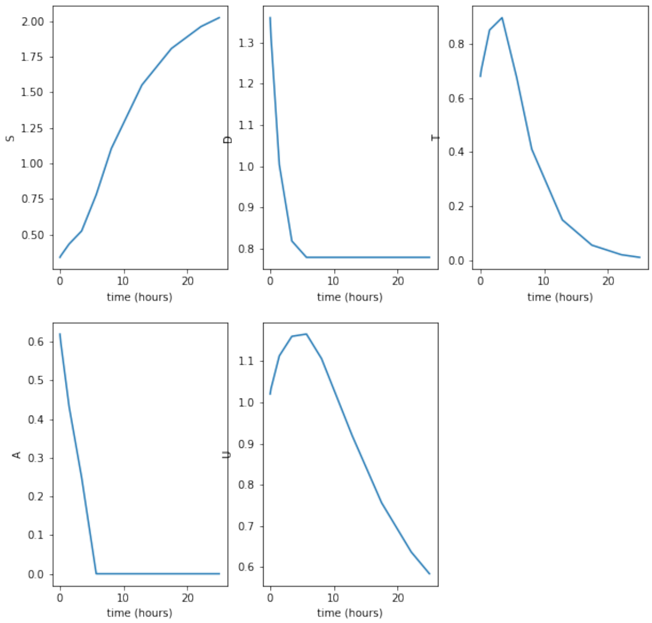


Looking at the msd in relation to the equation $msd(t) = 2nDt^{\alpha}$ it can be seen that for part A and B the graphs are close to being linear as expected.

Also as can be seen from the graphs α is less than 1 and sub-diffusion is taking place.

Task 4

As can be seen from the graphs below the period for the clock is approximately 21 hours.



Task 5

A

Considering the for a problem of this type there are two parameters, the intensity of each cell I and its position $\mathbf{x}$. However, the position of an object in 2D can be give by its position in the euclidean plane. For instance by taking the centre of the frame and drawing two perpendicular lines you can obtain a set of axis. This problem now has three parameters to consider an x co-ordinate a y co-ordinate and I the intensity.

Now by following the hint in the question using an adaptation of the mean squared distance can be used. For an image $I_1 = (x_{1_i}, y_{1_i}, I_{1_i})$ where $i \in [1, N]$ where N is the number of cells in the image. A suitable cost function would be:

$$cost(I_1) = \frac{1}{N} \sum_{i=1}^{N} ((x_1 - 0)^2 + (y_1 - 0)^2 + (i_1 - 0)^2)^{1/2}$$

This is a adaptation of the euclidean distance between two points but generalized into three dimensions, as to include the intensity of the cell as a factor, and with aspects of the MSD algorithm given in lectures. This cost function fixes a point in the centre and fixes the intensity as 0 and compares all images to this.

This method will be able to give the distance between two images, in a way such that images that are similar i.e. the cells are in similar positions and intensities, have a similar cost (defined by the function above) associated to them than images that are very different.

This can be used to solve the assignment problem by placing images that are most similar (have the smallest cost) next to each other.

Now the Hungarian algorithm can be used to solve the assignment problem of matching corresponding cells, using this cost function.

B

To deal with the problem the cells that are near the edge could be disregarded. To do this make a smaller frame the size of the largest frame that doesn't include any cells that are leaving or entering in all the images, and look at the cost matrix of that smaller frame (including points that are on the boundary of this smaller frame, none of them will be on the actual boundary). This should offer the most accurate way of measuring the cost as there are no cells that have a centre in a position of the cell that isn't the centre (i.e. the centre of half of a cell is the centre of that half not of the actual cell). This way does allow for the increase and decrease in the number of cells but also only allows for cells that have an accurate description (position and intensity) to be included.