MXB261 Problem Solving Task

Kassia Lembryk-Walsh (n11090677)

PART 1 INDEX

Discussion....................................pages 1-2
Figures 1-2....................................page 1
Figures 3-4....................................page 2
MXB261_PST_Part_1.m.................pages 4-6
accretion.m..................................pages 7-9

PART 2 INDEX

Discussion..................................page 3
Figures 5-7..................................page 3
MXB_261_PST_Part_2.m..............page 10
sample_from_dist.m...................pages 11-12

Part 1: A biased random walk

Case 1: Equal probabilities $(south = {}^{1}/_{3}, west = {}^{1}/_{3}, east = {}^{1}/_{3})$

Since there is no northward movement, particles in Case 1 have a greater opportunity ($p = {}^{2}/_{3}$) to move along the x-axis (greater variance). However, with an equal weighting to move each direction, the further from the starting position, the less likely it is for a particle to travel there.

Case 2: Faster fall, equal x-wise probabilities $(south = {}^{2}/_{3}, west = {}^{1}/_{6}, east = {}^{1}/_{6})$

Particles have a much higher probability of falling south than moving x-wise, meaning that particles on average will tend to land much closer to their starting position than Case 1 (lesser variance). X-wise movement is still equal, although their distance from starting position would tend to be smaller.

Case 3: Weighted westward $(south = {}^{3}/_{5}, west = {}^{3}/_{10}, east = {}^{1}/_{10})$

The probability of moving west is 3x greater than that of moving east, and therefore particles will drift westward as they fall (westward expected value).

Case 4: Weighted eastward $(south = {}^{3}/_{5}, west = {}^{1}/_{10}, east = {}^{3}/_{10})$

The probability of moving east is 3x greater than that of moving west, and therefore particles will drift eastward as they fall (eastward expected value).

Figure 1: Cases for P = 1 and N = 100

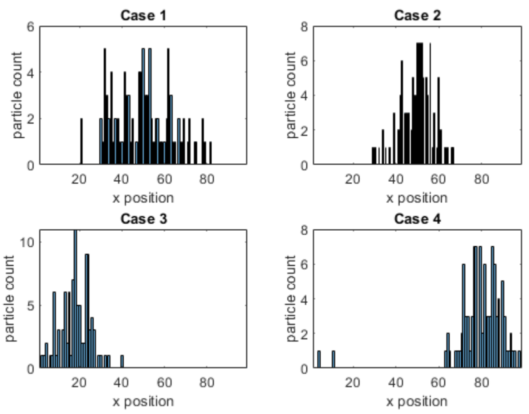
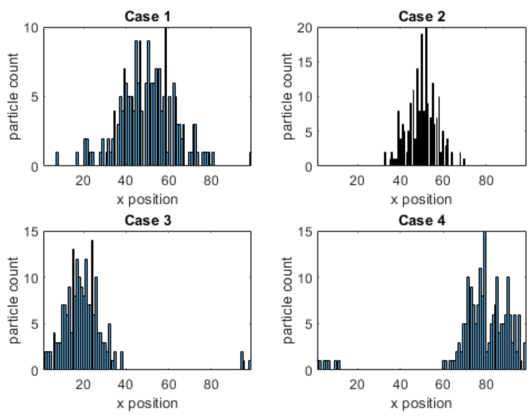


Figure 2: Cases for P = 1 and N = 200



For a **fixed starting position** of $x = 50$; Figure 1 (*above, left*) and Figure 2 (*above, right*) show normal-like distributions for all their cases. Note that the boundaries on the left and right are looping, the seemingly discontinuous heights in Cases 3 & 4 are in fact an expected part of their distribution. The least normal-like curve is Figure 1 Case 1; however, with a smaller number of histogram bins, it becomes visibly much closer to a normal distribution (a smaller sample size is more significantly affected by outliers, but the average is still normally distributed). Since the difference between them is only in the number of particles, the distributions in Figure 1 have a smaller column height and generally more outliers; but maintain the same underlying mean and variance.

- **Case 1** approximates a normal curve with $\mu = 50$ (most likely to land at starting x) and some standard deviation $\sigma = \sigma_1$ that would correspond to its south-ward probability.
- **Case 2** approximates a normal curve with $\mu = 50$ (still equal x-wise probabilities), and with a far greater chance of falling south instead of x-wise, particles will on average travel shorter distances, resulting in a smaller standard deviation, $\sigma_2 < \sigma_1$.
- **Case 3** approximates a normal curve, however with the westward weighting, the expected value is different. On average, 10 total moves will place a particle 3 units west. Expanding this over the total domain from $y = 99$ to $y = 1$, particles on average move 29.4 units west, putting the expected value at $\mu = 20.6$. This is reflected in the figures. Since the variance is directly associated with the chance of downwards movement, we can see that since $south_1 < south_3 < south_2$, therefore $\sigma_1 < \sigma_3 < \sigma_2$. From discussions with classmates, we determined this probably isn't the exact expected value since the particles can collide with each other, but with a maximum column height of 20, this is unlikely to be significant.
- **Case 4** has the same probabilities as Case 3, except weighted eastwards. As such, we can expect the average particle to move 29.4 units east from the starting position and the variance to remain equal, $\mu = 79.4$ and $\sigma_3 = \sigma_4$, which is reflected in the figures.

Figure 3: Cases for P = 'rand' and N = 100

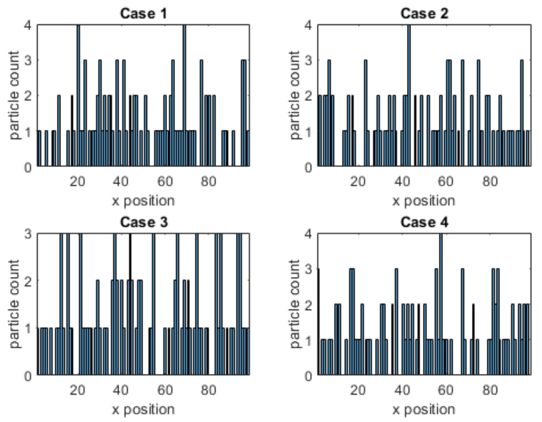
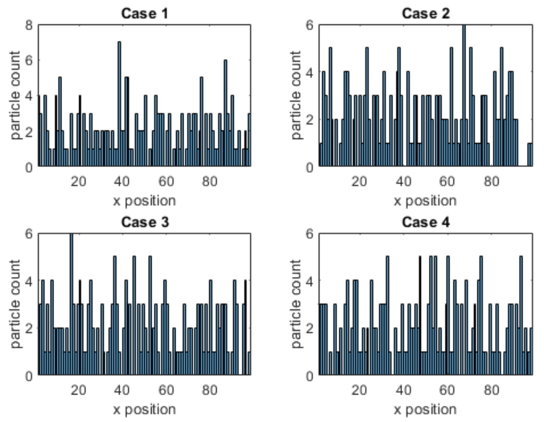


Figure 4: Cases for P = 'rand' and N = 200



Comparatively, for a **random starting position**, the figures are much less interesting. Figure 3 (*beneath paragraph, left*) and Figure 4 (*beneath paragraph, right*) appear as uniform distributions. This is expected. Although the final position of an *individual* particle might have a normal probability distribution, since the initial starting position is determined by a uniform probability distribution between 1 and 99, the chance of landing in any given end position is equally likely. Figures 3 and 4 again differ only in number of particles, reflected in higher column heights.

Part 2: Sampling from Experimental Data

As would be expected, samples generated from the probability distribution function of the given experimental data match closely to the experimental data. Since the samples were generated from a linear interpolation of the data, there might be slight inaccuracies – however, the element of randomness obscures this. As expected, the more bins in the histogram, the more Data0 and DataNew stray from similarity. This correlates well, the underlying distribution is theoretically the same, but with finer detail, randomness emphasises the outliers. This is most clearly communicated in the Kullback-Leibler measure; for 10 bins (Figure 5, *below*) the two distributions are very similar ($DKL_{Data0} = 0.0081$, and $DKL_{DataNew} = 0.0077$), becoming increasingly different as the number of bins increases. For 40 bins (Figure 7, *bottom of page*), DKL_{Data0} is 2.83x greater, and $DKL_{DataNew}$ is 2.73x greater.

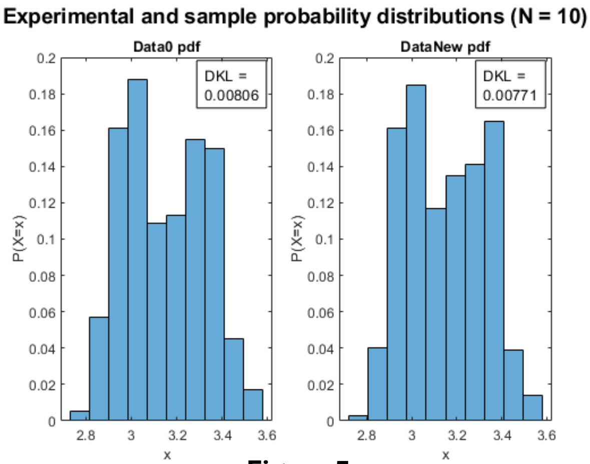


Figure 5

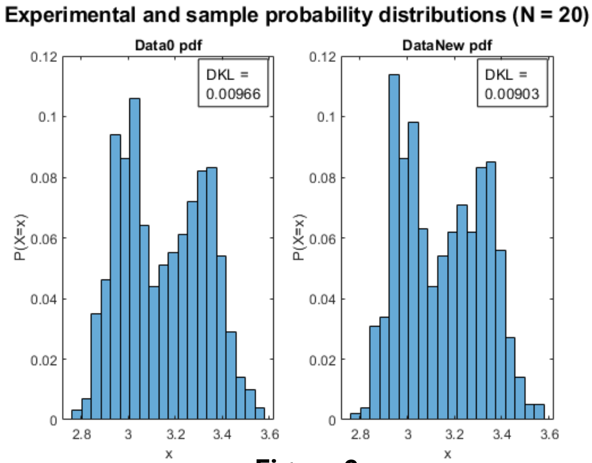


Figure 6

Interestingly, $DKL_{DataNew} < DKL_{Data0}$. I tested a few other seeds and found that this does not *always* hold true, but on average seems to be true. However, this might be confirmation bias. It would be best to run the simulation for a large range of random seeds to determine if this holds true.

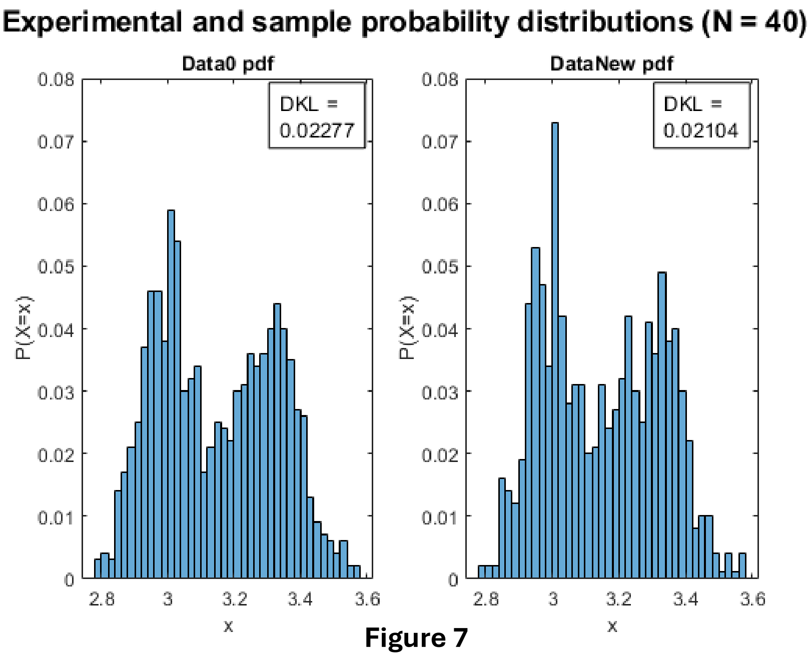


Figure 7