DIFFUSION AND RANDOM WALKS

ANNABEL M. EDWARDS AND LEWIS D. LUDWIG

ABSTRACT. The concept of diffusion is introduced and its importance discussed. The basic principles of diffusion are then discussed from the perspective of a random walk. Students discover that the a random walk distribution can be effectively described using the root mean-square distance and that the root mean-square distance increases linearly with the square-root of time. The classic equation for one dimensional diffusion $x_{rms} = (2Dt)^{1/2}$ is thus introduced and its consequence discussed. The Gaussian probability distribution is also developed.

1. OVERVIEW

This module is designed for students in an introductory chemistry course. The concept of diffusion is introduced and its importance discussed. The basic principles of diffusion are then discussed from the perspective of a random walk. Fick's first and second laws are not invoked. In the course of the module, the student develops his or her own Microsoft Excel spreadsheet using Excel's random number generator. The first goal of this spreadsheet exercise is to generate a histogram of the particle distances after a certain number of random walk steps left or right on a line. Through the self-designed spreadsheet, the student learns the importance of sample size in characterizing the distribution. The student is then led to discover that the resulting distribution can be effectively described using the root mean-square distance and that the root mean-square distance increases linearly with the square-root of time. The classic equation for one dimensional diffusion $x_{rms} = (2Dt)^{1/2}$ is thus introduced and its consequence discussed. For students with a background in calculus, the next section derives the Gaussian probability distribution using the mathematics behind a random walk. After plotting this continuous distribution, the students see that this result is the same as the finite distributions that they generated in their spreadsheet. Additionally, the formal derivation of the distribution leads to the same fundamental relationship between diffusion distance and time, $x_{rms} = (2Dt)^{1/2}$. In the follow up exercises, the students are asked to use their spreadsheet to investigate the importance of boundaries and not so random walks.

In this module, computational science is introduced through modeling a complex system with a discrete number of particles and the use of simple numerical methods to learn the basic behavior of a system. Students are led through the development of their own model and introduced to some basic programming commands in Excel

1

including the random number generator and the count function. In the concluding exercises, students are asked to tweak their developed model and interpret the results.

2. Introduction: Diffusion

Suppose you placed a sugar cube in the bottom of a glass of iced tea and just waited (and waited and waited) for the sweetness to evenly distribute over the entire glass of tea. How long would you need to wait? It turns out, long enough that you would certainly start looking for a spoon or straw to stir your tea. By the end of this module, you will discover how long you would need to wait if you relied only on the random motion of the molecules to evenly distribute the sugar. The motion of molecules due only to their random motion is called *diffusion.*

We must keep in mind, that the water, tea, and sugar molecules in the glass are in constant motion due to their thermal energy, but each step for each molecule is both very small and in a random direction. From our viewpoint, since we cannot see the motion of individual molecules, the sugar just travels slowly away from the sugar cube and extends out a progressively larger distance. However, this net or total motion away from the sugar cube is a series of very small and random steps. These random motions will lead to an even distribution of the sugar molecules throughout the glass so that every drop of the drink has the same number of water molecules, tea molecules, and sugar molecules. When this occurs, if the temperature throughout the glass is exactly the same, our tea will be in equilibrium or at a balance point. The molecules are still continually moving, but there will always be the same concentration of water, tea, and sugar at every location in the glass and all molecules will be at the same temperature.

Let us consider a few other examples. How long does it take for cooking aromas to spread out from the kitchen and fill the house? How long does it take a drop of dye to evenly color the whole glass of water? Our intuition is not incredibly long. The reason is that these molecules travel not only due to their random motion (diffusion), but these molecules are also carried on air or water currents that are caused by small differences in temperature in a room or in the glass of water. The air or water currents result when all of the molecules are trying to evenly distribute a temperature difference so the system can return to equilibrium. This type of movement is often called convective mixing. In the tea, cooking, and dye example, both convective mixing and diffusion cause the molecules to spread out evenly in the system.

Ultimately, it is the conditions of the system that determine how fast molecules spread out. The most important factor is how crowded (on the molecular level) the system is. Suppose a molecule moves in a straight line until it bumps into another molecule. If there are more molecules to bump into, the first molecule (whose movement we are tracking) will travel a shorter distance before it runs into a different molecule and changes direction (see Figure 1). A molecule will travel faster in a gas than in a liquid, because a gas is mostly empty space and there are not as many molecules for our traveling molecule to run into. Molecules in

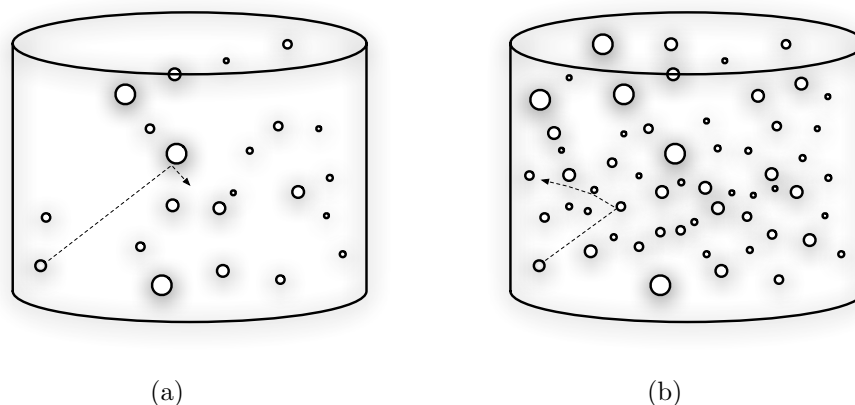(a)   (b)

FIGURE 1. A molecule in beaker (a) is more likely to travel farther
before a collision than in beaker (b).

more crowded environments take more time to travel a given distance because both
diffusion and convective mixing are slower in these environments.

In our everyday real world examples, it is hard to turn off convective mixing and
study the effect of just diffusion, because it is very hard to design a system that has
no small change in temperature. For the interested reader, we refer you to an exper-
imented created by Davis [3]. In that experiment there are a set of conditions that
do minimize convective mixing. Instead of looking at how a solute molecule travels
through water, the experiment explores how a solute molecule travels through a
gel. A solute molecule is the molecule in a mixture that is present in much smaller
amounts. The sugar in our tea example is the solute molecule. Water molecules
in liquid water can move easily around each other and they will move to try to
equalize a temperature difference. In this experiment, a gel is made by adding a
large molecule (a polymer molecule which is thousands of times larger than a water
molecule) to water. The polymer acts like a net that traps the water molecules pre-
venting the water molecules from moving about easily. Hence, the water molecules
cannot move all over to equalize a small change in temperature. This effectively
turns off water currents, so you can focus on just the random motion of the solute
molecule.

We will see that in a system that we can easily visualize , such as Davis' experiment
or the sugar and tea example, pure diffusion takes a very long time. This is because
we can only realize a change when our solute molecule has traveled a distance we
can discern with our eyes. In the cells in our body, the distance a solute molecule
has to travel is very small. Therefore diffusion is very important in biochemical
systems in living organisms. But remember diffusion relies only on random motion
that spreads out molecules from an area of high concentration (sugar in tea near
the sugar cube) to an area of low concentration (the rest of the tea in the glass).
Some cellular processes need to effectively move sugar from the rest of the tea in
the glass back to the sugar cube or move sugar outside of the water in the glass. In
this case, nature has designed proteins that help the molecules travel when diffusion

cannot get the job done. There are plenty of cases where diffusion is important and we will see one example of this in the final exercise.

3. Understanding the math

Before we can model the diffusion process and determine how long it will take for the sugar cube to dissolve in our glass of tea, we need to discuss some of the mathematics required to develop the model. The main mathematical notion employed is the *random walk*.

A random walk is a mathematical way to describe a random process of a sequence of discrete steps of fixed length. So, the path traced by a molecule as it travels in a liquid or gas, specifically our sugar molecule in tea, is described mathematically by a random walk. Molecular motion is not the only phenomenon that can be modeled by a random walk. The rise and fall of stock prices, the movement of a grazing cow, and a gambler's financial standing can all be model by random walks.

Let's consider a simple example. Imagine an inebriated walker leaning against a street lamp that is in the middle of a street running east to west (see Figure 2). He decides to start walking, and is as likely to take a step east (right) as west (left). So he has a 50% chance of stepping to the right and a 50% chance of stepping to the left. Each step is of the same length and once he reaches a new distance, he is still equally likely to step to the right or left to the next distance.
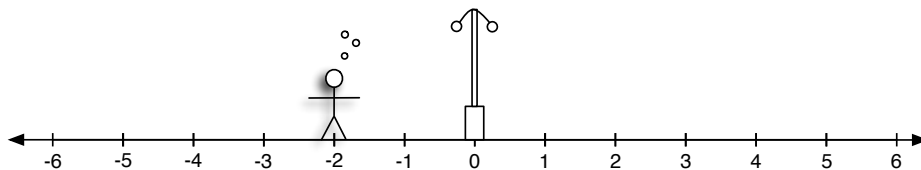

Figure 2. An inebriated person demonstrating a random walk.

We can actually illustrate the drunk's path in the following way. Start with the drunk at the lamp post. Flip a fair coin. If the coin lands on heads, he steps to the right one unit. If it lands on tails, he steps to the left one unit. Suppose we flip the coin four times and get H-T-T-T. This means we will land on -2.

Question 1. Suppose he starts at the lamp post and we flip the coin four times. Create a tree that lists all the possible outcomes of head tail combinations and the distances where he will land for each.

Question 2. Now suppose he starts at the lamp post and we flip the coin *five* times. Create a tree that lists all the possible outcomes of head tail combinations and the distances where he will land for each.

Question 3. From the above observations, what can you say about an even number of flips as opposed to an odd number of flips? That is, can you determine anything about his landing distance depending on whether he took an even number steps or an odd number of steps?

Now that we have some intuition about random walks, it should be no surprise that there is nothing special about our example with the drunk. Since he was only moving left or right, the drunken walk is an example of a one-dimensional random walk. The random motion of molecules in diffusion can also be modeled using a random walk. However, molecules move not only left or right, but up or down and forward and backward, so we should use a three-dimensional random walk to describe the motion of molecules. You can imagine keeping track of all of the possible motions of one molecule in three dimensions would get tedious with just a pencil and paper. If we want to model the behavior of a chemical system or the sugar spreading out in the tea glass, we need to follow the steps of more than just one molecule. Keeping track of possible distances of many molecules becomes nigh on impossible, without the counting system of a computer program. Computational chemists use computer programs to model diffusion, chemical reactions, protein folding, and other systems [1],[6].

Back to our dissolving sugar example. Let's get a handle on how big a system we are considering.

Question 4. How many molecules are in a sugar cube? Write down a guess. To find the actual number, you need some additional information. Assume the sugar cube has a mass of roughly 4.00 g. and one sugar molecule has an approximate mass of 5.69×10^{-22} grams.

Our goal is to build a simple model using Excel that helps us understand how diffusion, or molecules' random steps, leads to molecules spread through out a system. Due to the huge number of molecules, we are not going to use all of the molecules in our sugar cube in our simple model. But how many molecules do we need to get an accurate description of diffusion behavior? More than one, but less than the total amount. Will ten molecules be enough? Moreover, how many random steps of the molecules do we need to observe before we can establish a pattern for diffusion?

To simplify our model into something we can visualize, we will use a one-dimensional random walk. We will look at how a group of molecules disperses if their movement is restricted to the left or right on a line. The results we get from our one-dimensional model will allow us to develop some general relations for how long it takes molecules to travel via diffusion. These general results can be extended to understand the behavior of three dimensional systems. We will also restrict a molecules motion so that each step is the same distance, a length of one unit.

Question 5. What are the assumptions in our model? What parameters of our model do we need to determine as we work with the model?

4. Creating the model in Excel

First we need to find a way for Excel to flip a fair coin. The program has a built in function, RANDBETWEEN, that returns a random integer between the numbers you specify, including those numbers. In a new worksheet, type =RANDBETWEEN(1,2) in cell A4. The result should be a 1 or a 2. Each time anything is changed on the worksheet, Excel recalculates the function and chooses

at random 1 or 2. Give your worksheet a title in cell A1. Did the value in A4 change?

Exercise 1. To see how RANDBETWEEN works, create 50 random numbers using the RANDBETWEEN function. You may want to create a small tally sheet to keep track of your results (see Figure 3). Note, there is an easier way to have Excel recalculate. On a PC, simply press "F9" or on a Mac hold down "Command" followed by "=." Keep in mind, that a random number generator should have strings of just ones or twos, so don't be confused if it doesn't look like the values in A4 are not changing.

	A	B	C	D
1	Excel's random number generator			
2				
3	Random Number		Talling the number of	
4	1		times Excel chooses each	
5			value in A4	
6		Number	1	2
7		1	x	
8		2		x
9		3	x	
10		4	x	
11		5	x	
12		6	x	
13		7		x
14		8		x
15		9		x
16		10		x
17		11	x	
18		12	x	

FIGURE 3. An image of the random number generator function in Excel, with tally sheet.

How many ones and twos do you have? How does your distribution of ones and twos compare to your neighbor? Do you think the number generator is really random?

So now we have a way for the spreadsheet to randomly choose if our molecule will go to the right of left. Let's start with a system of five molecules and figure out how to have the spreadsheet keep track of the distance of each molecule after each molecule steps left or right.

Figure 4 gives an example of a way we can do this. In Column C, we use the spreadsheet's random number generator =RANDBETWEEN(1,2)." Therefore the value in column C will either be a 1 or a 2 and these values will change each time the worksheet is updated ("F9" or "Command" + "="). The value in column D depends on the value in column C. If column C equals 1 then column D equals -1 or the step we want our molecule to take is -1. Similarly if column C equals 2 then column D equals 1 or the step we want our molecule to take is 1. We need to use the spreadsheet's POWER function to do this. This function has the form POWER(number, power) and returns the result of a number raised to a power. So

column D either has $(-1)^1 = -1$ or $(-1)^2 = 1$. The New Position or column E is simply the start distance plus the step. So for cell E11 we enter "=B11+D11."

	A	B	C	D	E	F	G
1	One Dimensional Random Walk						
2	5 particles can step to the right or to the left						
3	The random numbers change with each update to the worksheet						
4	All particles start at zero						
5	Excel filps a coin and chooses "1" or "2" using the "=RANDBETWEEN(1,2)" funciton						
6	If random=1, the molecule's step is -1 ([-1]^1="Power(-1,D11)"						
7	If random=2, the molecule's step is 1 ([-1]^2="Power(-1,D11)")						
8	The new position is the sum of the start position and the step (using "=B11+D11")						
9							
10	Particle	Start position	Random	Step	New Position		
11	1	0	1	-1	-1		
12	2	0	2	1	1		
13	3	0	2	1	1		
14	4	0	2	1	1		
15	5	0	1	-1	-1		
16							

FIGURE 4. An image of the random number generator with five particles.

By copying the data and formulas from columns C through E, we can find the position of five particles after five steps (see Figure 5). Remember these steps are random, so each time the worksheet updates the position of each molecule will be different.

Particle	Start position	Random	Step	Position 1	Random	Step	Position 2	Random	Step	Position 2	Random	Step	Position 3	Random	Step	Position 4	Random	Step	Position 5
1	0	2	1	1	2	1	2	2	1	3	1	-1	2	1	-1	1	2	1	2
2	0	2	1	1	2	1	2	2	1	3	1	-1	2	1	-1	1	1	-1	0
3	0	2	1	1	1	-1	0	1	-1	-1	2	1	0	2	1	1	1	-1	0
4	0	1	-1	-1	1	-1	-2	1	-1	-3	1	-1	-4	1	-1	-5	1	-1	-6
5	0	1	-1	-1	2	1	0	1	-1	-1	2	1	0	1	-1	-1	1	-1	-2

FIGURE 5. An image of the random number generator with five particles.

Exercise 2. Use Excel to track the position of five particles after five steps. Your worksheet should look like Figure 5. Make sure to use the spreadsheet's copy and paste function to minimize typing.

This is a good start to modeling the one-dimensional movement of our molecules. However, we would like to keep track of more molecules so we can visualize the number of particles in each position.

We will now go through the steps to have the spreadsheet count the number of particles at distance 1 and distance -1 after one step, resulting in Figure 6. (Note we have increased the number of particles to 10.) In order to count the number of particles at distance 1 and the number at distance -1, we use the COUNTIF(range, criteria) function, which counts the number of cells within a range that meets a given criteria. So in cell C23 we enter "=COUNTIF(E11:E20,-1)." This will count the number of occurrences of the value -1 in the range of cells from E11 to E20. Similarly, we enter "=COUNTIF(E11:E20,1) in cell C24.

	A	B	C	D	E	F	G	H	I
1	One Dimensional Random Walk								
2	5 particles can step to the right or to the left								
3	The random numbers change with each update to the worksheet								
4	All particles start at zero								
5	Excel filps a coin and chooses "1" or "2" using the "=RANDBETWEEN(1,2)" function								
6	If random=1, the molecule's step is -1 ([-1]^1="Power(-1,D11)"								
7	If random=2, the molecule's step is 1 ([-1]^2="Power(-1,D11)")								
8	The new position is the sum of the start position and the step (using "=B11+D11")								
9									
10	Particle	Start position	Random	Step	Position 1				
11	1	0	1	-1	-1				
12	2	0	2	1	1				
13	3	0	1	-1	-1				
14	4	0	1	-1	-1				
15	5	0	1	-1	-1				
16	6	0	1	-1	-1				
17	7	0	2	1	1				
18	8	0	1	-1	-1				
19	9	0	2	1	1				
20	10	0	2	1	1				
21									
22		Position	Count						
23		-1	4						
24		1	5						

FIGURE 6. Counting number of particles at distance (1) and (-1) after 10 steps.

Exercise 3. Use Excel to create a worksheet that for a system of 10 particles counts the number of particles at distance (1) and (-1) after one step using the =COUNTIF function. Your worksheet should look like Figure 6.

Now we can ask Excel to create a column graph. Select cells G22 and G23. Insert a "2-D Column Graph and then go to the Window Select Data Source. On the right hand side of this box, select Edit under the Horizontal (Category) Axis Labels and highlight cells F22 and F23 (see Figure 7). Our labeled histogram is in Figure 8.

Exercise 4. Create a histogram that plots the number of particles at distance (1) and (-1) from your answers to Exercise 3. Your answer should be similar to Figure 8.

You now have the tools to create histograms for many molecules undergoing many steps.
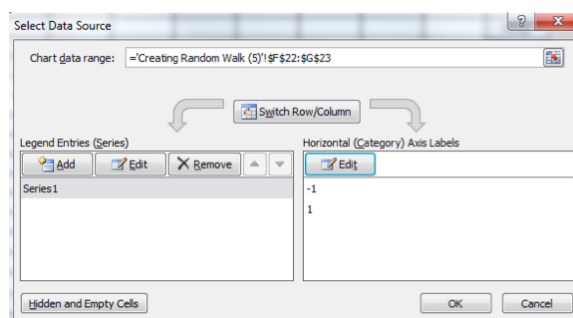
FIGURE 7. The Data Source window in Excel for a PC.

Exercise 5. What are the possible distances of a molecule after it has taken six steps? For ten molecules, create a histogram after six steps. You will need to use the COUNTIF function and ask Excel to find the number of molecules at each distance you determined above.

Update your Excel sheet so the random numbers change. Your histogram should change. If it does not, then check with your instructor to see where the error is. Do you see a pattern or does it just look random?
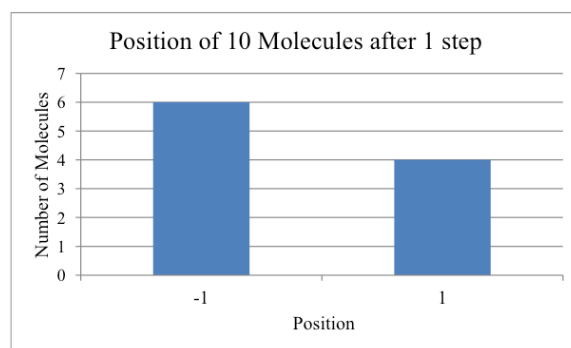

FIGURE 8. A histogram.

Exercise 6. Next create a duplicate copy of your worksheet and change this worksheet so you now create a histogram after six steps for 200 molecules. Make sure you are using Excel's copy and paste function.

Update your Excel sheet so the random numbers change. Again, your histogram should change. Do you see a pattern or does it just look haphazard?

You should start to see a general shape emerging that does not drastically change after each calculation. What is this shape?

Repeat the above but now create a histogram of 200 molecules after 10 steps.

Compare the shape of each histogram and value at distance zero for both your six step and ten step histogram. What is happening to the molecules? Where is the most likely place of finding the molecules?

We now have a visual representation of how the molecules spread out. If we increased the number of molecules to a thousand or a million we would see that our random steps lead to a very regular pattern or regular distribution.

While it is interesting to see the distribution that results from random motion of the molecules, our question was how long would it take for the sugar to spread throughout the glass. We need an equation that connects how far the sugar molecules travel to time. To find this equation we need to accomplish three tasks. First, we need to summarize all of the information in the distribution with one value of the distance. Second, we need to determine the functional dependence of this distance on the number of steps. Third, we need to find a way to link time to this equation and to determine the actual length of one step.

We will start with our first task. Why would we want to summarize all of the information in the distribution with one distance value? Our histograms contain a complete picture of how the molecules spread out, but we need to generate the whole plot to describe what the distribution looks like. Not only is generating the whole plot tedious, it is not useful if we want to easily describe how the distribution changes in time. Our next step will be to find a way to characterize all of the information in the distribution with one number. This number that characterizes the distribution will be the root mean-square distance traveled by all of sugar molecules. As we discuss how to characterize the distribution, you will understand why we use the root mean-square distance and not just a simple average distance.

We will start by using Excel to find the average distance traveled by each sugar molecules after each step. First we multiply the number of molecules at a given spot by the distance of that spot. For example, if our distribution has four molecules at distance 3 and five molecules at distance -2, then we would multiply four by three and five by negative two. Second we sum up all of the values of number times distance so we have $(4 \cdot 3) + (5 \cdot -2) = 2$. We then divide this overall sum by the total number of particles to find the average distance traveled by each molecule $(2/9) = 0.22$. Excel can efficiently help us with this math. Figure 9 shows an Excel sheet for calculating the average distance traveled in the first three steps by each molecule in a system of 200 molecules.

Exercise 7. Revisit the distribution you created for 200 molecules that have each taken ten steps. In Excel, calculate the average distance traveled by all of the molecules after each step for steps zero through ten. Then divide the average for all molecules by 200 to calculate the average distance each molecule traveled after one step, two steps, three steps ten steps. (Set up a table similar to the one in Figure 9.)

Number of molecules found at each position				
Position	0 Steps	1 step	2 steps	3 steps
-3	0	0	0	14
-2	0	0	38	0
-1	0	92	0	73
0	200	0	104	0
1	0	108	0	84
2	0	0	58	0
3	0	0	0	29

Calculating total distance traveled by 200 molecules				
	0 Steps	1 step	2 steps	3 steps
	"=0x0"	"=0x0"	"=0x0"	"=14x-3"
	"=0x0"	"=0x0"	"=38x-2"	"=0x0"
	"=0x0"	"=92x-1"	"=0x0"	"=73x-1"
	"=200x0"	"=0x0"	"=104x0"	"=0x0"
	"=0x0"	"=108*1"	"=0x0"	"=84*1"
	"=0x0"	"=0x0"	"=58x2"	"=0x0"
	"=0x0"	"=0x0"	"=0x0"	"=29x3"
SUM	0	16	40	56
Average distance traveled by each of the 200 molecules				
SUM/200 molecules	0	0.08	0.2	0.28

FIGURE 9. Calculating the average distance traveled in first three steps by each molecule in a system of 200.

Right away you might realize that the simple average is not a good way to characterize our distribution. The average distance traveled does not increase as the number of steps increases and is a number close to zero!

Why does the simple average give this unsatisfactory result and not a number that can help us describe how the distribution changes in time?

Next we will use Excel to find the root mean-square distance. First we multiply the number of molecules at a given spot by the square of the distance of that spot. For example, if our distribution has four molecules at distance 3 and five molecules at distance -2, then we would multiply four by nine (or three squared) and five by four (or negative two squared). Second we sum up all of the values of number times distance so we have $(4 \cdot 3^2) + (5 \cdot (-2)^2) = 56$. We then divide this number by the total number of particles to find the mean-square distance traveled by each molecule ($56/9 = 6.22$). Lastly, we take the square root of this number $\sqrt{6.22} = 2.49$ to obtain the root mean-square distance traveled by each molecule.

Again Excel can help us with this calculation for all of the sugar molecules. Figure 10 (on next page) shows an Excel sheet for calculating the root mean-square distance traveled in the first three steps by each molecule in a system of 200 molecules.

Number of molecules found at each position				
Position	**0 Steps**	**1 step**	**2 steps**	**3 steps**
-3	0	0	0	14
-2	0	0	38	0
-1	0	92	0	73
0	200	0	104	0
1	0	108	0	84
2	0	0	58	0
3	0	0	0	29

Calculating total squared distance traveled by 200 molecules				
	0 Steps	**1 step**	**2 steps**	**3 steps**
	"=0x(0)^2"	"=0x(0)^2"	"=0x(0)^2"	"=14x(-3)^2"
	"=0x(0)^2"	"=0x(0)^2"	"=38x(-2)^2"	"=0x(0)^2"
	"=0x(0)^2"	"=92x(-1)^2"	"=0x(0)^2"	"=73x(-1)^2"
	"=200x(0)^2"	"=0x(0)^2"	"=104x(0)^2"	"=0x(0)^2"
	"=0x(0)^2"	"=108*(1)^2"	"=0x(0)^2"	"=84*(1)^2"
	"=0x(0)^2"	"=0x(0)^2"	"=58x(2)^2"	"=0x(0)^2"
	"=0x(0)^2"	"=0x(0)^2"	"=0x(0)^2"	"=29x(3)^2"
SUM	0	200	384	544
Mean square distance traveled by each of the 200 molecules				
SUM/200 molecules	0	1	1.92	2.72
Root mean square distance traveled by each of the 200 molecules				
SQRT(SUM/200 molecules)	0	1	1.39	1.65

FIGURE 10. Calculating the root mean-square distance traveled in the first three steps by each molecule in a system of 200 molecules.

You will note that the root mean-square distance is very similar to the standard deviation. This makes sense because the standard deviation describes the spread of answers around a mean value. We want to describe the distribution of sugar molecules or the spread of our sugar molecules around the starting point. The formula for the standard deviation, σ, of n measurements is

$$\sigma = \frac{\sqrt{(x_1 - x_{ave})^2 + (x_2 - x_{ave})^2 + (x_3 - x_{ave})^2 + \ldots + (x_n - x_{ave})^2}}{n}.$$

In our example x_{ave} is zero and we have the same formula we used to calculate the root mean-square distance.

Exercise 8. Revisit again the distribution you created for 200 molecules that have each taken ten steps. In Excel, calculate the mean-square distance traveled by all of the molecules after each step for steps zero through ten. Then divide the average for all molecules by 200 to calculate the mean-square distance per molecule. Finally, take the square root of this value to find the root mean-square distance each molecule traveled after one step, two steps, three steps ten steps.(Set up a table similar to the one in Figure 10.)

You should see that the root mean-square distance each molecule traveled increases as the number of steps increases. We now have a succinct way of describing the

whole distribution of molecules. The root mean-square distance gives us the spread of how far the sugar molecules have traveled from their starting distance (zero in our model) and thus succinctly describes the histograms we created earlier.

We have found a way to summarize all of the information in the histogram with one distance value. We will then ask how does the root mean-square distance depend on the number of steps, n?

Exercise 9. Plot the root mean-square distance on the x-axis and the number of steps, n, on the x-axis.

You easily see that that the root mean-square distance does not increase linearly with time so we cannot say that the root mean-square distance is proportional to the number of steps, n. To find this dependence between x_{rms} and n, we will use the helpful properties of logarithms as shown by the following example. If our function is $a = 3b^2$ and we take the natural log of each side we have $\ln a = \ln(3b^2) = \ln 3 + \ln b^2 = \ln 3 + 2\ln b = 2\ln b + \ln 3$. Notice that by taking the natural log of each side and using the properties of logs, we have turned the initial equation, $a = 3b^2$, into an equation for a line, $\ln a = 2\ln b + \ln 3$. Therefore, if we plot $\ln a$ on the y-axis and $\ln b$ on the x-axis, we will have a line with a slope of 2 and a y-intercept equal to $\ln 3$. Now we return to our goal to determine how x_{rms} depends on n. In other words, we know $x_{rms} = (\text{constant})(n)m$ but we do not know the value of the constant or m. At this point, we want to find m and will wait to find the constant in the next section.

Exercise 10. Create a plot of $\ln x_{rms}$ versus $\ln n$. Add a trend line to the plot and include the equation for the trend line on the plot. The slope of this line is the m in our equation $x_{rms} = (\text{constant})(n)m$.

We now have a general equation that connects how far the sugar molecules travel to the number of steps: $x_{rms} = (constant)\sqrt{n}$ where x_{rms} is the root mean square distance and the proportionality *constant* we will determine next.

This final step will link time to this equation and set the units on x_{rms} to a meaningful value. Time enters our model through the number of steps. We have already assumed the each step is the same distance and we extend this assumption by saying that all steps take each molecule the same amount of time. We can write that

$$(1) \qquad\qquad n = (t/\tau)$$

where t=time, n=number of steps, and $\tau = $ the length of time for one step. Now we need to specify what the distance per step is in terms of physical quantities: we need to convert one step unit into an actual distance. For diffusion in liquids, the distance per step equals the diameter of the diffusing molecule. This step distance

is equivalent to stating that the molecule travels a distance equal to itself before it collides with another molecule. To do this, we need to perform a simple unit conversion where x_0 is the diameter of the diffusing molecule in centimeters and thus the step length in centimeters.

(2) $x_{rms}(\text{in cm}) = x_{rms}(\text{in step length}) * x_0(\text{cm/(step length)})$

For diffusion along a line, we can solve for x_{rms} in centimeters as a function of time by combining equation 1 and 2:

$$x_{rms} = x_0 \left(\frac{t}{\tau}\right)^{1/2} = \left(\frac{x_0^2 t}{\tau}\right)^{1/2}.$$

Chemists group x_0 and τ together into a single constant known as the diffusion constant.

$$D = \frac{x_0^2}{2\tau}$$

This equation is known as the Einstein-Smoluchowski equation. The diffusion constant, D, is specific to both the molecule that is diffusing (the sugar) and the system the molecule is diffusing in (the tea). This dual dependence makes sense because x_0 is the diameter of the diffusing molecule and the time per step, τ, depends on how fast the molecule moves through the system which depends on the density of the system and interactions between the diffusing molecule and system. The exact value of the diffusion constant is determined empirically through a wide variety of experiments. Most commonly, the diffusion constant is expressed in units of cm^2/sec. We are now at the point where we have completed the three steps and can write an exact equation for how a molecule will spread out along a straight line.

(3) $x_{rms} = (2Dt)^{1/2}$

The diffusion constant is the same value if we model diffusion in two dimensions (diffusion in a plane) or in three diffusions. However, to account for the added dimensions, the constant "2 in the above equation increases as the dimension increases. For diffusion in a plane or a two- dimensional random walk, $x_{rms} = \sqrt{4Dt}$. For diffusion in three dimensions such as the sugar in the tea glass, $x_{rms} = \sqrt{6Dt}$. Thus if we know a value of D for sugar in tea, then we can determine how long it will take for the sugar to be distributed evenly in the tea.

In the next section, you will see that we can find this same result by looking at the distribution generated by our histograms more formally using calculus.

5. Solving the sugar question

Now that we have an understanding of random walks and how they can model molecular motion in a straight line, let's return to our sugar molecule in the ice tea

question. Specifically, how far will our sugar molecule travel in a straight line in a given period of time (or steps)?

To answer this question, we must first find the probability of a particle being at distance x on our number line given it started at $x = 0$ and is moving along the number line according to a random walk (see Figure 11).
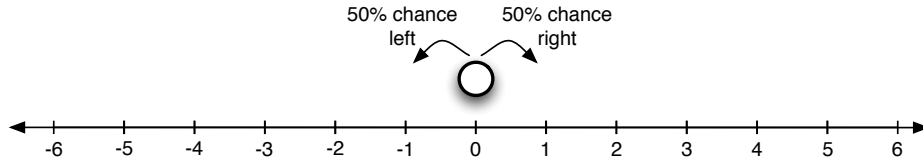

FIGURE 11. Particle starting at $x = 0$ has 50% chance of moving to the left or 50% chance of moving to the right.

Using this probability distribution, we then develop the diffusion equation, which will ultimate help us answer our sugar-tea question.

In order to find the probability distribution of our particle, we first count the number of ways our particle can be at distance x. To do this, we let

$$n \quad = \quad \text{total number of steps}$$
$$r \quad = \quad \text{total number of steps to the right}$$
$$l \quad = \quad \text{total number of steps to the left}$$
$$x \quad = \quad \text{integer distance of particle on number line.}$$

We now use some algebra to relate these expressions. We see that $x = r - l$, so $r = x + l$. Also, we have that $n = r + l$, so $l = n - r$. That is $r = x + n - r$. Solving for, we get $r = \dfrac{n + x}{2}$. Similarly , we can express l in terms of n and x, giving $l = \dfrac{n - x}{2}$. Hence the number of ways a particle can be at a distance x after n total steps is determined by the number of different sequences of rights and lefts from the starting point of $x = 0$ that land on x (recall Questions 1-3). This is the number of permutations of length n of the letters r and l. That is, the number of ways a particle can be at distance x is $l = \dfrac{n!}{r!l!} = \dfrac{n!}{\left(\frac{n+x}{2}\right)! \left(\frac{n-x}{2}\right)!}$. (It should be noted that $n + x$ and $n - x$ are always even as the molecule can only be at distance x if x and n are both even or both odd, and $n > x$.) Moreover, the probability of a specific sequence of r's and l's occurring is $\left(\dfrac{1}{2}\right)^n$. Thus the probability that the particle is at distance x, $P(x)$, is given by

$$(4) \qquad\qquad P(x) = \left(\frac{1}{2}\right)^n \frac{n!}{\left(\frac{n+x}{2}\right)! \left(\frac{n-x}{2}\right)!}.$$

Next, we approximate the above probability expression as a continuous function of x that can be plotted against our random walk data we generated in the worksheet

exploration so we can compare the experimental data against the theoretical approximation. Unfortunately, working with factorials is mathematically difficult, so we will work with the logarithm which can be obtained by a Taylor series expansion of $\ln P(x)$ about $x = 0$. Because we are only looking for an approximation, we consider the first three terms of the expansion. This gives

$$(5) \qquad \ln P(x) \approx (\ln P(x))_{x=0} + \left(\frac{d \ln P(x)}{dx} \right)_{x=0} + \frac{1}{2} \left(\frac{d^2 \ln P(x)}{dx^2} \right)_{x=0}$$

This approximation is reasonable as the odd terms of the full expansion go to zero and the even terms become very small very quickly when evaluated at zero.

The reason we use the logarithm expansion of $P(x)$ is so we can employ Sterling's approximation, $\ln n! \approx n \ln n - n$, to evaluate the derivative in the series expansion. To do this, we take the natural logarithm of Equation 4 and apply Sterling's approximation.

$$(6) \ln P(x) = \ln \left(\left(\frac{1}{2} \right)^n \frac{n!}{\left(\frac{n+x}{2} \right)! \left(\frac{n-x}{2} \right)!} \right)$$

$$(7) \qquad = n \ln \left(\frac{1}{2} \right) + \ln(n!) - \ln \left(\left(\frac{n+x}{2} \right)! \right) - \ln \left(\left(\frac{n-x}{2} \right)! \right)$$

$$(8) \qquad \approx n \ln \left(\frac{1}{2} \right) + n \ln(n) - n - \left[\left(\frac{n+x}{2} \right) \ln \left(\frac{n+x}{2} \right) - \left(\frac{n+x}{2} \right) \right]$$
$$\qquad \qquad - \left[\left(\frac{n-x}{2} \right) \ln \left(\frac{n-x}{2} \right) - \left(\frac{n-x}{2} \right) \right]$$

$$(9) \qquad = n \ln \left(\frac{1}{2} \right) + n \ln(n) - \frac{n+x}{2} \ln \left(\frac{n+x}{2} \right) - \frac{n-x}{2} \ln \left(\frac{n-x}{2} \right)$$

$$(10) \qquad = n \ln \left(\frac{1}{2} \right) + n \ln(n) - \frac{n}{2} \ln \left(\frac{n+x}{2} \right) - \frac{x}{2} \ln \left(\frac{n+x}{2} \right)$$
$$\qquad \qquad - \frac{n}{2} \ln \left(\frac{n-x}{2} \right) + \frac{x}{2} \ln \left(\frac{n-x}{2} \right)$$

While this may look somewhat messy, we are now able to take the necessary derivatives of $\ln P(x)$ in order to substitute them into the Taylor series approximation, Equation 5.

$$(11) \frac{d \ln P}{dx} = -\frac{n}{2} \left(\frac{2}{n+x} \right) \left(\frac{1}{2} \right) - \frac{1}{2} \ln \left(\frac{n+x}{2} \right) - \frac{x}{2} \left(\frac{2}{n+x} \right) \left(\frac{1}{2} \right)$$
$$\qquad \qquad - \frac{n}{2} \left(\frac{2}{n-x} \right) \left(-\frac{1}{2} \right) + \frac{1}{2} \ln \left(\frac{n-x}{2} \right) - \frac{x}{2} \left(\frac{2}{n-x} \right) \left(\frac{1}{2} \right)$$

$$(12) \qquad = \left(-\frac{n}{2} \right) (n+x)^{-1} - \frac{1}{2} \ln \left(\frac{n+x}{2} \right) - \frac{x}{2} (n+x)^{-1} + \frac{n}{2} (n-x)^{-1}$$
$$\qquad \qquad + \frac{1}{2} \ln \left(\frac{n-x}{2} \right) - \frac{x}{2} (n-x)^{-1}$$

Recall from our spreadsheet exploration, we found that $x = 0$ was the most visited location. Therefore, $x = 0$ should be a maximum. For this to occur, the derivative of $\ln P$ evaluated at $x = 0$ should be 0. Let's check this.

$$\frac{d \ln P}{dx}\bigg|_{x=0} = -\frac{1}{2} - \frac{1}{2} \ln\left(\frac{n}{2}\right) - 0 + \frac{1}{2} + \frac{1}{2} \ln\left(\frac{n}{2}\right) - 0$$
$$= 0$$

Question 6. In a similar fashion as above, show that

$$\frac{d^2 \ln P}{dx^2}\bigg|_{x=0} = -\frac{1}{n}$$

We are now ready to return to the Taylor approximation, Equaiton 5, and make the above substitutions to arrive at

(13) $$\ln P(x) = \ln(P(x)_{x=0}) + 0 + \frac{1}{2}\left(-\frac{1}{n}\right)x^2.$$

For ease of notation, let $\ln(P(x)_{x=0}) = P^*$. Then exponentiating both sides of Equation 13 we have

(14) $$P(x) = P^* \exp(-x^2/2n).$$

Now let's compute P^*. To insure that our probability distribution is properly normalized, we integrate over all probabilities to equal one. To find P^*, we integrate Equation 14 as follows:

(15) $$\int_{-\infty}^{\infty} P(x)dx = \int_{-\infty}^{\infty} P^* \exp(-x^2/2n)dx = 1.$$

Using an integration table, we see that $\int_{-\infty}^{\infty} \exp(-ax^2)dx = \sqrt{\pi/a}$, so we have

(16) $$\frac{1}{P^*} = \int_{-\infty}^{\infty} \exp(-x^2/2n)dx = \sqrt{2\pi n}.$$

Solving for P^* and substituting into Equation 14 we have

(17) $$P(x) = \frac{1}{\sqrt{2\pi n}} \exp(-x^2/2n).$$

Equation 17 describes the distribution of the final ending distance x for a particle that started at $x = 0$ and took n random steps to the left(-x) and to the right(+x).

Exercise 11. Create a plot of $P(x)$ as a function of x from $x = -10$ to $x = 10$ in steps equal to 1. You see that this is the same bell shaped curve as we found from the histogram. The probability curve is smooth and does not contain random variations. If we calculated a histogram for thousands upon thousands of molecules, then the histogram would be as smooth as the probability function.

So, how does this help us solve our original sugar-tea question: How far will our sugar molecule travel in a straight line in a given period of time (or steps)? From our spreadsheet exploration, we know that taking the average distance traveled for a molecule on a random walk does not tell us much since the average distance is $x = 0$. Moreover, if we try to sum the length of the path, steps to the right would be positive, but steps to the left would be negative, giving us a meaningless average.

Instead, we consider the *root mean-square* (rms) distance.

We can use the probability distribution to calculate the average distance traveled. The following development of the avergae distanced traveled is based on McQuarrie[5]. For a continuous distribution such as our $P(x)$, we would find x_{ave} using

$$(18) \qquad\qquad x_{ave} = \int_{-\infty}^{\infty} xP(x) \ dx$$

To understand where this equation comes from consider a specific example of a non-continuous or discrete probability distribution. In a carnival game, a child tosses a ball and depending on where the ball lands, the child earns a certain number of points. The worker who runs the game, keeps track of the number of times the ball lands in each area so he can calculate the average number of points a child will likely earn. The worker found that 40% of the time the ball landed in the one point area, 25% of the time the ball landed in the two point area, 30% of the time the ball landed in the three point area, and 5% of the time the ball landed in the four point area. From this information, you can see that it is easy to calculate the average score by multiplying the percent of the time a ball lands in each area by the number of points designated to that area. We find that $0.40*1+0.25*2+0.3*1+0.05*4 = 2$ and that the average points earned is two. We can express this equation as a more general formula

$$x_{ave} = \sum_{1}^{4} xP(x)$$

where $P(x)$ is our discrete probability distribution and x is the number of points. This equation is easily generalized to a continuous probability distribution by assuming that the distance between different x values becomes infinitely small so that the sum is replaced with an integral. The limits of integration are the limits of the x values.

Now we can take the probability distribution we derived from our random walk example (equation 17) and substitute it into equation 18. With a little calculus or by remembering that the integral of an odd function (one that is exactly inverted about the origin) we see that

$$(19) \qquad\qquad x_{ave} = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi n}} \exp(-x^2/2n) \ dx = 0$$

This is exactly what we found in our spreadsheet exploration that the average distance traveled is zero because the molecules can move to the left and right with equal probability. Using the spreadsheet, we also realized that the root mean-square distance is a much more useful quantity to describe how the molecules spread out.

Therefore, we will next use the probability distribution to calculate the root mean-square distance. First we note that the standard notation for the average or mean of the x^2 values is $\langle x^2 \rangle$ where the angled brackets indicate an average value. The general formula for finding the mean-square distance is by analogy with equation 18 is

$$(20) \qquad \langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 P(x) \; dx$$

Because we want to find the root mean-square distance we have

$$(21) \qquad \langle x^2 \rangle^{1/2} = x_{rms} = \sqrt{\int_{-\infty}^{\infty} x^2 P(x) \; dx}$$

If we plug equation 17 into this equation, we find

$$(22) \qquad x_{rms} = \sqrt{\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi n}} \exp(-x^2/2n) \; dx} = \sqrt{n}$$

This result is exactly where we were in after exercise 10 after our spreadsheet investigation but we have now have calculated the functional dependence of x_{rms} on n instead of finding it from trial and error. From the discussion following the spreadsheet exercise, we know that we can link the number of steps, n, to time and can find the x_{rms} in terms of a real distance and not just step lengths to find for diffusion along a line that $x_{rms} = (2Dt)^{1/2}$. Again following the logic of this earlier discussion, we have for two dimensional diffusion, $x_{rms} = \sqrt{4Dt}$ and for diffusion in three dimensions we have $x_{rms} = \sqrt{6Dt}$.

6. Summary and Additional Exercises

We have now developed a model and successfully used the concept of a random walk to find an equation that describes the distance an average molecule in a large number of molecules diffuses as a function of time. Let's answer the sugar question! How long will it take for the sugar to spread throughout the glass of tea?

Exercise 12. At 25°C, how long will it take a sugar to spread out 5.0 cm in all three dimensions? Remember the root mean-square distance characterizes the spread of the diffusing molecule and this is diffusion in three dimensions. The diffusion constant of sugar in water is roughly 5×10^{-6} cm^2 s^{-1}.

Exercise 13. This problem is adapted from Tinico et. al [7]. Myoglobin is a protein that stores oxygen in cells. The diffusion coefficient for myoglobin in water at 20°C is 11×10^{-7} cm^2 s^{-1}. In half a second, how far would an average myoglobin molecule diffuse along a line? How does this distance compare to the size of a cell?

A cell's cytoplasm is more crowded than pure water. Would you expect the time it takes for myoglobin to diffuse in a cell to be longer or shorter than in pure water?

Exercise 14. What happens if the random walk is not quite random? To test this on the Excel sheet change the "Random" column from "=RANDBETWEEN(1,2)" to "=RANDBETWEEN(1,9)."

(1) What does this change on Excel effectively do?
(2) What is your prediction for what will happen to your distribution?
(3) What would be a real world example of a not quite random walk?
(4) Change your Excel sheet and test your prediction.

Exercise 15. Will our modeled system ever reach equilibrium? What is needed? What is present in the sweet tea example that is missing from our system? Equilibrium is only established in a closed system our model did not include the boundaries of the glass.

We can add boundaries to our model by assuming a wall exists at distance 4 and -4. We have just picked these values to illustrate our point. Clearly a glass is more than 8 steps wide. We need a way to tell our model that if a particle is at a distance 4 then the particle can only step left and if a particle is at a distance -4 then the particle can only step right.

Make a histogram for 200 particles undergoing 14 steps but change how you calculate the "Step" column. Instead of setting STEP = POWER(-1, random number) use Excel's IF function. This function has the form =IF(condition, returned value if condition is true, returned value if condition is false). Set STEP = IF(ABS(B12)=4,-4/B12,POWER(-1,C12)).

What does this command tell Excel to do? Excel looks at the initial position (B12). If the absolute value of this position equals 4 meaning that the particle is either at position -4 or 4, then the IF statement is TRUE and either the STEP is forced to be +1 (if the initial position is -4) or the STEP is forced to be -1 (if the initial position is 4). The "returned value if condition is true" is set to equal to (-4/B12). Thus if the particle is at -4, then -4/B12 = -4/-4 = 1 and the particle is forced to step right (STEP set at 1). Thus if the particle is at 4, then -4/B12 = -4/4 = -1 and the particle is forced to step left (STEP set at -1). If the particle is not at the boundary so that the IF statement is FALSE, then we use our normal expression for STEP determined by the random number generator, POWER(-1,C12), where C12 contains the random number.

Set up your Excel worksheet, and observe the histogram. Does it appear that the system is now approaching a constant concentration? How could we make our model more closely approximate equilibrium?

REFERENCES

[1] Anta, Juan A., *Random walk numerical simulation for the solar cell applications*, Energy Environ. Sci., **2** (2009), 387-392.

[2] Bromberg, Sarina and Dill, Ken A. 2003. *Molecular Driving Forces*. Garland Science, New York, NY.

[3] Davis, L. C., *Diffusion Confusion*, Journal of Chemical Education, **73** (1996), no. 8, 824-825.

[4] Engel, Thomas and Reid, Philip, *Physical Chemistry*, second edition, Prentice Hall, 2010.

[5] McQuarrie, 1983 . *Quantum Chemistry* Univ. Science Books, pages 89-92.

[6] Faísca, Patricia F. N., *The nucleation mechanism of protein folding: a survey of computer simulation studies*, Journal of Physics: Condensed Matter, Volume 21 **37** (2009), 373102.

[7] Tinoco, I., Sauer, K., Wang, J., Puglisi, J. 2002. *Physical Chemistry: Principles and Applications in Biological Sciences*. Prentice Hall, page 274.

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, DENISON UNIVERSITY, GRANVILLE, OH 43023 EDWARDSA@DENISON.EDU, LUDWIGL@DENISON.EDU