

Chapter 5

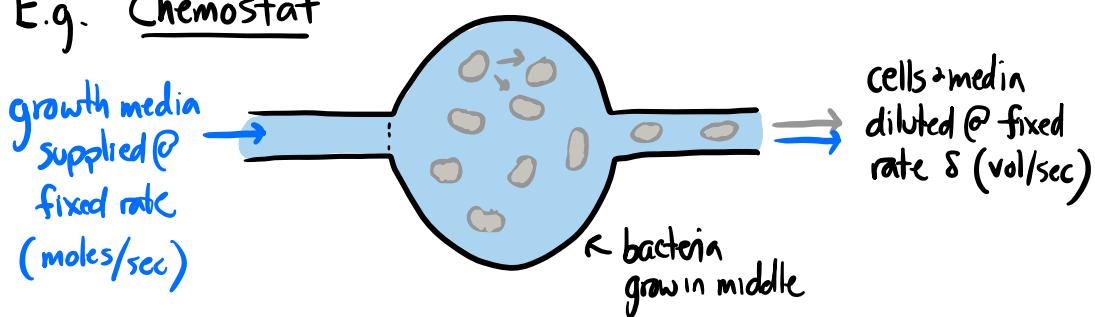
Microscopic Models and the Diffusion Limit

5.1 Microscopic models of evolution

In the previous chapter, we developed a *microscopic model* for the dynamics of a mutation's frequency, based on an idealized version of a serial dilution experiment. However, there are many other microscopic models we could have considered.

Experimentally motivated models. Some are experimentally motivated like Chapter 4. An important example is the *chemostat*, which is a device used for maintaining continuously growing cells at a fixed growth rate.

E.g. "Chemostat"

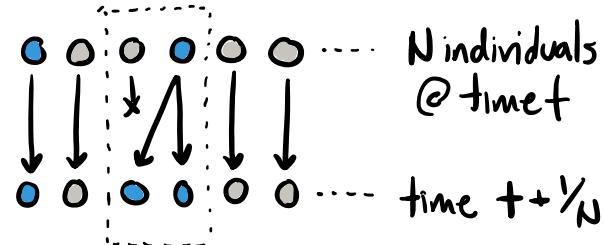


In this model, nutrients are continuously supplied to the (well-mixed) culture vessel at a fixed rate, and cells and media are removed (i.e. diluted out) at a fixed rate δ (measured in volume/sec). The bacteria will therefore grow until they reach a steady state size \bar{N} , where the growth rate of the bacteria is exactly balanced by δ .¹ The chemostat is appealing from both a theoretical and physiological standpoint, because unlike our serial dilution model, there is no temporal variation in growth throughout the day. Despite these advantages, chemostats can be tricky to set up in practice, and this has tended to limit the length of time and number of replicate populations that can be evolved in this way.

Mathematically motivated models. Other microscopic models are defined purely mathematically. Many of the most commonly studied models in population genetics fall into this latter category. These are sometimes referred to as *ball-and-urn models* or *bean bag genetics*, since they attempt to abstract away most of the underlying biology. Some notable examples include the *Moran model*:

¹You can explore the dynamics of this model in more detail in Problem 7 on Homework 1.

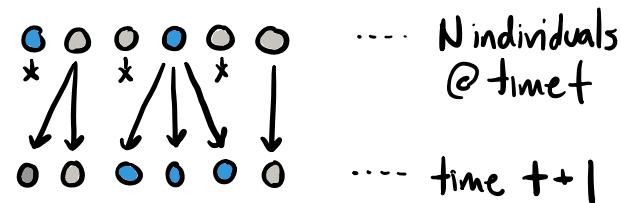
E.g. "Moran model"



- ① pick 2 @ random to "compete"
- ② one replaces other w/ prob $\propto e^s$

where individuals replace each other one at a time. Another common example is the Wright-Fisher model,

E.g. "Wright-Fisher model"



- ① parent of each new cell
selected w/ weight $\propto e^{s_i}$

which is similar to the Moran model, except that all of the individuals in the new generation are chosen at the same time. Both of these models differ from the chemostat and serial dilution examples above in that the total population size is fixed. For a population with a single mutant type, the Wright-Fisher dynamics yield a binomial update rule,

$$f(t+1) \sim \frac{1}{N} \cdot \text{Binomial} \left(N, \frac{f(t)e^s}{1 - f(t) + f(t)e^s} \right), \quad (\text{s.i})$$

which is similar — but not identical to — the Poisson update rule in our serial dilution model in Chapter 4.

One of the reasons that the Wright-Fisher model is so popular is that Eq. (??) admits an exact result for the mean and variance of $f(t)$ in the absence of natural selection ($s = 0$). Similar to the serial dilution model, we find that

$$\mathbb{E}[f(t)] = \mathbb{E}[f(t-1)] = \dots = f_0 \quad (5.2)$$

Using the following property of the Binomial(N, p) distribution:

$$\begin{aligned} \mathbb{E}[n(N-n)] &= \sum_{n=0}^N n(N-n) \binom{N}{n} p^n (1-p)^{N-n} \\ &= N(N-1)p(1-p) \sum_{n=1}^{N-1} \binom{N-2}{n-1} p^{n-1} (1-p)^{N-n-1} \\ &= N(N-1)p(1-p), \end{aligned} \quad (5.3)$$

we can derive a similar recursion for the so-called *heterozygosity*:

$$\mathbb{E}[f(t)(1-f(t))] = \left(1 - \frac{1}{N}\right) \mathbb{E}[f(t-1)(1-f(t-1))], \quad (5.4)$$

which implies that

$$E[f(t)(1-f(t))] \approx f(0)(1-f(0))e^{-t/N} \quad (5.5)$$

The heterozygosity decays as $t \rightarrow \infty$, which makes sense because we know that the mutation must eventually either fix ($f = 1$) or go extinct ($f = 0$). The timescale of this process ($t \sim N$) is consistent with the heuristic argument we made in Chapter 4.

5.2 Emergence of Universality and the Diffusion Limit

Equations (5.4) and (5.5) are about it as far as exact results go, even for such ridiculously simple “bean bag genetics” models. This is a sobering thought — if the

simplest models are this hard, how could we hope to make progress for anything remotely resembling a real biological organism (e.g. influenza)?

At the same time, it might come as a surprise to learn that the field of population genetics routinely applies versions of these simple models (particularly the Wright-Fisher example) to genomic data from real biological populations (e.g. humans) — and they often do a surprisingly good job. Why does this work at all? After all, it's pretty clear that humans are definitely not reproducing according to the Wright-Fisher diagram above.

In the following sections, we'll start to get a partial answer to both of these questions by introducing the *diffusion limit of population genetics*. This is one of my favorite results in classical population genetics, and has some deep connections to the concepts of *universality*, *coarse-graining*, and the *renormalization group (RG)* in physics. Along the way, we will also develop the mathematical concept of a *stochastic differential equation*, which will be an important theoretical tool that we'll use throughout the rest of the course. To do so, it will be helpful to first take a brief detour from our evolutionary applications, and revisit the classical mathematical problem of a discrete random walk.

5.2.1 Detour: ordinary random walks

Let's start by considering the simpler problem of a *discrete-time random walk*. Let $x(t)$ denote the position of a particle at time $t = 0, 1, 2, \dots$, etc. In each timestep, the position of the particle is incremented by an independent random variable Δx_t , which is Gaussian distributed with mean zero and constant variance:

$$x(t+1) = x(t) + \Delta x_t, \quad \Delta x_t \stackrel{i.i.d.}{\sim} \text{Gaussian}(0, \sigma^2). \quad (5.6)$$

Starting from a given value of $x(0)$, this recursion generates a random sequence of positions, $x(0) \rightarrow x(1) \rightarrow \dots x(t)$, similar to our mutation frequency model in Chapter 4. Using the fact that sums of independent Gaussian random

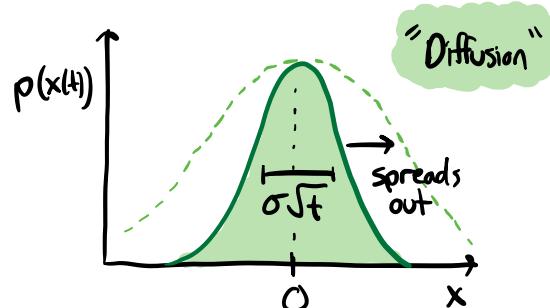
variables are also Gaussian,

$$\sum_i \text{Gaussian}(\mu_i, \sigma_i) \sim \text{Gaussian}\left(\sum_i \mu_i, \sum_i \sigma_i^2\right), \quad (5.7)$$

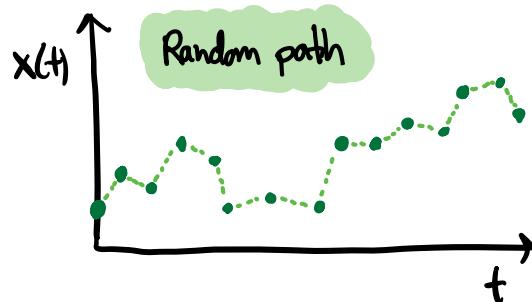
we can conclude that the position of the particle is also a Gaussian

$$x(t) = \Delta x_0 + \Delta x_1 + \dots + \Delta x_{t-1} \sim \text{Gaussian}(0, \sigma^2 \cdot t) \quad (5.8)$$

with a width that grows $\propto \sqrt{t}$. These dynamics are known as **Brownian motion** or **diffusion**; in this course, we will use the phrase **physical diffusion** so that we can distinguish it from the population genetic version we will eventually introduce below.



In addition to the marginal distribution of $x(t)$, we can also write down the probability of an entire path $x(0) \rightarrow x(1) \rightarrow \dots \rightarrow x(t)$,



by noting that the increments in successive timepoints are all independent Gaussians:

$$p(x(1), \dots, x(t)|x(0)) = \prod_{i=1}^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x(i)-x(i-1))^2}{2\sigma^2}} \quad (5.9)$$

We can think of this as a simple form of a *path integral*.

What happens if the increments are not Gaussian distributed? Suppose that we generalize the model slightly, so that

$$\Delta x_t \stackrel{i.i.d.}{\sim} p(\Delta x) \quad \text{w/ } \langle \Delta x \rangle = \mu, \quad \text{Var}(\Delta x) = \sigma^2, \quad (5.10)$$

for some more general distribution $p(x)$. From the central limit theorem (Eq. 2.37 in Chapter 2), we can conclude that for a broad range of distributions, $x(t)$ will approach the same Gaussian form as in Eq. (5.8) when the number of timesteps t is sufficiently large²:

$$x(t) \approx \Delta x_0 + \Delta x_1 + \dots + \Delta x_{t-1} \approx \text{Gaussian}(\mu \cdot t, \sigma^2 \cdot t) \quad (5.11)$$

This is a textbook application of the CLT. However, a fact that is less commonly emphasized for sums like Eq. (5.11) is that the CLT can also apply *locally* for sub-intervals of length $\delta t \gg 1$:

$$x(t) = \underbrace{\Delta x_0 + \dots + \Delta x_{\delta t-1}}_{\approx \text{Gaussian}(\mu \cdot \delta t, \sigma^2 \cdot \delta t)} + \dots + \underbrace{\Delta x_{t-\delta t} + \dots + \Delta x_{t-1}}_{\approx \text{Gaussian}(\mu \cdot \delta t, \sigma^2 \cdot \delta t)}. \quad (5.12)$$

Thus, if we *coarse-grain* over some intermediate timescale δt with $\delta t \gg 1$ but $\delta t \ll t$, we can rewrite our recursion as

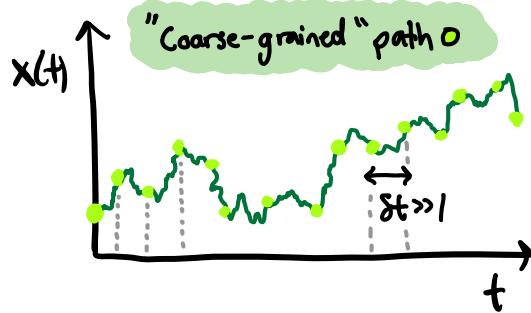
$$x(t + \delta t) \approx x(t) + \text{Gaussian}(\mu \cdot \delta t, \sigma^2 \cdot \delta t), \quad (5.13)$$

²How large will depend on the details of $p(x)$, as we will explore in Problem 2 of Homework 1

or alternatively,

$$x(t + \delta t) = x(t) + \underbrace{\mu \cdot \delta t + \sqrt{\sigma^2 \delta t} \cdot Z_t}_{\delta x(t)}, \quad (5.14)$$

where the Z_t are independent standard Gaussians with $\langle Z_t \rangle = 0$ and $\langle Z_t^2 \rangle = 1$. This generates a sequence of positions $x(0) \rightarrow x(\delta t) \rightarrow x(2 \cdot \delta t) \rightarrow \dots x(t)$ that is similar to the original model in Eq. (5.10), but evaluated at only a *subset* of the original timepoints.



Notation: It is common to re-express the update rule in Eq. (5.14) using the notation of a *stochastic differential equation (SDE)*,

$$\frac{\partial x}{\partial t} = \underbrace{\mu}_{\text{deterministic part}} + \underbrace{\sqrt{\sigma^2} \cdot \eta(t)}_{\text{stochastic part}} \quad (5.15)$$

(also known as a *Langevin equation*). The interpretation of Eq. (5.15) is that the first term represents the deterministic contribution (i.e. the behavior of $x(t)$ in the absence of noise) while the second term represents the stochastic contribution due to the *Brownian noise term* $\eta(t)$. SDEs have subtle mathematical properties if you take them too seriously.³ In this course, we will treat

³The difficulties rapidly become apparent if we try to calculate the “derivative” $\delta f / \delta t$ using the recursion in Eq. (5.14). The $\sqrt{\delta t}$ scaling of the Z_t term means that it will diverge when we divide by $\delta t \rightarrow 0$ (i.e., $x(t)$ is not even differentiable in the traditional sense). These difficulties are surmountable if we’re willing to generalize our definition of a derivative; see Oksendal’s *Stochastic differential equations: an introduction with applications* for more details.

the SDE notation in Eq. (5.15) solely as a notational shorthand for the series expansion in Eq. (5.14). This concrete definition will be sufficient to derive all the results we'll need here.

Universality. Using our recursion in Eq. (5.14), we can write down a formula for the probability of an arbitrary coarse-grained path $x(0) \rightarrow x(\delta t) \rightarrow \dots x(t)$:

$$p(x(\delta t), \dots, x(t)|x(0)) \approx \prod_{i=1}^{t/\delta t} \frac{1}{\sqrt{2\pi\sigma^2\delta t}} e^{-\frac{(x(i\delta t)-x(i\delta t-\delta t)-\mu\delta t)^2}{2\sigma^2\delta t}} \quad (5.16)$$

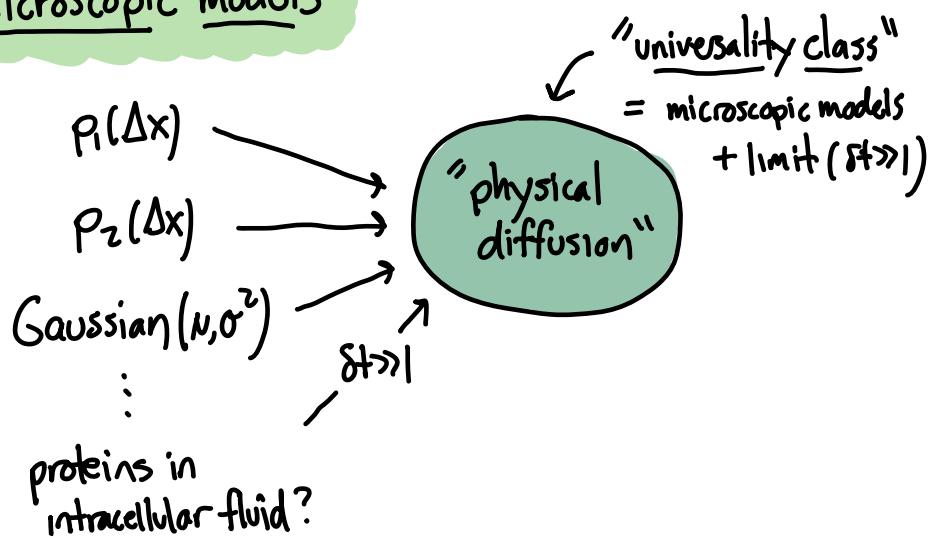
We can also write this in differential notation:

$$\begin{aligned} p(x(\delta t), \dots, x(t)|x(0)) &\approx (2\pi\sigma^2\delta t)^{-\frac{t}{2\delta t}} \cdot e^{-\sum_{i=1}^{t/\delta t} \frac{\left(\frac{x(i\delta t)-x(i\delta t-\delta t)}{\delta t}-\mu\right)^2}{2\sigma^2}\delta t} \\ &\approx \mathcal{D}(x) \cdot e^{-\int_0^t \frac{\left(\frac{\partial x}{\partial t}-\mu\right)^2}{2\sigma^2}} \end{aligned} \quad (5.17)$$

which yields a path integral with an effective Lagrangian $\mathcal{L}(\dot{x}, x) = \frac{(\dot{x}-\mu)^2}{2\sigma^2}$.

If we focus only on the coarse-grained timepoints ($t = \delta t, 2\delta t, \dots$), the probability of the coarse-grained path in Eq. (5.16) is identical to the Gaussian model in Eq. (5.9) — even when the underlying distribution $p(\Delta x)$ was not Gaussian. This shows that for a large class of $p(\Delta x)$, random walks have similar statistical properties when viewed over sufficiently long timescales ($t \gg 1$ and $\delta t \gg 1$). We can represent this with a diagram,

Microscopic models



where many different microscopic models correspond to the same **universality class** (defined to be a collection of microscopic models + a particular limit). This implies that we can use any of the microscopic models to predict the behavior of any of others in the same class on timescales $\delta t \gg 1$ — a good choice is to use the model that we can actually solve (e.g. the Gaussian model in this case).

Note that “universal” is a slight misnomer here. It’s not that the paths from two distributions $p_1(\Delta x)$ and $p_2(\Delta x)$ are truly *identical* — they’re only similar in the limit that δt is sufficiently large. In fact, we expect to be able to tell them apart if we look at sufficiently short timescales (i.e. $\delta t \sim 1$, or when the CLT no longer applies).

This concept of universality plays a central role in modern physics.⁴ However, it is worth appreciating how different it is from some of the laws of physics that are taught in introductory physics classes. A shift from an inverse-squared law to an inverse-cubed law in gravity or electrostatics famously leads to qual-

⁴for a broader review, see H.E. Stanley, “Scaling, universality, and renormalization: Three pillars of modern critical phenomena,” *Rev Mod Phys*, 71(2), S358 (1999).

itatively different physics⁵. In contrast, our present results show that a similar change from $|x|^2 \rightarrow |x|^3$ in the exponent of the Gaussian produces the same random paths at long times. Somewhat paradoxically, the law of diffusion is so important in physics precisely because it is insensitive to these microscopic details — it allows us to make quantitatively accurate predictions for a range of experimental systems using simple toy models like Eq. (5.6), despite their obvious disconnect from reality. If we could show that a similar effect also emerges in evolutionary contexts, it would go a long way towards addressing the questions we posed at the beginning of this section.

5.2.2 Diffusion of mutation frequencies

Physical diffusion comes up a lot in biophysics: molecular timescales are very fast, so coarse-graining over lots of microscopic movements is often a good approximation if we are interested in timescales relevant to proteins, cells, and larger organisms. How might similar concepts apply in the context of evolutionary problems? The basic idea is somewhat similar — this time exploiting the fact that evolutionary phenomena take place on very long timescales (much longer than a single generation). This suggests that diffusion-like behavior might emerge when we coarse-grain over large numbers of generations.

We can start by writing the frequency trajectory from our serial dilution model as a sum over the changes that occur during each daily cycle:

$$\Delta f(t) \equiv f(t + \Delta t) - f(t) = \frac{N_2}{N_1 + N_2} - f(t) \quad (5.18)$$

⁵see J Dorling, “Henry Cavendish’s deduction of the electrostatic inverse square law from the result of a single experiment,” *Studies in History and Philosophy of Science*, Part A 4.4 (1974): 327-348.

and N_1 and N_2 defined as in Chapter 4,

$$N_2 \sim \text{Poisson} \left(\bar{N}_0 \cdot \frac{f(t)e^{s\Delta t}}{1 - f(t) + f(t)e^{s\Delta t}} \right) \quad (5.19)$$

$$N_1 \sim \text{Poisson} \left(\bar{N}_0 \cdot \frac{1 - f(t)}{1 - f(t) + f(t)e^{s\Delta t}} \right) \quad (5.20)$$

We can then write

$$f(t) = f(0) + \Delta f(0) + \Delta f(\Delta t) + \dots + \Delta f(t - \Delta t) \quad (5.21)$$

The primary difference from the simple random walk in Eq. (5.10) is that the mean and variance of $\Delta f(t)$ now depend on the current value of $f(t)$. This means that if we define the coarse-grained increment,

$$\delta f(t) \equiv \Delta f(t) + \Delta f(t + \Delta t) + \dots + \Delta f(t + \delta t - \Delta t), \quad (5.22)$$

we cannot directly apply the central limit theorem from Chapter 2...