BE/APh161: Physical Biology of the Cell
Homework 3
Due Date: Wednesday, January 29, 2020

"You can't depend on your eyes when your imagination is out of focus." - Mark Twain

This problem set is largely devoted to practice with scaling arguments like those introduced in class. The idea is to develop simple scaling arguments to figure out answers to seemingly impossible problems such as "what is the tallest mountain allowed from physical principles on the planet Mars?" In class, I developed such an argument as our first approach to discovering how diffusion time depends upon the length scale over which diffusion occurs. The second main thread is to explore the binomial distribution that arose in our thinking about diffusion.

1. Migration of the bar-tailed godwit

Animal migrations are one of the greatest of interdisciplinary subjects, bringing together diverse topics ranging from animal behavior to the physics of navigation to the metabolism required for sustained long-distance travel. The bar-tailed godwit is a small bird that each year travels between Alaska and New Zealand on the same kind of incredible nonstop voyage taken by happy tourists in modern long-distance jetliners. During a visit to New Zealand's South Island, I had the chance to see these amazing birds in Okarito Lagoon with a naturalist guide who claimed that over the course of their ten-day, ten-thousand kilometer trip, these migratory birds lose 1/3 of their body mass. In this problem, we make a series of simple divide-and-conquer estimates to see whether this claim might be true.

(a) Using dimensional-analysis arguments, work out how the drag force experienced by flying godwits depends upon the density of air, the speed of the birds and the size of the birds. Specifically, work out the coefficients α, β and γ in the expression

$$F_{drag} = \text{const.}\rho^{\alpha} v^{\beta} L^{\gamma}. \tag{1}$$

(b) Work out the power expended by the bar-tailed godwit to overcome the drag force. Then, work out the total energy expended during the ten-day migration in overcoming this drag force.

(c) Given that burning fat yields 9 kcal/g, work out the number of grams of fat that would need to be burned to sustain the ten day flight of the bar-tailed godwit.

2. The height of mountains on Mars

As a prelude to thinking about the buckling force of one-dimensional rods in the context of animal legs, we examined the physics behind mountain height. Using a simple relationship between mountain height and the weight of a column of rock and the stress needed to crush rock, we made an estimate of mountain height. In this problem, we use the observed height of the Olympus Mons on Mars which is 22 km high, to estimate the gravitational acceleration on the Red Planet.

(a) The estimate given in class was very hand wavy. In this part of the problem, let's do better. Specifically, the street fighter approach adopted in class argued that the mountain is a cylinder. Now consider a conical mountain with base radius R and height h and improve our earlier estimate. Explain the relationship you choose between h and R by commenting on some real world mountains.

(b) Given the scaling estimate for mountain height derived above, work out the ratio of mountain heights on Mars and those on Earth. Make sure to state all of your assumptions in constructing this ratio and then solve for g_{Mars}.

(c) As a second approach, use the observed $g_{Mars} \approx 3.7$ m/s^2 to make an estimate of the height of the tallest mountain on Mars.

3. The length scale of morphogen gradients

Later in the course, we are going to introduce the important and fascinating

topic of reaction-diffusion equations as a window onto the process of pattern formation. One of the outcomes of the careful analysis we will do there is the existence of solutions to the equations describing morphogen dynamics that lead to morphogen gradients. In this problem, we exploit the skills we have been working out on scaling analysis to figure out how the length scale of the morphogen depends upon the two key parameters we imagine matter, namely, the diffusion coefficient of the morphogen proteins and their degradation times.

(a) Imitate the scaling analysis we have performed in class to find an expression for the length scale l in terms of the parameters D and τ that describe the diffusion and degradation of the morphogen protein, respectively.

(b) Given a diffusion constant of $D = 10\ \mu m^2/s$ and a degradation time $\tau = 50$ min, work out the length scale of a morphogen in a fly embryo.

4. Averages and the binomial distribution

In class we worked out the average position of a one-dimensional random walker by evaluating the expression $\langle x \rangle = \langle (2n_r - N)a \rangle$. This required us to evaluate the average

$$\langle n_r \rangle = \sum_{n_r=0}^{N} n_r p(n_r, N), \tag{2}$$

where for the case of the binomial distribution,

$$p(n_r, N) = \frac{N!}{n_r!(N-n_r)!} p^{n_r} (1-p)^{N-n_r}. \tag{3}$$

(a) Imitate the analysis given in class and show that the probability distribution is normalized, namely,

$$\sum_{n_r=0}^{N} p(n_r, N) = \sum_{n_r=0}^{N} \frac{N!}{n_r!(N-n_r)!} p^{n_r} (1-p)^{N-n_r} = 1 \tag{4}$$

by invoking the binomial theorem.

(b) Similarly, imitate the analysis given in class and show that the average position of the random walker is given by $\langle x \rangle = \langle (2n_r - N)a \rangle$ and use

$$\langle n_r \rangle = \sum_{n_r=0}^{N} n_r p(n_r, N), \tag{5}$$

to evaluate the average. Specifically, find

$$\langle n_r \rangle = \sum_{n_r=0}^{N} n_r p(n_r, N) = \sum_{n_r=0}^{N} n_r \frac{N!}{n_r!(N-n_r)!} p^{n_r}(1-p)^{N-n_r} = Np. \tag{6}$$

To do so, justify and use the approach I introduced that says

$$\langle n_r \rangle = p\frac{\partial}{\partial p} \sum_{n_r=0}^{N} \frac{N!}{n_r!(N-n_r)!} p^{n_r} q^{N-n_r}. \tag{7}$$

(c) Finally, we go after the result we really wanted, namely,

$$\langle x^2 \rangle = \langle (n_r - n_l)^2 a^2 \rangle = a^2 \langle (2n_r - N)^2 \rangle. \tag{8}$$

Expand the term in brackets and then use the same trick as in the previous part, namely,

$$\langle n_r^2 \rangle = p\frac{\partial}{\partial p} p \frac{\partial}{\partial p} \sum_{n_r=0}^{N} \frac{N!}{n_r!(N-n_r)!} p^{n_r} q^{N-n_r}, \tag{9}$$

to obtain your result for $\langle x^2 \rangle$. Make sure you connect to what we did in class and explain the meaning of this result for the excursion of the walker.

5. Counting molecules with the binomial distribution

(a) Begin by reading the paper by Rosenfeld *et al.* entitled "Gene Regulation at the Single-Cell Level" (posted on the website with the homework) and write a one paragraph commentary on the paper with special reference to how they used the binomial partitioning as a way to count repressor proteins. What is the experiment they did and what were they trying to learn?

In the rest of the problem we work out for ourselves the ideas about binomial partitioning introduced in the Rosenfeld *et al.* paper in order to

consider the concentration of mRNA or proteins as a function of time in dividing cells. In particular, the point of this problem is to work out the concentration of mRNA or protein given that we start with a single parental cell that has N copies of this mRNA or protein (in the experiments of Golding *et al.* they watch the mRNA dilution effect while in the experiments of Rosenfeld *et al.* this is a fluorescently-labeled transcription factor). In the Rosenfeld experiment, at some point while the culture is growing, the production of the protein is stopped by providing a chemical in the medium and then the number of copies per cell is reduced as a result of dilution as the cells divide.

Interestingly, this problem opens the door to one of the most important themes in physics, namely, that of fluctuations. In particular, as the cells divide from one generation to the next, each daughter does not really get $N/2$ copies of the protein since the dilution effect is a stochastic process. Rather the partitioning of the N proteins into daughter cells during division follows the binomial distribution. Analyzing these fluctuations can actually lead to a quantification of the number of copies of a protein in a cell.

(b) If we think of the N copies of the protein as being divided between the two daughters with N_1 going to daughter 1 and $N - N_1$ going to daughter 2, write the probability distribution $p(N_1, N)$. Next, work out the expected fluctuations in the partitioning process after each division by noting that the fluctuations can be written as $\sqrt{< (N_1 - N_2)^2 >}$, where N_1 and N_2 are the number of proteins that end up in daughter cells 1 and 2, respectively. Show that $\sqrt{< (N_1 - N_2)^2 >} = \sqrt{N}$.

(c) Next, look at the Rosenfeld paper and explain how measuring fluorescence variations can be used to calibrate the exact number of copies of the fluorescent protein in a cell. Specifically, assume that the fluorescence intensity in each cell can be written as $I = \alpha N$, where α is an as-yet unknown calibration factor and N the number of proteins in the cell. Explain what this equation means and why you think it is justified. Derive an expression relating I_1, I_2 and I_{tot} using the result of part (b). Make a plot of $\sqrt{< (I_1 - I_2)^2 >}$ versus I_{tot} and explain how to get the calibration factor α from this plot.

(d) Now we are going to repeat the Rosenfeld experiment numerically in

order to *fit* the calibration factor. Consider a fluorescent protein such that the calibration factor between the intensity and the number of fluorophores is 50, that is $I = 50N$. Generate intensity data by choosing $N_1 + N_2 = 10, 50, 100, 1000$ and 5000 and for each case, "partition" the proteins from the mother cell to the two daughters 100 times (i.e. as if you are looking at 100 mother cells divide for each choice of the protein copy number). Then, make a plot of the resulting $\sqrt{< (I_1 - I_2)^2 >}$ vs I_{tot} just as we did analytically in the previous problem. What I mean is that you need to make a plot of all of your simulation results. Then, do a fit to your "data" and see how well you recover the calibration factor that you actually put in by hand. Plot the fit on the same graph as all of the "data".