

AnnouncementsOffice hours this week:

4-5 PM on Tuesday the 6th (right after section!)

7-8:30 PM on Wednesday the 7th, virtual at <https://harvard.zoom.us/j/98197473635>.

or by appointment

Section next week:

3-4 PM on Tuesday the 13th at Mallinckrodt 217

Cool paper(s) I'm reading:<https://rdcu.be/cURgL>

DNA-PAINT MINFLUX nanoscopy: New letter from Stefan Hell's group combining MINFLUX with DNA-PAINT (2 different techniques for breaking the diffraction limit in optical imaging) to achieve multi-color super-resolution images with very high resolution (<3 nm localization precision). Highly recommended if you're interested in super-resolution imaging.

<https://pubmed.ncbi.nlm.nih.gov/30944477/>

Genetic compensation triggered by mutant mRNA degradation: OK, this isn't as directly relevant to the course materials, but I thought it was pretty wack: Mutating mRNAs to add a premature termination codon (PTC) causes cells to compensate by upregulating the expression of mRNAs with sequence homologies. The authors report a mechanism in which the mutated mRNA triggers the COMPASS complex to add epigenetic markers at genes with similar sequences, leading to increased expression.

Problem 1: Review of Taylor series & rules for approximations

Often, biophysical models are too complex to be evaluated across their entire domains analytically, but we can still gain valuable insights by trying to find a function that captures the behavior well in a local region of interest. This problem takes you through the process of making such approximations using a tool you probably learned in calculus: Taylor series. These tricks will come up frequently in lectures and on the homework, and you will likely need them if you work with an analytical model for your final project. If you already feel comfortable with these tools, you can skip this problem.

Part 1: Math review: Consider a function $f(x)$ that is infinitely differentiable at some real or complex number a . Then its Taylor series is given by the power series:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \dots \quad (1)$$

$$= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n. \quad (2)$$

To practice, calculate the first three terms of the Taylor series around $x = 0$ of the following functions:

1. e^x

2. $\sin(x)$

Part 2: Making approximations OK, so now we have a way to represent $f(x)$ as an infinite series, but this isn't really an easier form to work with. To approximate $f(x)$ in the vicinity of a , we can truncate its Taylor series and keep only the first few terms. When we talk about a "first-order approximation" of a function, we mean keeping only up to the linear term of the Taylor series. The second-order approximation means keeping up to the quadratic term, and so on.

How do we know how many terms to keep in our approximation? Typically, we want to keep as few terms as we can get away with while still being able to gain insight into our function. In practice, you should truncate the series at the lowest-order term to start. If you find that all of your terms vanish, you can go back to this step and add the next lowest-order term. One critical note here is that all approximations made in solving the same problem need to be consistent. If you use a first-order approximation in one step, you cannot use a second-order one in a future step.

Below are a list of some common Taylor approximations around $x = 0$ encountered in the physical sciences (excluding the two you solved for above). You can use this as a reference for the rest of the semester:

1. $\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$

2. $\tan(x) = x + \frac{x^3}{3} + \frac{2x^5}{15} + \dots$

3. $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$

4. $(1+x)^n \approx 1 + nx$

Problem 2: Alternative derivation of random walk statistics

This section will guide you through an alternative derivation of the mean and variance of a particle performing a random walk.

a) Consider an ensemble of N particles. We denote position of the i th particle after the n th step of a random walk is $x_i(n)$. At each step, the particle is equally likely to take a step of size Δx to the left or to the right. Write down a recursive relation for $x_i(n)$ and use it to solve for the average position of the entire ensemble of particles $\langle x(n) \rangle$.

b) Now do the same for the mean squared position of all particles in the ensemble $\langle x^2(n) \rangle$. How does your answer scale with the number of steps?

c) Let's now relax the requirement that the particle be equally likely to step to the left and to the right. Denote the probability that the particle steps to the right p and the probability that it steps to the left $q = 1 - p$. The probability that a particle steps exactly k times to the right is then given by the binomial distribution¹:

$$P(k; n, p) = \frac{n!}{k!(n-k)!} p^k q^{n-k}. \quad (3)$$

From this function, compute $\langle k \rangle$ and $\langle k^2 \rangle$. Use these results to compute the mean displacement $\langle x(n) \rangle$ and mean squared displacement $\langle x^2(n) \rangle$ of the particle. Check that your answers reduce to the ones you derived above when $p = q = 0.5$.

¹This probability distribution will come up later in the class. Keep an eye out!

Problem 3: Some more back-of-the-envelope calculations

These questions are left deliberately open-ended. You should feel free to look up any numbers you think might be useful.¹

a) How does the size of a protein compare to that of the mRNA that encodes it?² Consider both the lengths when fully extended and the effective size in solution. For the second quantity, a useful approximation that we will prove when we cover polymers is that the end-to-end distance is roughly the geometric mean of the contour length and the persistence length, $\sqrt{\ell_p L_C}$.

b) What is the information content of the human genome in bytes? How does the density of information (information/weight) compare to that of a modern hard drive?

c) What is the concentration (units of molarity) of DNA nucleotides in a human eukaryotic cell? Assume the cell is in the G1 phase, i.e. it has not yet replicated its genome.

¹The Bionumbers website is often helpful for these sorts of questions: <https://bionumbers.hms.harvard.edu/search.aspx>

²This estimate is complicated by the fact that mRNAs have both 5' and 3' untranslated regions. You can either ignore this fact or examine how your answer changes as you vary the length of these UTRs.