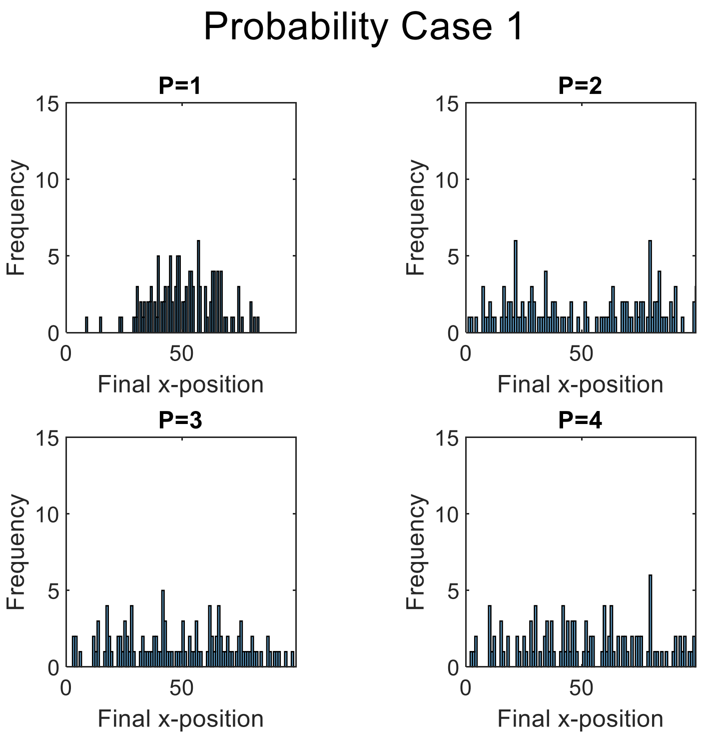
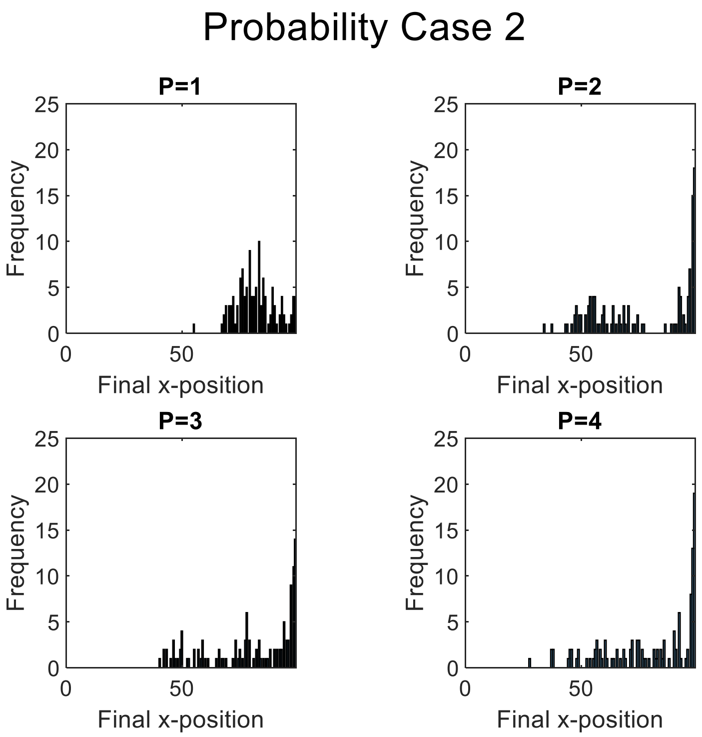
MXB261 ASSIGNMENT 1

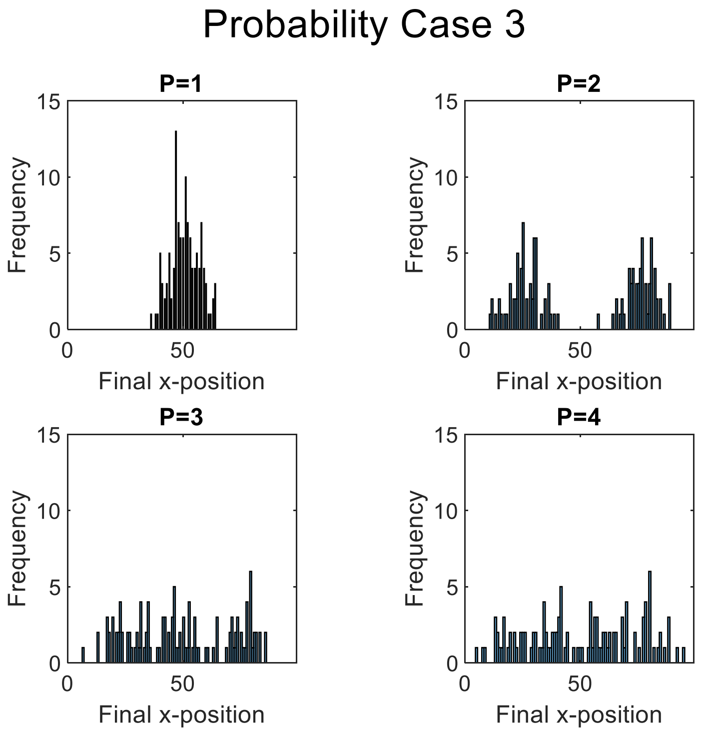
LOK SZE VICTORIA CHEN n10226427

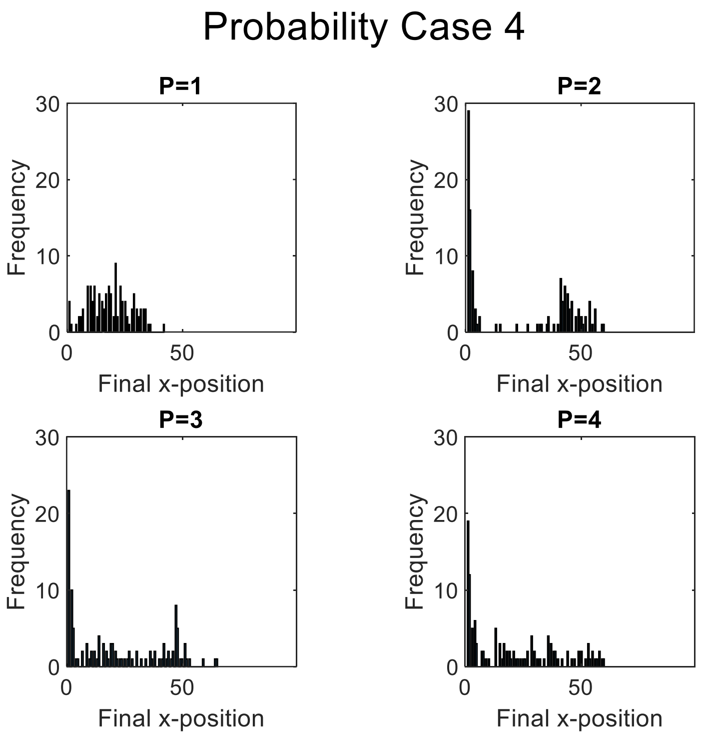
PART 1

Plot figures in MATLAB showing the distribution of heights for 1, 2, 3, and 4 start positions, for each of the probability cases.









Maximum height in each category:

Number of start positions	1	2	3	4
Max Height, case (i)	6	6	5	6
Max Height, case (ii)	10	18	14	19
Max Height, case (iii)	13	7	6	6
Max Height, case (iv)	9	29	23	19

Discussion

It can be observed from histograms that the particles in most walks tend to cluster around the starting x-position. As a result, the number of peaks in each histogram is often the number of starting positions in that simulation.

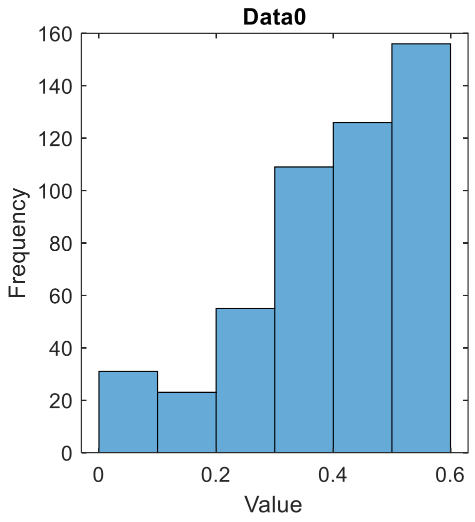
When the probability bias toward the south is higher compared to the east and west, the peaks that form are sharper. Due to the reflective boundary, the particles which would've otherwise fallen off the edges stack along the side of the domain.

For this run, Case (i), which had equal bias between south, west and east, has the lowest frequencies out of all the cases. Case (iv), which had a high bias toward south, has highest max height.

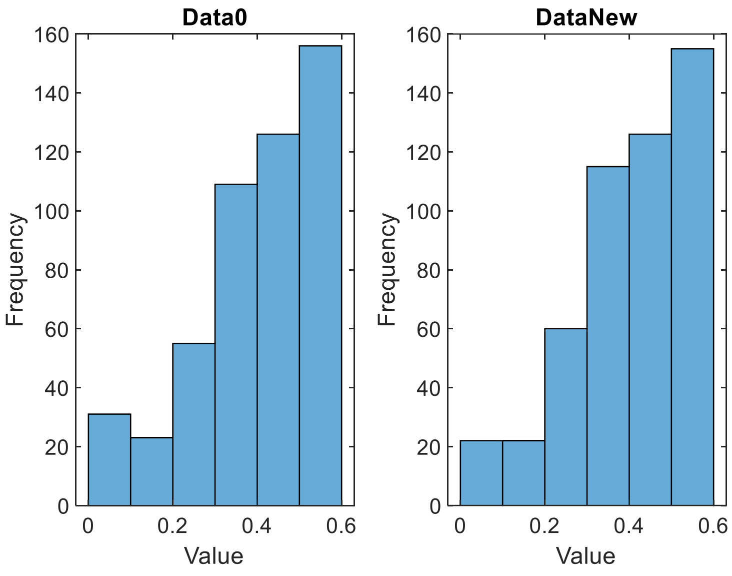
PART 2

SAMPLING FROM EXPERIMENTAL DATA

Data0 plotted with `histogram` on MATLAB, using 6 bins:



The probability distribution of Data0 and DataNew are plotted as follows:



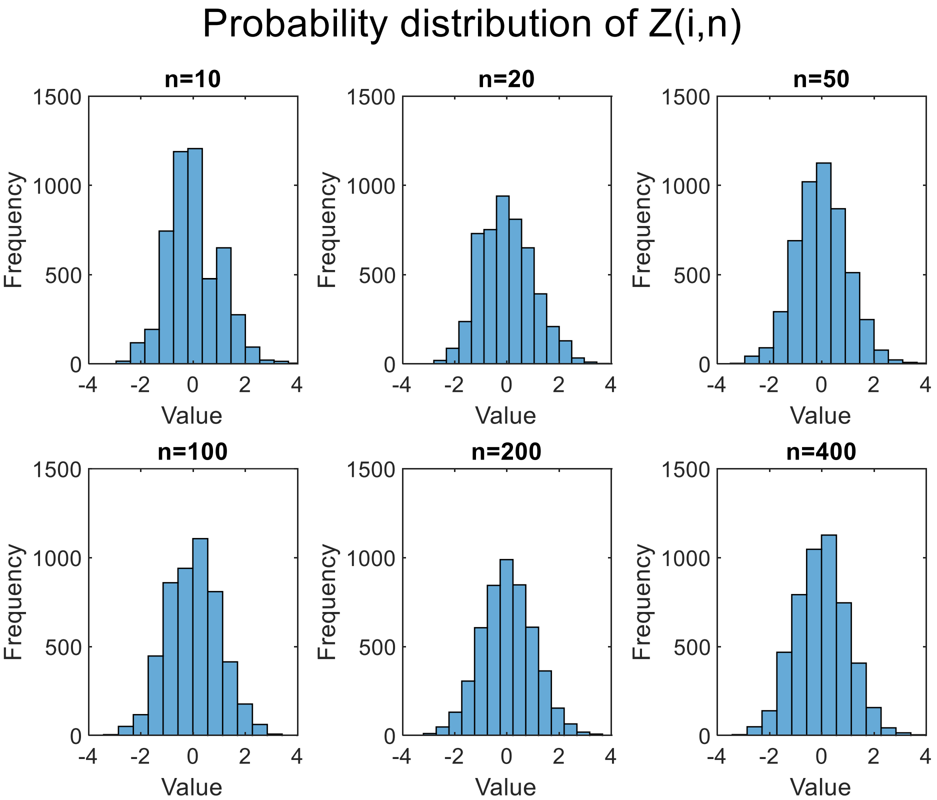
The KL measure between the two probability distributions, calculated with `computeKLD.m`, is 0.0041.

Discussion

500 new random variables were generated successfully from the distribution of Data0. As indicated visually by the histograms, probability distribution of Data0 is very close to the probability distribution of DataNew. The KL measure, which is very close to 0 at 0.0041, further supports this. To achieve an even closer probability distribution and decrease the KL measure, the number of new random variables generated should be increased.

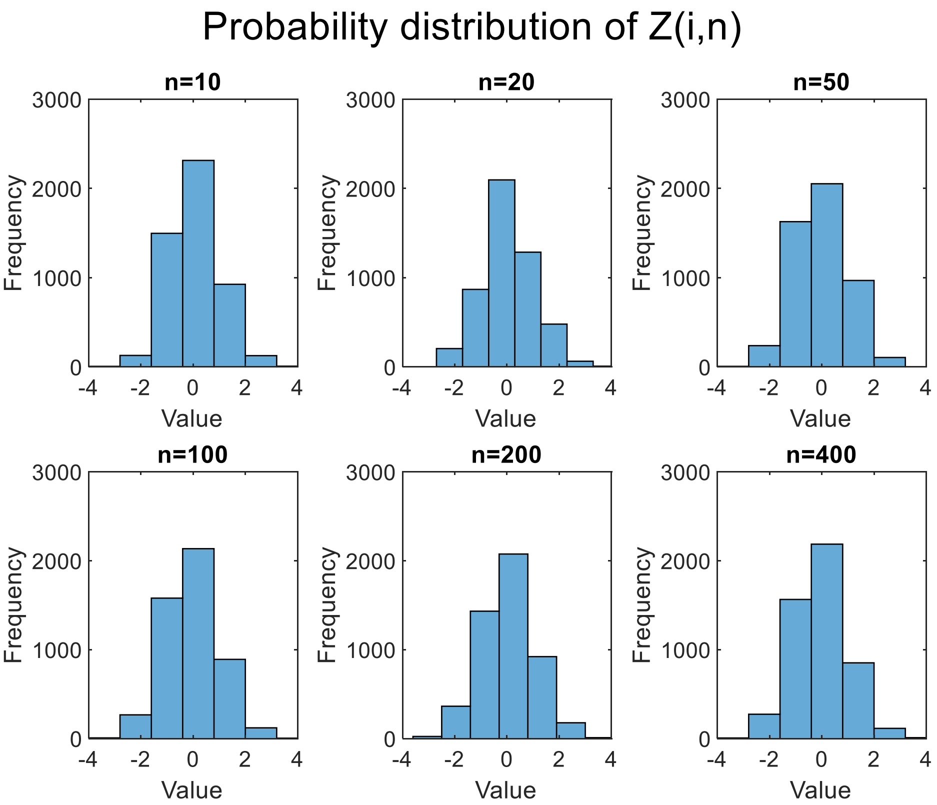
SAMPLING RANDOM NUMBERS FROM POISSON DISTRIBUTION

For each n, a histogram of $Z_{i,n}$ was plotted, using 14 bins:

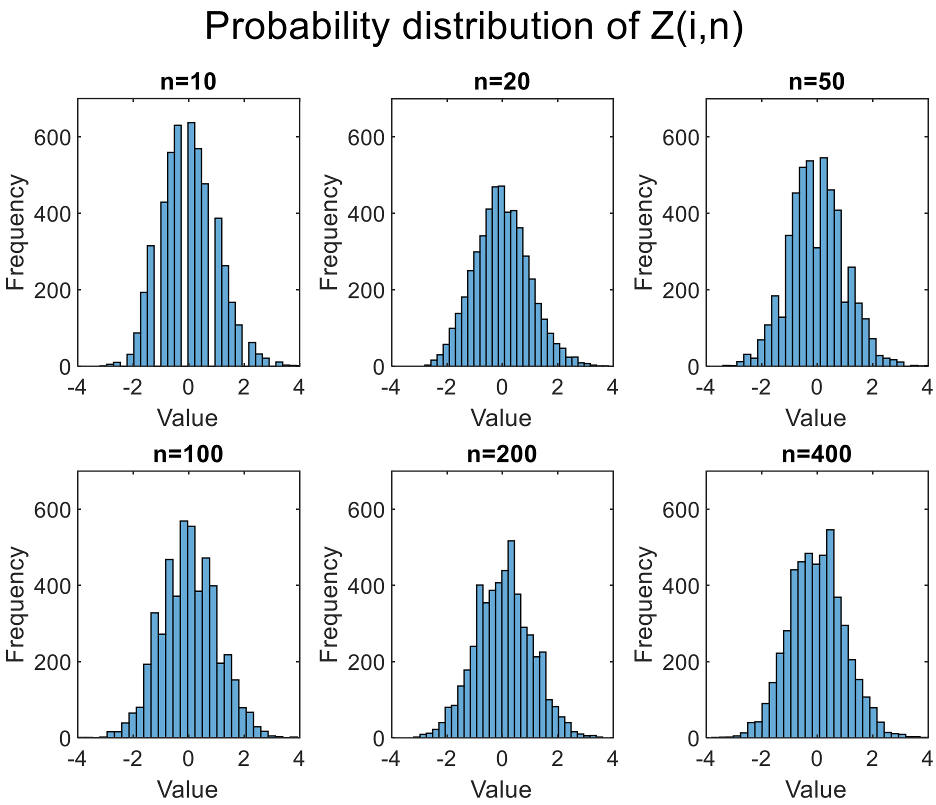


Why use 14 bins?

Picking an appropriate number of bins is important when interpreting data. A bin count that is too low can result in bins which are too wide, obfuscating important details about the distribution. As evidenced in the figure below, using 7 bins makes the normal distribution become less apparent.



Conversely, bins which are too narrow creates unnecessary noise. The histograms below are plotted with 30 bins. There are spikes and drops in the distribution just by coincidence because the points are so precise.



What is the limiting distribution P_∞ of the sequences of Poisson samples, as *n* increases?

The limiting distribution P_∞ of the sequences of Poisson samples, as n increases, is the normal distribution.

Complete the table, to show how the KL measure varies according to the parameter *n*.

The KL measure between the samples and the standard normal distribution is calculated with `poisson_distribution.m`.

n	10	20	50	100	200	400
KL measure from normal distribution	0.0886	0.0985	0.0400	0.0813	0.0244	0.0185

Discussion

It is evident that as n increases, the KL measure between the normalised Poisson(n) samples and the normal distribution approaches 0. The trend will be more apparent if more values of n and/or higher values of n are tested.