This document is a work in progress intended for use in this course only.

QUANTITATIVE PRINCIPLES IN BIOLOGICAL SYSTEMS
Instructor: Po-Yi Ho
2026 Spring

LECTURE 1. CHEMOTAXIS AND RANDOM WALKS

Introduction

This course is my attempt to answer the following common questions: What do we want to understand about biological systems? Are there quantitative principles? What has modeling done for biology? What more can modeling do for biology? Ideally after this course, we will have developed some intuition for the quantitative principles such that our feeling of surprise for various biological systems will have shifted.

The materials are intended for advanced undergraduates and graduate students. We aim to provide a common ruler for researchers from diverse backgrounds to reason about biological systems. Mathematical and biological concepts will be introduced in a just-in-time and just-enough manner. The goal is to highlight the questions and principles that will empower us to work together with each other.

Syllabus and learning goals

To build this common language, we will cover a broad array of systems, selected because they highlight the emerging principles. These systems include some of the most influential topics in the past few years, such as predicting the next pandemic strain and neural networks. Different systems necessitate different quantitative tools, in this sense the diversity of tools mirrors biological diversity. This breadth might cause some confusion, so I want to explicitly highlight our learning goals.

The first goal is to convince us that biological systems follow quantitative principles. The principles are dense in terms of the number of concepts involved. To convey these principles, we will spend a lot of time on not just the concepts themselves, but the key connections among them.

If biological systems follow quantitative principles, then a corollary is that interdisciplinary research on biological systems is not simply mix and match. In this sense, being able to intuit the key principles and ingredients involved is a rare and valuable research skill.

How do we develop this skill? Our third learning goal is to convince ourselves that we can do it. For systems as complex as the ones we will cover, often the most practical way to understand them is to jump in, because there are too many factors to sort out at the outset. We will take this approach in class by emphasizing the actionable research questions that lead to the application or discovery of the various concepts and principles. This pragmatism also translates to our approach for math and problem sets.

In sum, I want to show you several research threads that are fundamental, historied, and beautiful. They demonstrate that there are models for biological systems that are neither too detailed (sometimes criticized as “stylized facts” or “just statistics”) nor not detailed enough (sometimes criticized as “don’t worry models” or “fact-free biology”). When models hit the sweet spot, as we shall see, they can become principles, at which point they demystify our naïve surprises and raise new questions.

This course is under active development. Please ask questions and engage actively. I welcome all feedback! Last year we received perfect scores from all students but one – but unfortunately not a lot of feedback on how to improve.

Our first model system
Much of this lecture and the next are inspired by William Bialek’s Biophysics.

Teaching philosophy on math
Before we go through our first derivation, I want to explicitly state our learning goals on math. To accomplish the learning goals of this course in one semester, we will only focus on fundamental concepts and key derivations in class. Unfortunately, we will not have the time to go over extensive explanations, practice examples, or calculations that are important but disconnected from the research context. Nonetheless, during the derivations, we will highlight keywords and conceptual links that will allow you to fill in the details in your own time. It is worth reiterating that the goal is not for all of us to become modelers and theorists. Rather, the goal is to spend enough time to develop both a basic knowledge map so that we will know what to look up when encountering a mathematical problem in our research, as well as an appreciation for the rigor involved in mathematical models of biological systems.

Teaching philosophy on problem sets (and large language models)
For these purposes, problem sets are essential. The problems are meant for us to get practice on applying the quantitative principles discussed in class. They might contain aspects that are unclear, which is just like research. However, they are not completely open-ended.

I want to be explicit about our policy on LLMs. We believe that they are excellent tools to help us understand the materials covered, but we should be careful to avoid falling into an illusion of understanding. This trap for learning is not new and the solution is tried and true: Ensure that you understand the materials by engaging actively with your peers. In fact, I have found that LLMs did not affect the distribution of learning outcomes at all. Grades from last year’s class were still normally distributed with values that reflected effort and curiosity rather than the extent of LLM use.

Random walks by combinatorics
The next sections are inspired by Ariel Amir’s Thinking Probabilistically.

With the logistics out of the way, let’s return to the random walk problem. To develop intuition for the random motion of bacteria, let’s consider the simplest possible one-

dimensional random walk where at each time step, a particle moves left or right one space with equal probability. We can numerically simulate this process with one line of pseudocode

$$x_{t+1} = x_t + \xi_t,$$

where x_t is the position of the particle at time t and ξ_t is a random variable taking values of +1 or -1 with equal probability at each time step. More generally, this equation is a simple case of a stochastic differential equation, which we will discuss in more detail soon. For now, let's simulate it and derive some basic properties.

What is the mean value of the position of the particle after N steps? From symmetry,

$$\langle x_N \rangle = 0,$$

where the bracket $\langle \quad \rangle$ denotes the average over many random instances.

The mean has to be zero but we do not expect the particle to be exactly at the origin. What is the typical distance of the particle from the origin after N steps? Suppose that the particle is at position x at time N, then at the next time step, the mean squared distance is

$$[(x+1)^2 + (x-1)^2]/2 = x^2 + 1.$$

Thus, the mean squared distance increases by 1 after every step, so we have

$$\langle x_N^2 \rangle = N.$$

The typical distance scales like the square root of time.

What about the probability distribution of positions? Let's calculate it using nothing but combinatorics. The probability to be a distance M away after N steps is

$$P(M) = \frac{1}{2^N}\binom{N}{R} = \frac{1}{2^N}\frac{N!}{R!\,(N-R)!},$$

where R is the number of steps to the right, with $M = R - (N - R)$. To evaluate this expression, we can use Stirling's formula to approximate factorials, $N! \approx \sqrt{2\pi N}(N/e)^N$, which gives us

$$P(M) \approx \frac{1}{\sqrt{2\pi}} e^{\cdots}.$$

We can make progress by realizing that $M \ll N$, and so

$$P(M) \approx \frac{2}{\sqrt{2\pi N}} e^{-\frac{M^2}{2N}}$$

Thus, the distribution is approximately Gaussian. Using nothing but brute force counting, we answered Pearson's question in a simple case. In this sense, as Laplace said, "probability theory is nothing but common sense reduced to calculations".

Specifying how random

The above example is a case where our intuition works. However, in probability and for complex systems, our intuition will often fail. Let's see a curious example. While not directly related to bacterial chemotaxis, this example will highlight important mathematical concepts.

Consider a real-life collection of seemingly random numbers like the population sizes of the countries in the world, the set of physical constants, the lengths of rivers, stock prices, and so on. What is the probability distribution of the first digit?

Perhaps surprisingly, it is not uniform. Instead, the probability of the first digit d closely follows Benford's law

$$P(d) = \log_{10}\left(\frac{d+1}{d}\right).$$

The probability for the first digit to be 1 is larger than that for 2, and so on.

Benford's law is surprising because we might have assumed that the underlying distributions of these variables are uniform. However, for many natural datasets, it is the logarithm of the variable that is uniformly distributed. If x is such a variable, then the probability distribution $p(x) \propto 1/x$, and

$$P(d) \propto \int_d^{d+1} p(x)dx = \log\left(\frac{d+1}{d}\right).$$

Two common features of these natural datasets are that they have units and that they span many orders of magnitude. These features suggest that $p(x)$ must be invariant under a change of scale, which then implies that $p(x) \propto 1/x$, and Benford's law follows.

The takeaway is that whenever we encounter a random process, we should first specify exactly how random.

Brownian motion, diffusion, and the Einstein relation

Let's return to bacterial chemotaxis. What might be the molecular level mechanism that underlies the random motion of a bacterium? Without the knowledge we have today about bacteria, one might guess at first that the mechanism is the collection of random collisions between the random walker and the many particles that make up the liquid. To test this idea, let's derive an equation that would describe this process, which is known as Brownian motion.

Since the motion is random, we are interested in the quantity $p(x,t)$, the probability density to find the walker at position x at time t. Let's consider discrete time as before, with interval τ, but take the jump size to be continuous. Then the probability at $t+\tau$ can be written as

$$p(x, t+\tau) = \int p(x-s,t)q(s)ds,$$

where $q(s)$ is the probability distribution for jump size s (as a result of the collection of random collisions in an interval τ), and the integral expresses the probability for the walker to move to position x. Moreover, assuming that p is smooth, we can expand about x to write

$$p(x-s,t) \approx p(x,t) - \frac{\partial p}{\partial x}s + \frac{1}{2}\frac{\partial^2 p}{\partial x^2}s^2 + \cdots.$$

The integral of the linear term is zero due to symmetry, so we find that

$$p(x,t+\tau)-p(x,t)\approx\frac{1}{2}\frac{\partial^2 p}{\partial x^2}\int s^2 q(s)ds+\cdots.$$

Finally, dividing by τ, we find the diffusion equation

$$\frac{\partial p}{\partial t}=D\frac{\partial^2 p}{\partial x^2},$$

where D is known as the diffusion coefficient and is linked to the jump size distribution by

$$D=\frac{\langle s^2\rangle}{2\tau}.$$

We can check that the diffusion equation is solved by the normal distribution

$$p(x,t)=\frac{1}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}},$$

where the variance $\sigma^2=2Dt$ increases linearly in time, as in the one-dimensional walker case.

In this way, we have linked the microscopic distribution of jump sizes to the macroscopic diffusion coefficient. However, the distribution of jump sizes is not easily measurable, nor has it been linked to other physically measurable properties. To answer whether Brownian motion can explain the random movement of a bacterium, let's derive the Einstein relation which relates the diffusion coefficient to the viscosity of the fluid.

Consider a collection of Brownian particles of mass m suspended in a viscous fluid under gravity. In this setting, gravity and diffusion counteract each other, generating a concentration profile that can be measured, and also shown from the ideal gas law, to be exponential

$$p(x)\propto e^{-\frac{mgx}{kT}},$$

where x here denotes the height. Then, diffusion leads to a flux up because diffusion tends to homogenize the concentration. To see this effect, consider the continuity equation, which just states that particle number is locally conserved,

$$\frac{\partial p}{\partial t}=-\frac{\partial J}{\partial x},$$

where J is the flux of the particles. Combining the continuity with the diffusion equation, we find the phenomenological Fick's law

$$J=-D\frac{\partial p}{\partial x},$$

stating that the flux is proportional to the concentration gradient, with proportionality constant equal to the diffusion coefficient.

We also know that in highly viscous fluids, gravity pulls the particles down at a constant speed μmg, where μ is the mobility of the particle. From fluid dynamics, we know that for a spherical particle of radius r, the mobility

$$\mu=\frac{1}{6\pi\eta r},$$

where η is the viscosity of the fluid, a measurable quantity. Thus, the flux due to gravity is μmgp. Combining with Fick's law, we find the Einstein relation

$$D = \mu kT = \frac{kT}{6\pi\eta r}.$$

This equation relates the diffusion coefficient to the viscosity of the fluid. Note that the factors mg cancel, reflecting the fact that this relation can be derived not just for gravity but for any force. The Einstein relation is an example of a fluctuation-dissipation relation – here, diffusion is the fluctuation, and the drag from the fluid is the dissipation.

Before returning to bacterial chemotaxis, I want to highlight some terms and concepts that we just encountered: stochastic differential equations, Langevin equations, the diffusion equation, Fokker-Planck equations.

Bacterial chemotaxis is an active diffusion process

The Einstein relation tells us that the diffusion coefficient of a bacterium-sized particle in water is on the order of 1 um^2/s, much smaller than the apparent diffusion coefficient observed! The solution to this apparent paradox is simple: The diffusion coefficient of dead bacterial cells is indeed around 1 um^2/s, i.e. live bacteria are undergoing active diffusion. They are spending energy to reach diffusion coefficients on the order of 100 um^2/s.

Note that without knowing the length or time scales involved, it would have been difficult to distinguish between the random trajectories of live versus dead bacterial cells, since in both cases the displacement scales with the square root of time. This feature is known as self-similarity, and is an example of coarse-graining and universality, recurring concepts in modeling biological systems.

We can approximate the energy that the bacteria must be spending to diffuse at 100 um^2/s. Needless to say, the energy must be a significant amount, which begs the question of why are they doing that?

Life at low Reynolds number

Chemotaxis

Poisson processes

To see how an exponential distribution might arise, let's consider an everyday example: bus timings. For bus station A, buses are dispatched every 6 minutes. For bus station B, the manager throws a die every minute and dispatches a bus if they roll a 6.

How many buses are dispatched per day? For station A, 240 buses. For station B, the probability for k buses to be dispatched is

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k},$$

where $n = 1{,}440$ is the number of die throws per day and $p = 1/6$ is the probability to dispatch per throw. This distribution is known as the binomial distribution. The average number is still 240.

You arrive at a random time to the bus station. How long must you wait for a bus? For station A, 3 minutes. For station B, the probability to wait for T minutes is

$$P(T) = (1-p)^{T-1}p.$$

This distribution is known as the geometric distribution. The binomial distribution describes the number of events, while the geometric distribution describes the time between events. What is the mean of the geometric distribution?

Rather than evaluating the discrete combinatorics, let's develop our intuition by taking time to be continuous such that the number of trials $n \to \infty$ with the expected number of events per interval $\lambda = np = rt$ fixed, where r is the rate at which buses are dispatched. In this setting, the quantity $P(k,t)$, the probability to have dispatched k buses by time t, obeys

$$\frac{\partial P(k,t)}{\partial t} = rP(k-1,t) - rP(k,t),$$

except at $k = 0$, where

$$\frac{\partial P(0,t)}{\partial t} = -rP(0,t).$$

At the beginning, $P(0,0) = 1$. Then,

$$P(0,t) = e^{-rt}.$$

For $k = 1$, we get a linear differential equation with solution

$$P(1,t) = rte^{-rt}.$$

We can iterate for all $k = 2$, which suggests that for all k,

$$P(k,t) = e^{-rt}\frac{(rt)^k}{k!}.$$

We can also take the interval to be given and write the distribution in terms of λ,

$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

This distribution is known as the Poisson distribution. We have also just encountered a master equation, a set of first-order differential equations describing the probability of the system to occupy each of a discrete set of states. We will encounter master equations again in the next lecture. The Poisson distribution arises in many contexts, as we shall see. It has the special property that its variance and mean are both equal to λ, the expected number of events in the interval. In this setting, the Poisson distribution is the continuous analogue of the binomial distribution.

What is the continuous analogue for the geometric distribution of intervals between events? The probability that the waiting time is larger than t is

$$P(T > t) = P(k = 0, t) = e^{-rt}.$$

Thus, the cumulative distribution function, or the probability that the waiting time is less than or equal to t is

$$P(T \leq t) = 1 - e^{-rt}.$$

The probability distribution is its derivative,

$$P(t) = re^{-rt}.$$

This distribution is known as the exponential distribution, and its mean is $1/r$. Returning to bus timings, the average wait time for station B is also 6 minutes.

In hindsight, this result should have been obvious, since station B is memoryless. More precisely, the probability for the waiting time T to be larger than $s + t$, given that it is already larger than s, is

$$P(T > s + t | T > s) = \frac{P(T > s + t; T > s)}{P(T > s)}.$$

Here, the left-hand side is the conditional probability and the numerator is the joint probability that both events occur. In this case, the numerator is just equal to $P(T > s + t)$, and so for the exponential distribution, we find that

$$P(T > s + t | T > s) = e^{-r(s+t)+rs} = e^{-rt} = P(T > t).$$

Precision of molecule counting

Sampling a small volume to obtain the concentration