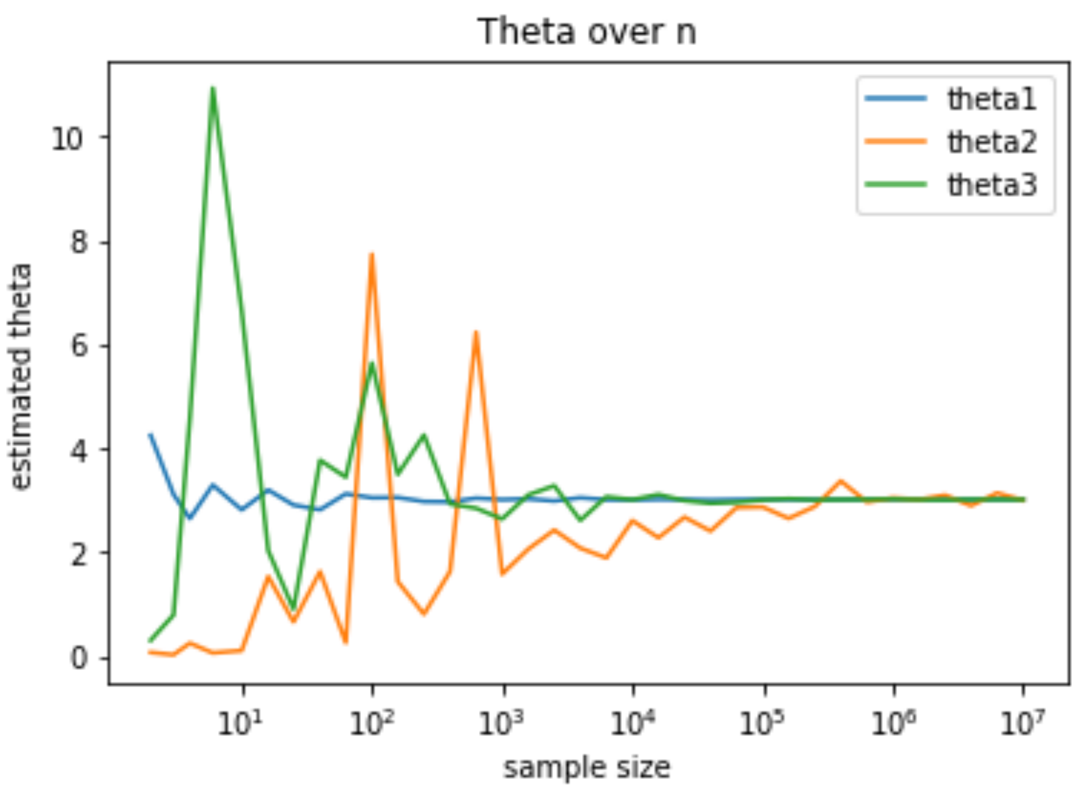
Stats202C-Project1 Report

1. Importance sampling and the effective number of samples:

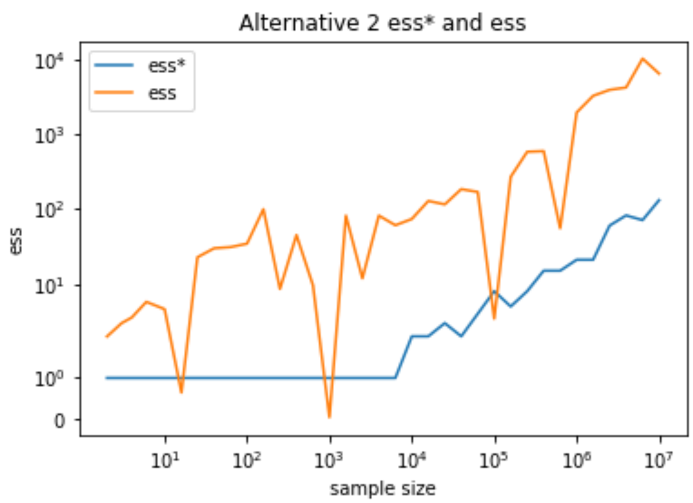
    a) The figure of comparing the convergence rates of the three Alternatives is shown below. Before running the experiment, I guessed Alternative 3 is more effective than Alternative 2 because Alternative 3 has a larger standard deviation of 4 compared to standard deviation of 1 in Alternative 2, which will lead to a higher chance of falling into the region of the target distribution.

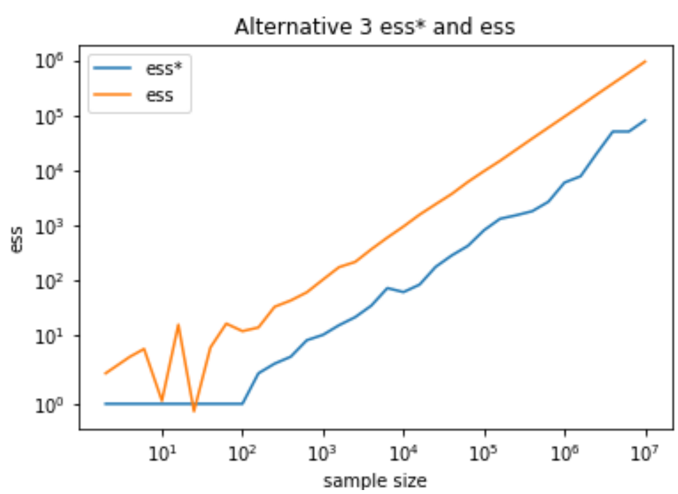
    From the figures we can see that the theta converges fastest by drawing samples directly from the target distribution (theta1, blue line). The second fastest is theta3(green line), which converges after approximately 10^4 samples are drawn. Theta2 converges the slowest(orange line).



b) To estimate the "effective sample size" for each alternative, first we use ess∗(n1) = n1 as the truth, since the samples in Alternative 1 are each "effective" samples of size 1 as they are directly drawn from the target distribution. ess* (n2) and ess*(n3) are the sample sizes required to reach the estimated errors as the same level in Alternative1.
The approximation to ess* (n2) and ess*(n3) -- ess (n2) and ess(n3), are calculated as the suggested estimate in BZ.

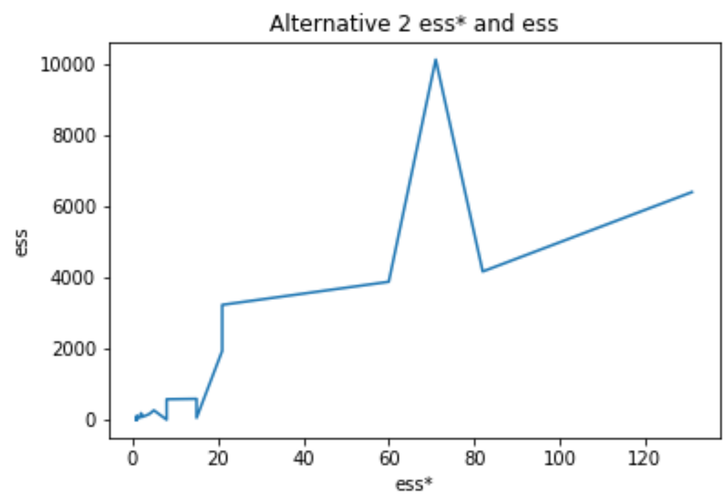
From the two figures below we can see that ess is higher than ess* in both figures. This shows that the approximation of the effective sample size is higher than the true sample size. So the estimator suggested in BZ is too large for an effective sample size.

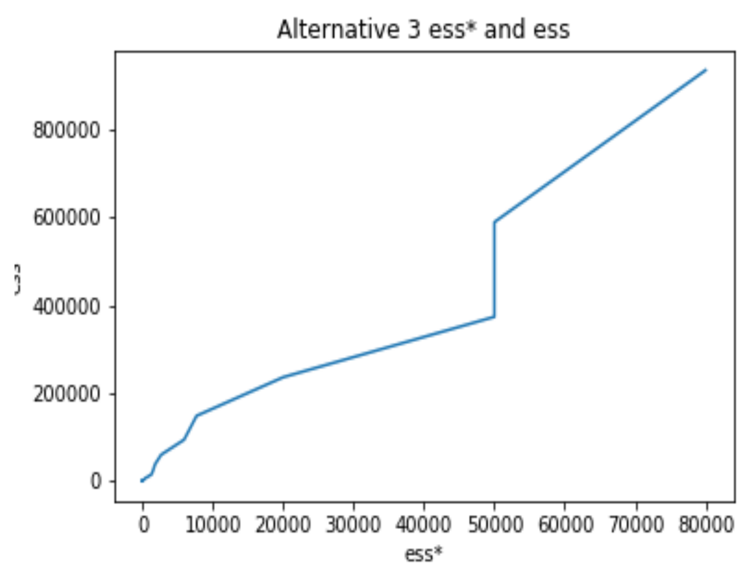




c) Plot ess*(n_2) over ess(n_2), and ess(n_3) over ess*(n_3) .

From the two plots below we can see that, the ratio of ess* to ess for Alternative 2 is approximately 50, and this ratio is approximately 10 for Alternative 3. This indicates that for a given sample, Alternative 3 requires a smaller effective sample size than Alternative 2 would require. This also conforms with the findings from part a)—Alternative 3 converges faster than Alternative2 since Alternative 3 is more effective in generating effective sample sizes.





2. Estimating the Number of Self-Avoiding Walks in an (n+1)*(n+1) Grid.

a)I used three designs in the textbook. I sampled M=10^7 samples for each design. K--the estimated number of SAW, for each design is shown below.

Design1: uniformly choose a direction at each step.

K = 3.463*10^25

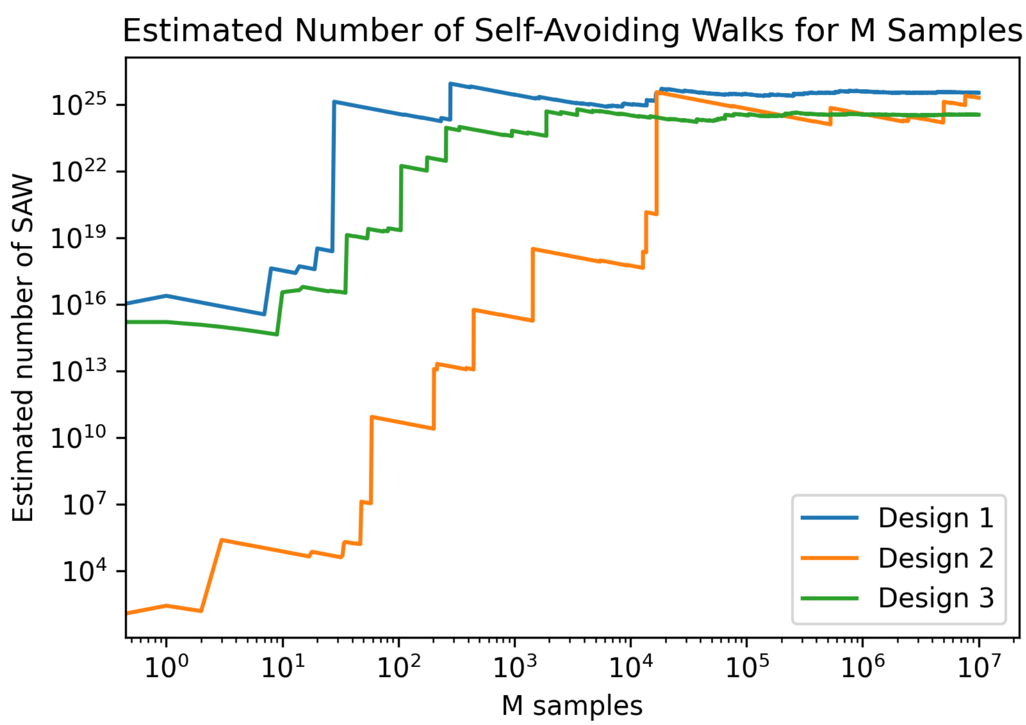
Design2: introduce an early termination rate of 0.1

K = 2.047*10^25

Design3: for walk longer than 50, generate five childrens and reweight each children by a factor of 1/5.

K = 3.539*10^24

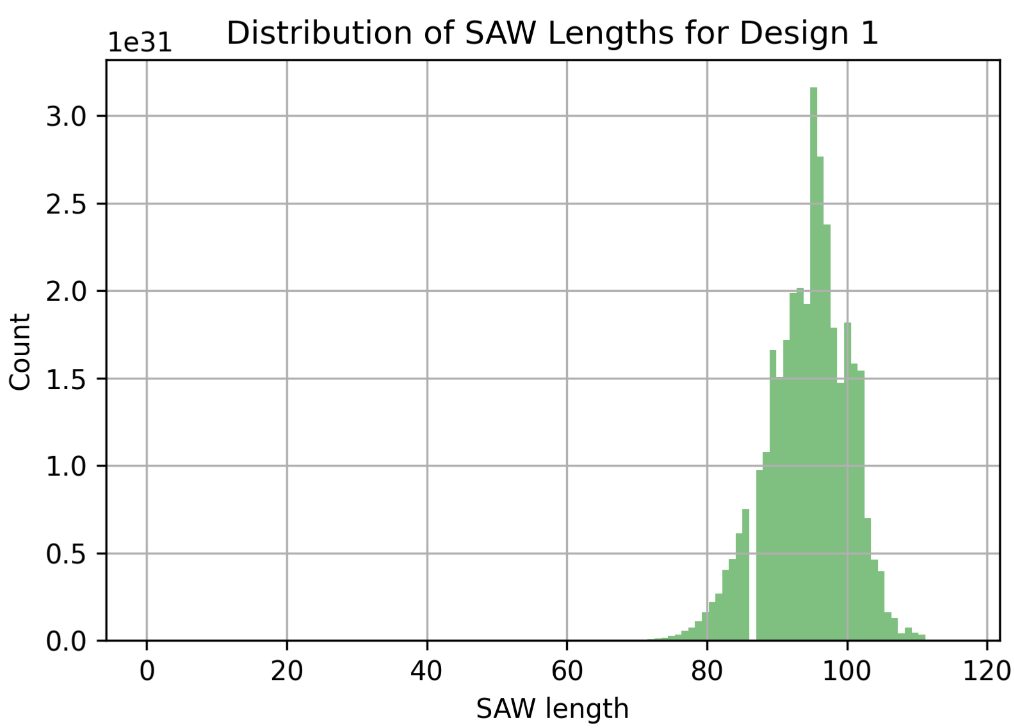
The plot for K against M is shown below:

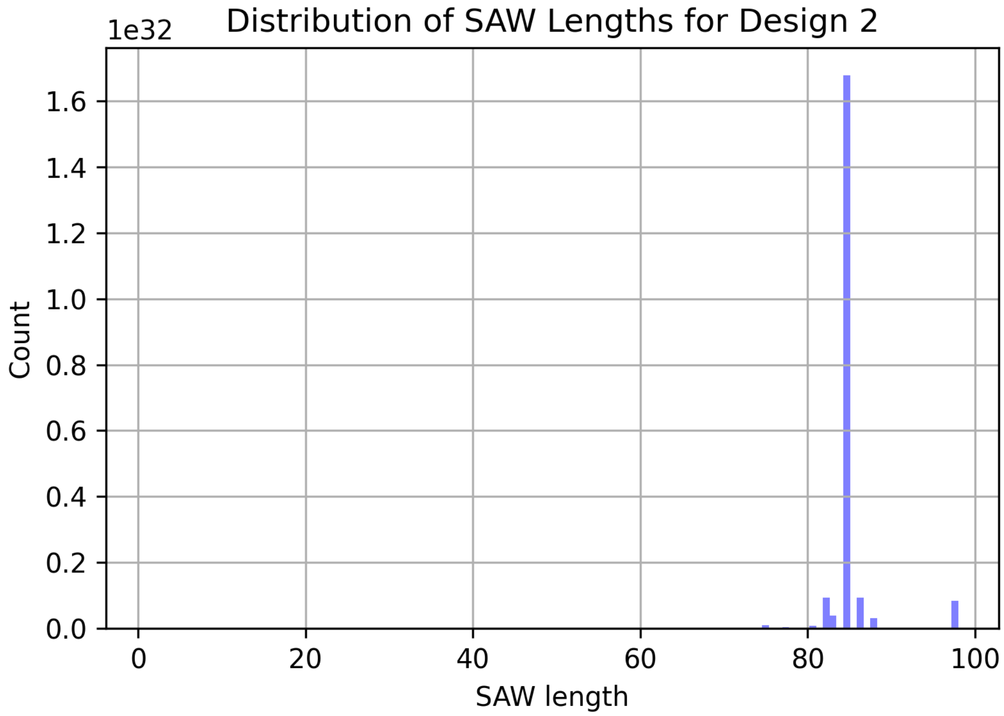


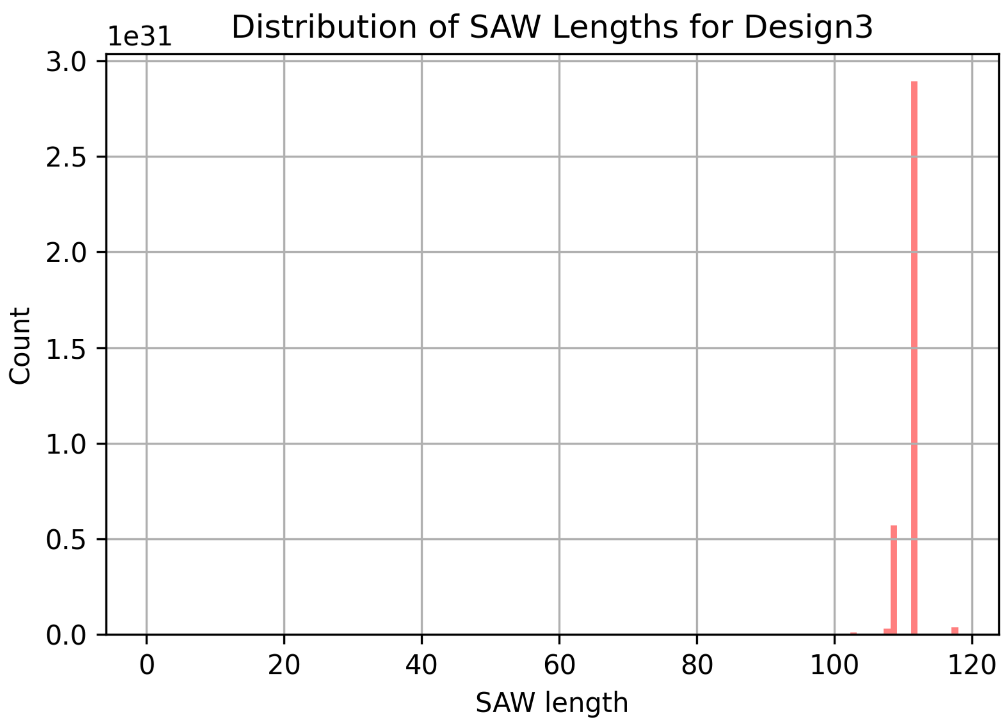
From the plot we can see the convergence rate for Design 1 and Design 3 are similar. But Design 2 has a much slower convergence compared to Design 1 and Design 3. This is because Design 2 imposes an early termination rate and so the paths are much shorter compared to the other designs. Shorter paths have lower weights, hence the convergence are slow.

b) I used design 1 to calculate the total number of SAWs that start from (0, 0) and end at (n, n). The result I got is 1.579*10^25, which is higher than the true value of 1.5687 x 10^24.

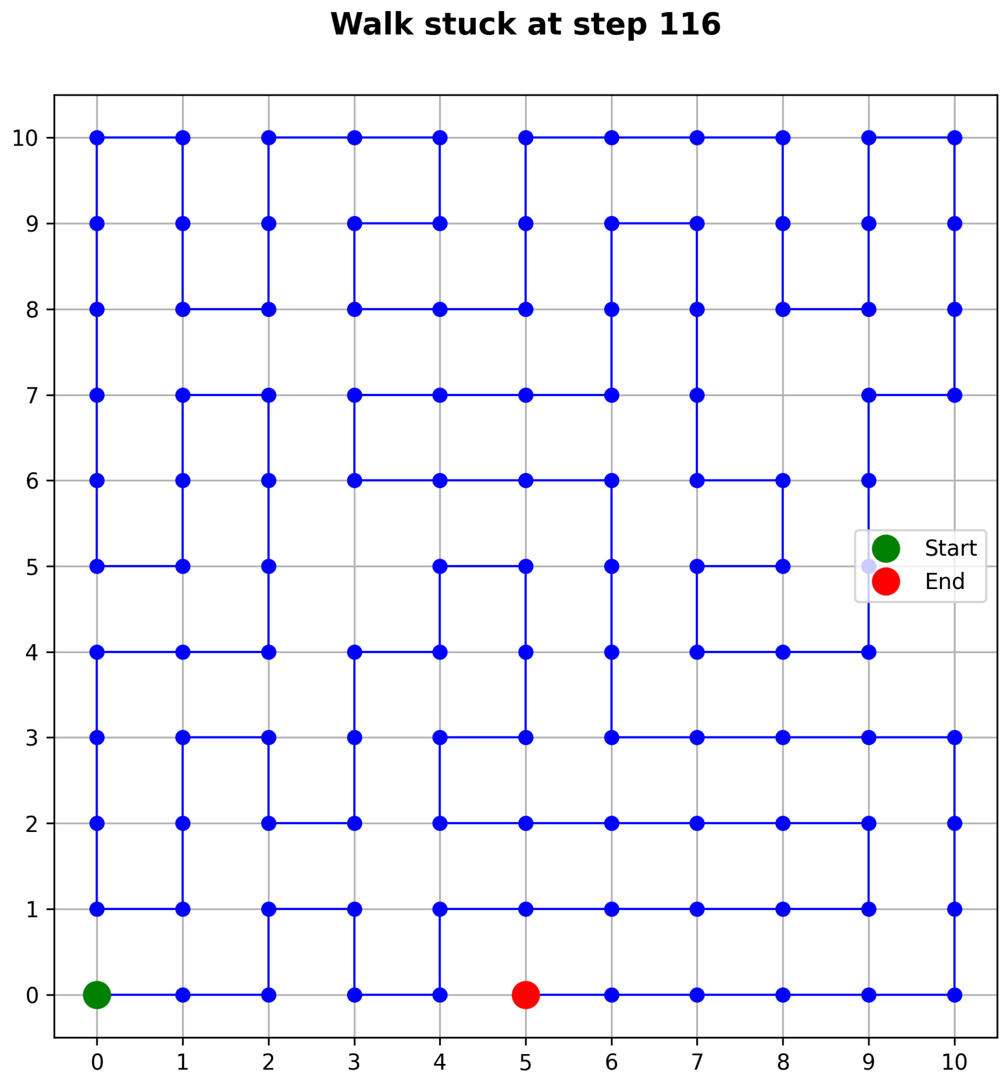
c)For each experiment in part a), I plotted the distribution of lengths N of the SAWs in a histogram. We need to use the weight of the SAWs in calculating the histogram. From the plot below we can see that the distribution of Design 1 is approximately Gaussian because Design 1 has no constrain on the length of its walk. The distribution of Design 2 is highly concentrated at roughly a SAW length of 85, as the early termination probability eliminated longer walks. Design 3 is highly concentrated at roughly a SAW length of 112 since Design 3 favored longer walks by generating children.



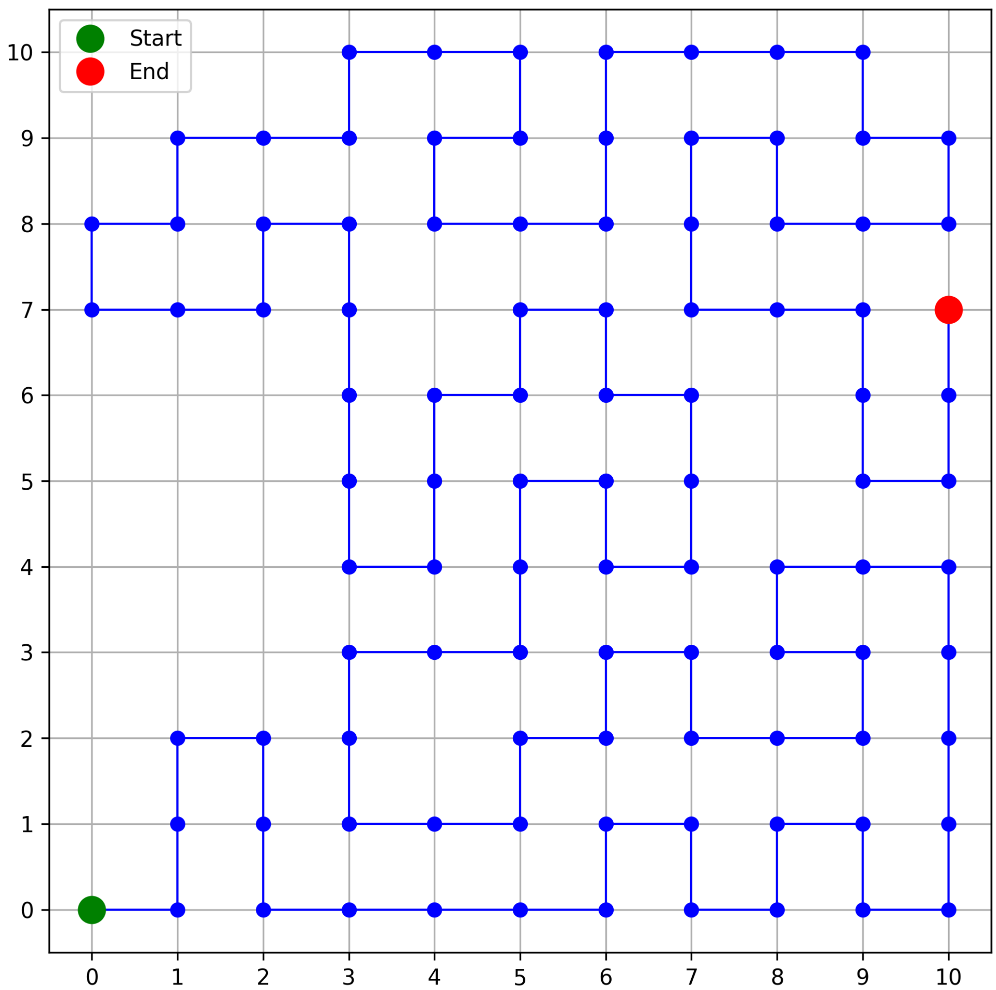

Distribution of SAW Lengths for Design 2
1e32
Count
SAW length
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
0
20
40
60
80
100


Distribution of SAW Lengths for Design3
1e31
Count
SAW length
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
20
40
60
80
100
120

The longest walks for each of the Design 1,2 and 3 from part a) are shown below respectively.



Walk stuck at step 98



Walk stuck at step 118