BE/APh161: Physical Biology of the Cell
Homework 4
Due Date: Wednesday, February 2, 2022

“We must travel in the direction of our fear.” - John Berryman

1. Laws of Cellular Growth Dynamics.

Much of our understanding of the natural world is couched in the language of the subject now known as “dynamical systems.” In a nutshell, the idea is to write down equations that tell us how some variable(s) of interest change in time. Often, this ends up being written in the form of coupled differential equations. Perhaps the most important and simplest of such dynamical systems is the law of exponential growth (or decay), relevant to thinking about the early stages of growth of a culture of cells, for example. In my recorded vignette “Laws of Cellular Growth,” I give a brief introduction to the way we can write dynamical equations that describe the evolution of the size of a population of cells ($N(t)$) as a function of time. In this problem, you are going to revisit the discussion I give there by solving for the dynamics of a population of bacterial cells both analytically and numerically.

(a) In this first part of the problem, write down and justify the differential equation for a population of dividing bacteria that is not limited by nutrient availability. Write an analytic solution. In addition, write a code in Python that integrates the equation over time. The basic idea for solving a differential equation of the form

$$\frac{dx}{dt} = f(x, t), \tag{1}$$

is time stepping. Here we propose you use the most naive method which instructs us to write the solution at time t as

$$x(t) = x(t - \Delta t) + f(x, t - \Delta t)\Delta t, \tag{2}$$

where the time step Δt is much smaller than any time scale in the system. The structure of for loops permits us to solve this numerically for any “well behaved” $f(x, t)$. Use this algorithm to solve your growth equation numerically and plot your numerical solution on the same graph as the analytic solution.

(b) The logistic growth equation is a phenomenological way of curbing unchecked growth and is written as

$$\frac{dN}{dt} = kN\left(1 - \frac{N}{K}\right), \tag{3}$$

where K is the so-called carrying capacity of the population. As in part (a), find an analytic solution for this equation and then find a numerical solution as well. Estimate the parameters relevant to your solution by thinking about a saturated bacterial culture. Comment on the meaning of the carrying capacity.

2. Fluorescence Recovery After Photobleaching by Pencil and Paper and by Computation.

NOTE: relevant vignettes to watch are those about diffusion.

In this problem, we are going to consider a "one-dimensional" cell. Of course, this sounds contrived, but really we are saying that the fluorescence only depends upon a single coordinate. We will consider the long axis of bacterial cells as the region to be photobleached. So, we will think of a region of length $2L = 4\ \mu m$ that initially has uniform fluorescence. We then photobleach (i.e. destroy the fluorescence) between $-a$ and a, with $a = 0.5\ \mu m$. Consider the concentration in the unbleached region to be $c_0 = 1\ \mu M$, and let the diffusing molecules have a diffusion coefficient of $10\ \mu m^2/s$. For each section below, we will use a different approach to working out the dynamics of the recovery process.

(a) FRAP by coin flips. In this part of the problem, you are going to write a simulation code that takes random walkers that start either in the region $-L$ to $-a$ or a to L and flip coins and let them jiggle around. For each such walker, the only rule you will need is that if on a given flip they try to leave the region from $-L$ to L, you will reflect them off the walls. The goal is to do 100s of such simulations and then plot the concentration as a function of position for different time points. After one time step, almost all of the walkers will be in the unbleached regions. But over time, more and more molecules will have ventured into the photobleached region. Your goal is to get the full profile of the independently diffusing molecules. Make plots of the concentration as a function of the number of steps. If the lattice parameter you use is $d = 40\ nm$, this will mean that you have 100 such lattice points. You can reconcile your simulation time step, the lattice parameter and the diffusion coefficient through the relation $D = d^2/\tau$, where τ is the time step.

(b) FRAP by math. For this part of the problem, I am going to explicitly walk you through the steps and your job is to really carefully demonstrate that everything works and holds together, showing all of the steps. To compute the recovery curves, we first solve the diffusion equation

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} \tag{4}$$

for the concentration of fluorescent molecules $c(x,t)$, with the initial concentration after photobleaching given by

$$c(x,0) = \begin{cases} c_0 & \text{for } -L \text{ to } -a \\ 0 & \text{for} -a \text{ to } a \\ c_0 & \text{for } a \text{ to } L. \end{cases} \tag{5}$$

We also impose the boundary condition $\partial c/\partial x = 0$ for $x = -L$ and $x = L$, which says that the flux of fluorescent molecules vanishes at the boundaries of the one-dimensional cell (no material flows in or out). This mimics the real-life situation with fluorescent proteins confined to the volume of the cell, to the cell membrane, or to some other subcellular structure.

To solve the diffusion equation with the prescribed initial and boundary conditions, we begin by expanding the concentration profile $c(x,t)$ in terms of cosine functions using "Fourier series,"

$$c(x,t) = A_0(t) + \sum_{n=1}^{\infty} A_n(t)\cos\left(\frac{x}{L}n\pi\right). \tag{6}$$

This expansion guarantees that the boundary conditions are met, namely each of the functions $A_n(t)\cos(xn\pi/L)$ has vanishing first derivatives with respect to x at $x = \pm L$. Furthermore, since the initial concentration profile takes the same values for positive and negative x, it is readily expanded in cosine functions since the concentration profile is symmetric about the origin. The solution of the diffusion equation now boils down to finding the functions $A_n(t)$ such that both the diffusion equation and the initial condition are satisfied.

To proceed, we substitute the series expansion of $c(x,t)$ into the diffusion equation. This yields

$$\frac{\partial A_0}{\partial t} + \sum_{n=1}^{\infty} \frac{\partial A_n(t)}{\partial t} \cos\left(\frac{x}{L} n\pi\right) = D \sum_{n=1}^{\infty} \left[-A_n(t)\frac{n^2\pi^2}{L^2}\right] \cos\left(\frac{x}{L} n\pi\right), \tag{7}$$

which, due to the orthogonality property of the cosine functions for different n (see Equation 11 below), turns into a set of independent differential equations,

$$\begin{array}{c} \frac{\partial A_0}{\partial t} = 0 \\ \frac{\partial A_n}{\partial t} = -\frac{Dn^2\pi^2}{L^2} A_n(t) \quad (n \geq 1) \end{array} \tag{8}$$

Show that the solution to each one of these (infinite in number) equations is an exponential function

$$A_n(t) = A_n(0)\mathrm{e}^{-(Dn^2\pi^2/L^2)t}, \tag{9}$$

which when substituted into Equation 6 gives

$$c(x,t) = A_0(0) + \sum_{n=1}^{\infty} A_n(0)\mathrm{e}^{-(Dn^2\pi^2/L^2)t} \cos\left(\frac{x}{L} n\pi\right). \tag{10}$$

Make sure you demonstrate this. The final piece of the puzzle is the determination of the constants $A_n(0)$.

To compute the initial amplitudes of the cosine functions, we resort to the orthogonality property of these functions, namely,

$$\int_{-L}^{L} \cos\left(\frac{x}{L} n\pi\right) \cos\left(\frac{x}{L} m\pi\right) \mathrm{d}x = L\delta_{n,m}. \tag{11}$$

In particular, multiply both sides of Equation 10 by $\cos(m\pi x/L)$ for different values of m, and then integrate over x to derive the equations

$$\begin{array}{c} A_0(0) = \frac{1}{2L} \int_{-L}^{L} c(x,0)\mathrm{d}x \\ A_n(0) = \frac{1}{L} \int_{-L}^{L} c(x,0) \cos\left(\frac{x}{L} n\pi\right) \mathrm{d}x \quad (n \geq 1) \end{array} \tag{12}$$

for the initial amplitudes. Substitute the initial concentration profile, $c(x,0)$, into these equations, and perform the integrals, to show that

$$\begin{array}{c} A_0(0) = c_0 \frac{L-a}{L} \\ A_n(0) = -2c_0 \frac{\sin(n\pi a/L)}{n\pi} \quad (n \geq 1) \end{array} \tag{13}$$

Put these results back into the derived formula for $c(x,t)$, Equation 10 and show that the solution for the concentration profile as a function of time is given by

$$c(x,t) = c_0\left[1 - \frac{a}{L} - \sum_{n=1}^{\infty} \frac{2\sin(n\pi a/L)}{n\pi} \mathrm{e}^{-(Dn^2\pi^2/L^2)t} \cos\left(\frac{x}{L}n\pi\right)\right]. \tag{14}$$

Make a plot of your resulting concentration profile as a function of time for several different times. Also, make sure you illustrate how your result depends upon how many terms you keep in the series. Obviously, you can't do an infinite number of terms. Note that at long times, such that t is much greater than L^2/D, which is the diffusion time for a box of length L, the concentration profile tends to a constant value equal to $c_\infty = c_0(1 - a/L)$. This can be understood in a very simple way. Namely, at long times, we expect diffusion to make the concentration profile uniform over the $2L$ interval. Show that the fact that the number of fluorescent molecules does not change in time leads to the equation

$$c_\infty(2L) = c_0[2(L-a)], \tag{15}$$

which gives the computed value of the concentration at long times.

(c) FRAP by chemical master equation. In the vignette entitled "Diffusion: Master Equation" I wrote down the evolution equation

$$p(x,t+\Delta t) = p(x,t) + (k\Delta t)p(x-a,t) + (k\Delta t)p(x+a,t) - (k\Delta t)p(x,t) - (k\Delta t)p(x,t). \tag{16}$$

In that vignette, I argued that the equation as written is the basis of a very nice way to numerically investigate diffusion problems. Here you will consider a 4 μm long cell that is discretized into 100 boxes. As you did in the previous two parts of the problem, you are going to integrate the chemical master equation by starting with the initially bleached profile and then plotting the concentration as a function of time.

3. The Standard Candle: Counting Proteins with Partitioning Statistics.

NOTE: to do this problem, the vignettes "Biological puzzles and probabilistic thinking," "Carboxysomes and binomial partitioning" from the 2021 edition of the course will be helpful.

(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 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 $\sqrt{< (N_1 - N_2)^2 >} = \sqrt{N}$. When I do this calculation, I find it convenient to write $N_2 = N - N_1$. Basically, this reduces the problem to having to calculate $\langle N_1 \rangle$ and $\langle N_1^2 \rangle$ since once you have those two quantities you can evaluate $\sqrt{< (N_1 - N_2)^2 >}$.

(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".