

Chapter 5

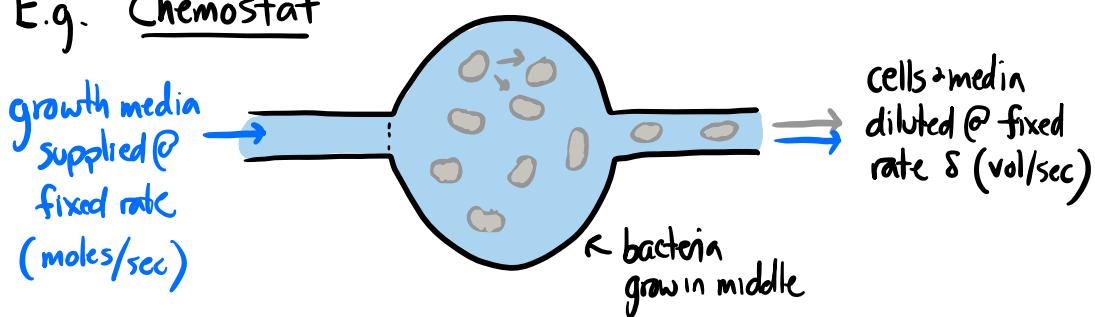
Microscopic Models and the Diffusion Limit

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"

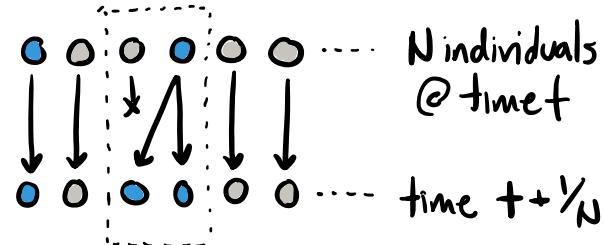


In this model, nutrients are continuously supplied to the (well-mixed) culture vessel at a fixed rate, and cells and media are removed (i.e. diluted out) at a fixed rate δ (measured in volume/sec). The bacteria will therefore grow until they reach a steady state size \bar{N} , where the growth rate of 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.

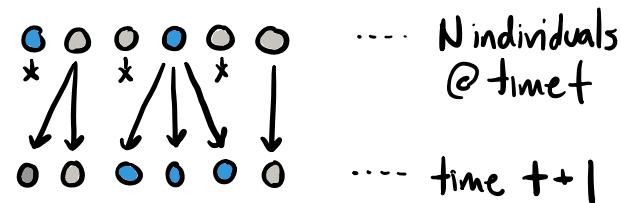
E.g. "Moran model"



- 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. (5.1) 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 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

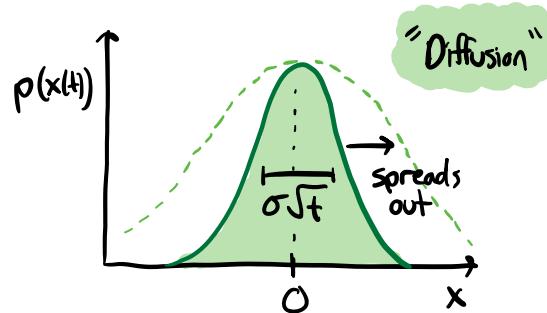
variables are also Gaussian,

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

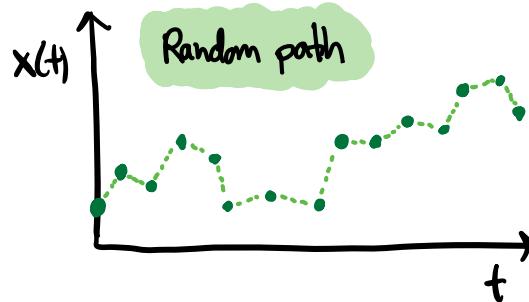
we can conclude that the position of the particle is also a Gaussian

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

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 or **spatial 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

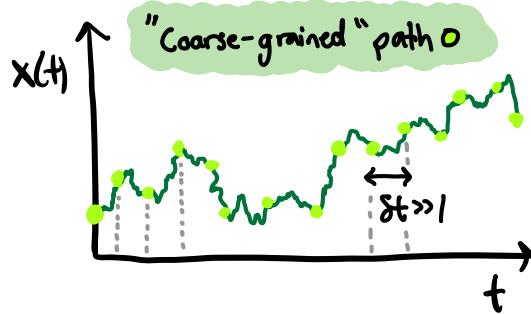
$$x(t + \delta t) \approx x(t) + \text{Gaussian}(\mu \cdot \delta t, \sigma^2 \cdot \delta t), \quad (5.13)$$

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* term $\eta(t)$. SDEs have subtle mathematical properties if you take them too seriously.³. In this course, we will treat

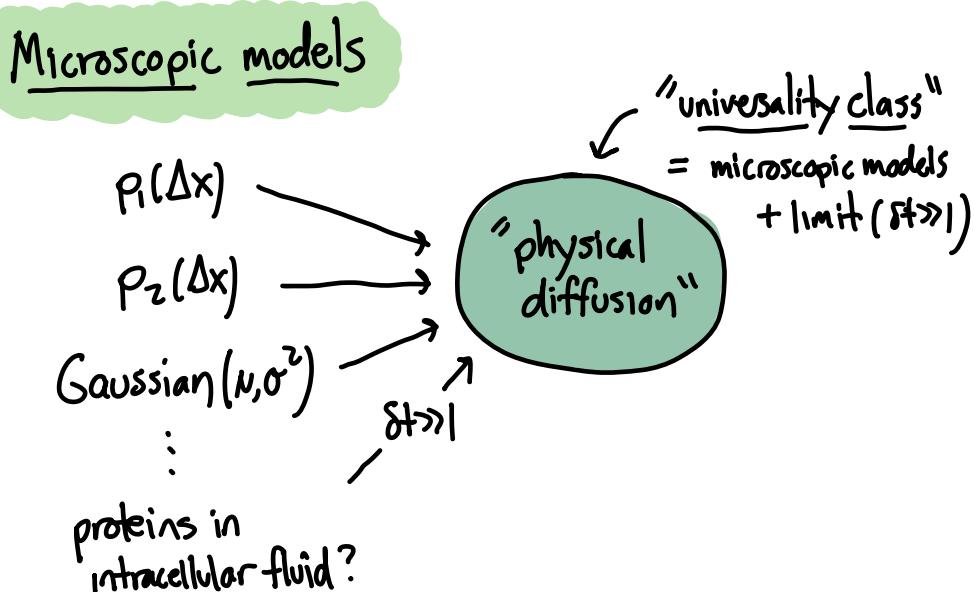
³The difficulties rapidly become apparent if we try to calculate the “derivative” $\delta f / \delta t$ using the recursion in Eq. (5.14). The $\sqrt{\delta t}$ scaling of the Z_t term means that it will diverge when we divide by $\delta t \rightarrow 0$ (i.e., $x(t)$ is not even differentiable in the 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 $x(0) \rightarrow x(\delta t) \rightarrow \dots x(t)$:

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

If we focus only on these coarse-grained timepoints ($t = \delta t, 2\delta t, \dots$), then the probability of the 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 jump distributions $p(\Delta x)$, the resulting 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,



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 the coarse-graining timescale 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).

The concept of universality plays a central role in modern physics.⁴ However, it is worth appreciating how different it is from the more well-known “laws of physics” that are taught in introductory physics classes. For example, a shift from an inverse-squared law of gravity to an inverse-cubed law leads to qualitatively different physics⁵. In contrast, our results above show that a similar-looking change from $|x|^2 \rightarrow |x|^3$ in the exponent of the Gaussian distribution 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

Spatial diffusion comes up a lot in physics. This makes sense: molecular timescales are very fast, so coarse-graining over many microscopic events is often a good

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).

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.

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 some 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:

$$\Delta f_t \equiv f(t + \Delta t) - f(t) = \frac{N_2}{N_1 + N_2} - f(t), \quad (5.17a)$$

where N_1 and N_2 are defined as in Chapter 4,

$$N_2 \sim \text{Poisson} \left(\overline{N}_0 \cdot \frac{f(t)e^{s\Delta t}}{1 - f(t) + f(t)e^{s\Delta t}} \right), \quad (5.17b)$$

$$N_1 \sim \text{Poisson} \left(\overline{N}_0 \cdot \frac{1 - f(t)}{1 - f(t) + f(t)e^{s\Delta t}} \right). \quad (5.17c)$$

We can then write

$$f(t) = f(0) + \Delta f_0 + \Delta f_{\Delta t} + \dots + \Delta f_{t-\Delta t} \quad (5.18)$$

The primary difference from the simple random walk model in Eq. (5.10) is that the mean and variance of Δf_t now depend on the current value of $f(t)$:

$$\langle \Delta f_t \rangle \equiv \mu[f(t)], \quad \text{Var}[\Delta f_t] \equiv \sigma^2(f(t)). \quad (5.19)$$

This means that if we define the coarse-grained increment,

$$\delta f(t) \equiv \Delta f_t + \Delta f_{t+\Delta t} + \dots + \Delta f_{t+\delta t-\Delta t}, \quad (5.20)$$

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

However, as long as the individual increments (Δf_t) are small, we might hope that $f(t)$ would still be a good approximation for $f(t + i\Delta t)$ in the near-term — simply because there has not been enough time for large changes to accumulate. This motivates our basic approach: we wish to find a coarse-grained time interval δt that is simultaneously (i) large enough that the central limit theorem could apply to the net change within a sub-interval, but (ii) small enough that the mean and variance of the individual increments remain approximately constant:

$$\mu[f(t + \delta t)] \approx \mu[f(t)], \quad \sigma^2[f(t + \delta t)] \approx \sigma^2[f(t)] \quad (5.21)$$

We could then apply the central limit theorem within an interval to obtain,

$$\delta f(t) \approx \text{Gaussian}(\mu[f(t)] \cdot \delta t, \sigma^2[f(t)] \cdot \delta t), \quad (5.22)$$

or

$$f(t + \delta t) = f(t) + \mu[f(t)] \cdot \delta t + \sqrt{\sigma^2[f(t)] \cdot \delta t} \cdot Z_t, \quad (5.23)$$

which is a simple generalization of the classical random walk model in Eq. (5.14). Our goal now is to understand whether (and under what conditions) this idea actually works.

Our basic approach will rely on the self-consistency arguments and series expansions that we introduced in Chapter 2. We can break this argument into 4 key steps. We will illustrate each step with the serial dilution model in Eq. (5.17), but the same approach can be used for other microscopic models as well.⁶

Step 1 (identify the relevant limits): Our argument relies on the individual increments (Δf_t) being small, so we must first identify the parameter regimes where this will be a good approximation. For the serial dilution model in Eq. (5.17), we see that Δf_t will be small if the following two conditions are met:

⁶We will work through another example in Problem 2 of Homework 2.

- (i) The means of the Poisson distributions defining N_2 and N_1 must be close to $\bar{N}_0 f(t)$ and $\bar{N}_0(1 - f(t))$, respectively. This will be true if the fitness differences within a single cycle are sufficiently small:

$$s\Delta t \ll 1. \quad (5.24)$$

- (ii) The random draws from these Poisson distributions must be close to their means (i.e., the random variables must be of the “Case 1” form from Section 2.2.1). For a Poisson distribution, this will be true if the means themselves are sufficiently large:

$$\bar{N}_0 f(t) \gg 1, \quad \bar{N}_0(1 - f(t)) \gg 1, \quad (5.25)$$

which implies that a large number of cells of each type are expected to survive the dilution step.⁷ It is often convenient to combine these conditions into a single expression:

$$\bar{N}_0 f(t)[1 - f(t)] \gg 1, \quad (5.26)$$

which will be valid in the same limits.

When the conditions in Eqs. (5.24) and (5.26) are met, the random value of $N_2/(N_2 + N_1)$ will be close to the original frequency $f(t)$, and the increment Δf_t will be small.

Step 2 (calculate the leading-order behavior for 1 cycle): Using the limits we identified above, we can calculate the leading-order contributions to $\mu[f(t)]$ and $\sigma^2[f(t)]$ for a single dilution cycle. The algebra becomes somewhat tedious, but the basic idea follows from the series expansions discussed in Chapter 2.

⁷If you’re worried about what happens when the mutant or wildtype is represented by just a few cells — you should be! We will talk about this later when we discuss low-frequency dynamics and asymptotic matching.

When $s\Delta t \ll 1$, we can Taylor expand the argument of the Poisson distribution to obtain:

$$\begin{aligned} \frac{fe^{s\Delta t}}{1 - f + fe^{s\Delta t}} &\approx \frac{f[1 + s\Delta t + \dots]}{1 - f + f[1 + s\Delta t + \dots]} \\ &\approx f + \underbrace{s\Delta t f(1 - f)}_{\text{leading-order contribution}} + \text{h.o.t.} \end{aligned} \quad (5.27a)$$

where we are using f as a shorthand for $f(t)$. A similar expansion for the wild-type frequency yields

$$\frac{1 - f}{1 - f + fe^{s\Delta t}} \approx 1 - f + \underbrace{-s\Delta t f(1 - f)}_{\text{leading-order contribution}} + \text{h.o.t.} \quad (5.27b)$$

Similarly, when $\bar{N}_0 f$ and $\bar{N}_0(1 - f)$ are large, we can use the Gaussian approximation to the Poisson distribution to write:

$$N_2 \approx \bar{N}_0 \underbrace{[f + s\Delta t f(1 - f)]}_{\text{leading-order contributions}} + \sqrt{\bar{N}_0 f} Z_2 + \text{h.o.t.} \quad (5.28a)$$

and

$$N_1 \approx \bar{N}_0 \underbrace{[1 - f - s\Delta t f(1 - f)]}_{\text{leading-order contributions}} + \sqrt{\bar{N}_0(1 - f)} Z_1 + \text{h.o.t.} \quad (5.28b)$$

where Z_1 and Z_2 are independent standard Gaussians, and where we have again kept only the leading-order contributions in the (joint) limit that $s\Delta t$ is small and $\bar{N}_0 f(t)$ and $\bar{N}_0(1 - f(t))$ are large. We can add these expressions to calcu-

late the total population size at the beginning of the next cycle:

$$\begin{aligned} N_1 + N_2 &\approx \bar{N}_0[1 - f - s\Delta t f(1 - f)] + \sqrt{\bar{N}_0(1 - f)}Z_1 \\ &\quad + [f + s\Delta t f(1 - f)] + \sqrt{\bar{N}_0 f}Z_2 \\ &\approx \bar{N}_0 + \sqrt{\bar{N}_0(1 - f)}Z_1 + \sqrt{\bar{N}_0 f}Z_2 \end{aligned} \quad (5.29)$$

which is independent of s as expected.⁸ We can then take the ratio between N_2 and $N_1 + N_2$ to calculate the frequency of the mutant at the beginning of the next cycle:

$$\begin{aligned} f(t + \Delta t) &\approx \frac{f + s\Delta t f(1 - f) + \sqrt{\frac{f}{\bar{N}_0}}Z_2 + \dots}{1 + \sqrt{\frac{1-f}{\bar{N}_0}}Z_1 + \sqrt{\frac{f}{\bar{N}_0}}Z_2 + \dots} \\ &\approx f(t) + s\Delta t f(1 - f) + \sqrt{\frac{f}{\bar{N}_0}}Z_2 - f\sqrt{\frac{1-f}{\bar{N}_0}}Z_1 - f\sqrt{\frac{f}{\bar{N}_0}}Z_2 + \dots \\ &\approx f(t) + s\Delta t f(1 - f) + \sqrt{\frac{(1-f)^2 f}{\bar{N}_0}}Z_2 - \sqrt{\frac{f^2(1-f)}{\bar{N}_0}}Z_1 + \dots, \end{aligned} \quad (5.30)$$

We can use the Gaussian sum rule in Eq. (5.7) to combine the last 2 terms into a single random variable,

$$\sqrt{\frac{(1-f)^2 f}{\bar{N}_0}}Z_2 - \sqrt{\frac{f^2(1-f)}{\bar{N}_0}}Z_1 = \sqrt{\frac{f(1-f)}{\bar{N}_0}}Z_t, \quad (5.31)$$

where Z_t is a new standard Gaussian. This yields an approximate expression for the frequency change over a single cycle

$$\Delta f_t \approx \underbrace{s\Delta t \cdot f(t)[1 - f(t)]}_{\mu[f(t)]} + \sqrt{\underbrace{\frac{f(t)[1 - f(t)]}{\bar{N}_0}}_{\sigma^2[f(t)]}} \cdot Z_t \quad (5.32)$$

⁸In our serial dilution model, selection can only change the relative frequencies of the strains, since the bottleneck size is fixed by the experimenter.

that is valid in the limit that $s\Delta t \ll 1$ and $\bar{N}_0 f(t)[1 - f(t)] \gg 1$.

Step 2 (add up contributions over multiple cycles): If we assume that $f(t + k\Delta t) \approx f(t)$ within a coarse-grained time interval, we can add up the contributions from the individual daily cycles in Eq. (5.32) to obtain:

$$\begin{aligned} f(t + \delta t) &= f(t) + \Delta f_t + \dots + \Delta f_{t+\delta t-\Delta t} \\ &\approx f(t) + s f(t)[1 - f(t)]\delta t + \sqrt{\frac{f(t)[1 - f(t)]\delta t}{\bar{N}_0 \cdot \Delta t}} \cdot Z_t \end{aligned} \quad (5.33)$$

At this point, it is common to define a set of *effective parameters*

$$N_e \equiv \bar{N}_0 \cdot \Delta t, \quad s_e \equiv s, \quad (5.34)$$

such that Eq. (5.33) can be written in the standard form,

$$f(t + \delta t) = f(t) + \underbrace{s_e f(t)[1 - f(t)]}_{\text{natural selection}} + \underbrace{\sqrt{\frac{f(t)[1 - f(t)]\delta t}{N_e}} \cdot Z_t}_{\text{genetic drift}}, \quad (5.35)$$

which decomposes the change in frequency into a deterministic portion from natural selection and a stochastic component due to genetic drift. Using the SDE notation in Eq. (5.15), this can be equivalently written as

$$\frac{\partial f}{\partial t} = \underbrace{s_e f(1 - f)}_{\text{natural selection}} + \underbrace{\sqrt{\frac{f(1 - f)}{N_e}} \cdot \eta(t)}_{\text{genetic drift}}. \quad (5.36)$$

Step 4 (check self-consistency): Finally, we have to check that the various assumptions we made along the way are self-consistent with each other.

To verify that the means and variances of the Δf_t are constant within a coarse-grained time interval, we can check that this holds at the two endpoints:

$$\mu[f(t + \delta t)] \approx \mu[f(t)], \quad \sigma^2[f(t + \delta t)] \approx \sigma^2[f(t)] \quad (5.37)$$

Using the forms for $\mu[f(t)]$ and $\sigma^2[f(t)]$ derived above, we see that the relative error will be small provided that

$$\delta f \ll f(t), \quad \delta f \ll 1 - f(t) \implies \delta f \ll f(t)[1 - f(t)] \quad (5.38)$$

Using the update rule in Eq. (5.35), we see that this condition will be satisfied if

$$\delta t \ll 1/s, \quad \delta t \ll N_e f(1 - f) \quad (5.39)$$

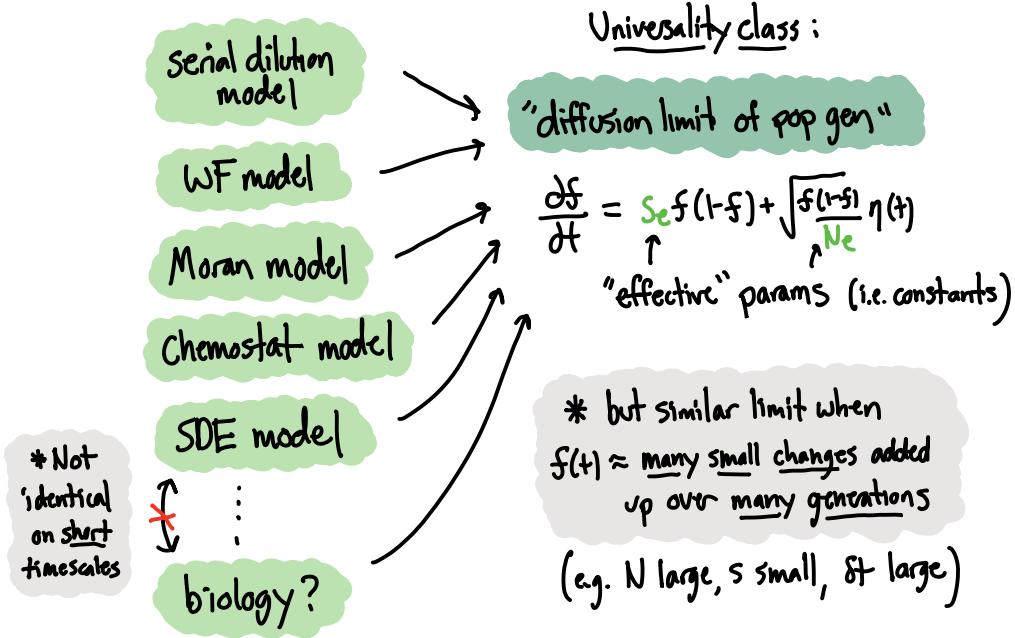
Since the coarse-grained timescale δt must be at least as large as a single cycle, we must also have

$$\delta t \gtrsim \Delta t \quad (5.40)$$

We see that both conditions can be satisfied if the fitness differences are small ($s_e \rightarrow 0$) and the population sizes are large ($N_e \rightarrow \infty$), while their product $N_e \cdot s_e$ can take on any value between $-\infty$ and ∞ . This is known as the **diffusion limit of population genetics**. When these conditions hold, the coarse-grained dynamics of $f(t)$ can be described by the stochastic differential equation Eq. (5.36).

Universality. We demonstrated this result for our simple serial dilution model in Chapter 4. With a similar amount of work, one can show that it holds for a number of other microscopic models, including the Wright-Fisher, Moran, and chemostat models above. This suggests that the diffusion limit captures the behavior of a broader universality class:

Microscopic models



which applies whenever the coarse-grained dynamics can be described by Eq. (5.36) for an appropriate⁹ pair of constants s_e and N_e . This model is sometimes known as the *Wright-Fisher diffusion process* or *Kimura's diffusion model*. (We might also call it the *single-locus diffusion model* to distinguish it from the longer-genomes we will analyze later in the course.) These dynamics constitute a different universality class than the spatial diffusion model in Eq. (5.15), since the mean and variance have a different functional dependence on $f(t)$. Thus, despite their visual similarities, we will see that mutation trajectories will behave in fundamentally different ways than the trajectories of diffusing particles. Understanding these differences will be a central focus of the next few lectures.

Similar to the random walk example above, the microscopic models that fall

⁹Note that these effective parameters can have a complex relationship with the actual population sizes and fitness benefits of the underlying microscopic model. Equation (5.34) provides the mapping for our simple serial dilution model, but this connection must generally be determined on a case-by-case basis by repeating the calculation above. In practice, since the microscopic parameters are usually unknown, the effective parameters are often treated as an empirical fitting parameter.

into the Wright-Fisher diffusion class are *not* equivalent to each other in general; they are only equivalent when we coarse-grain over sufficiently long timescales and large numbers of individuals. The Wright-Fisher diffusion is not the only universality class that arises in this limit, but it is a common one (and arguably the simplest). This provides a partial explanation for why toy models like the Wright-Fisher model (or their generalizations) can sometimes be useful for describing real data, despite their obvious departures from reality. It does so by changing the question from “*is the Wright-Fisher model an accurate representation of the real population?*” to “*does the real population fall into the same universality class as the Wright-Fisher model?*” Our coarse-graining derivation also shows that there are inherent limitations to this approach: we naturally expect the diffusion approximation to break down on short timescales, when the details of the underlying birth-and-death process start to become important.

Who is approximating whom? This last observation raises an important point. In the population genetics literature, the diffusion model in Eq. (5.36) is sometimes framed as an *approximation* to the Wright-Fisher model, and that “exact” results for the latter would be more desirable. Our universality diagram above illustrates how this might not provide many practical benefits when it comes to making connections with data. In the regimes where the differences between the Wright-Fisher model and the diffusion process are important, it is likely that the differences between the Wright-Fisher model and the biology will be equally important as well. In these cases, a more detailed model of the specific population will often be preferable to an exact solution of the Wright-Fisher model. This is an important consideration whenever we wish to study aspects of the biology that deviate from the diffusion limit.

5.2.3 Traditional derivation of the Wright-Fisher diffusion

The derivation above is somewhat different from the traditional derivation of the diffusion limit in population genetics. I think the coarse-graining version places a greater emphasis on the underlying physical assumptions, and it is often

easier to extend to more complex scenarios (as we'll see on Homework 2). Nevertheless, it is still worth presenting the traditional derivation for completeness, since it is often referenced in the literature.

The standard derivation starts from a general continuous-state Markov process,

$$x_0 \rightarrow x_1 \rightarrow \dots \rightarrow x_t, \quad (5.41)$$

characterized by the single step transition probability:

$$\Pr[x_i \rightarrow x_{i+1}] = p_1(x_{i+1}|x_i). \quad (5.42)$$

We can then consider the probability density, $p(x, t|x_0)$, which represents the probability of observing the system in position x at time t , given that it started at position x_0 at $t = 0$. By considering all the positions in the previous timestep, we can derive a recursive formula for this probability distribution,

$$p(x, t + 1|x_0) = \int p_1(x|x')p(x', t|x_0) dx' \quad (5.43)$$

$$= \int p_1(x|x - \Delta x)p(x - \Delta, t|x_0) d\Delta x, \quad (5.44)$$

where we have changed variables in the last line from the previous position (x') to the jump size ($\Delta x = x - x'$). If we expand this expression to first order in the time increment and second order in the jump size¹⁰, we can obtain a partial differential equation for the probability density,

$$\frac{\partial p(x, t|x_0)}{\partial t} = -\frac{\partial}{\partial x} [\mu(x) \cdot p(x, t|x_0)] + \frac{\partial^2}{\partial x^2} \left[\frac{\sigma^2(x)}{2} \cdot p(x, t|x_0) \right] \quad (5.45)$$

¹⁰The details are somewhat technical (and not particularly enlightening). They are presented in the appendix at the end of this chapter.

where the functions $\mu(x)$ and $\sigma^2(x)$ are defined by

$$\begin{aligned}\mu(x) &\equiv \int \Delta x \cdot p_1(x + \Delta x | x) d\Delta x \\ \sigma^2(x) &\equiv \int \Delta x^2 \cdot p_1(x + \Delta x | x) d\Delta x\end{aligned}\tag{5.46}$$

This is often known as the **Fokker-Planck equation** (or the **forward equation**), and it is equivalent to the Langevin equation:

$$\frac{\partial x}{\partial t} = \mu(x) + \sqrt{\sigma^2(x)} \cdot \eta(t)\tag{5.47}$$

or the recursive update rule:

$$x(t + \delta t) = x(t) + \mu(x)\delta t + \sqrt{\sigma^2(x)\delta t} \cdot Z_t\tag{5.48}$$

In the case of the Wright-Fisher model (or our serial dilution model), the mean and variance are given by Eq. (5.32), so the Fokker-Planck equation becomes

$$\frac{\partial p(f, t)}{\partial t} = -\underbrace{\frac{\partial}{\partial f} [sf(1-f)p(f, t)]}_{\text{natural selection}} + \underbrace{\frac{\partial^2}{\partial f^2} \left[\frac{f(1-f)}{2N_e} \cdot p(f, t) \right]}_{\text{genetic drift}}\tag{5.49}$$

This is sometimes known as **Kimura's equation**. It is equivalent to the Langevin equation in Eq. (5.36) or the recursive update rule in Eq. (5.35).

5.2.4 Incorporating spontaneous mutations

It is straightforward to repeat the derivation in Section 5.2.2 to allow for spontaneous mutations. We leave it as an exercise to show that the resulting SDE is given by

$$\frac{\partial f}{\partial t} = \underbrace{sf(1-f)}_{\text{selection}} + \underbrace{\mu(1-f) - \nu f}_{\text{mutation}} + \underbrace{\sqrt{\frac{f(1-f)}{N}} \cdot \eta(t)}_{\text{genetic drift}}\tag{5.50}$$

to lowest order in s, μ, ν , and $1/N$.

5.3 Appendix

5.3.1 Traditional derivation of the Fokker-Planck equation

To derive the Fokker-Planck equation from Eq. (5.44), it is useful to consider the generating function of the probability density,

$$H(z) \equiv \int e^{-zx} p(x, t|x_0) dx \quad (5.51)$$

By integrating both sides of our recursion relation, we have

$$\begin{aligned} H(z, t+1) &= \int dx d\Delta x e^{-zx} p(x - \Delta x, t) p_1(x, x - \Delta x) \\ (\text{define } y = x - \Delta x) \quad &= \int dy d\Delta x e^{-zy} e^{-z\Delta x} p(y, t) p_1(y + \Delta x|y) \\ (\text{relabel } y \rightarrow x) \quad &= \int dy d\Delta x e^{-zy} e^{-z\Delta x} p(y, t) p_1(y + \Delta x|y) \end{aligned} \quad (5.52)$$

We can then Taylor expand the integrand in the limit that Δx is small. The right hand side becomes:

$$\begin{aligned} &\approx \int dx d\Delta x \cdot e^{-xz} \left[1 - z\Delta x + \frac{(z\Delta x)^2}{2} \right] p(x, t) p_1(x + \Delta x|x) \\ &= \int dx e^{-zx} p(x, t) \int d\Delta x \left[1 - z\Delta x + \frac{(z\Delta x)^2}{2} \right] p_1(x + \Delta x|x) \\ &= \int dx e^{-zx} \left[1 - z\mu(x) + \frac{z^2\sigma^2(x)}{2} \right] p(x, t) \end{aligned} \quad (5.53)$$

where we have used the definitions of $\mu(x)$ and $\sigma^2(x)$ in Eq. (5.46). Integrating by parts then yields

$$= \int dx e^{-zx} \left\{ p(x, t) - \frac{\partial}{\partial x} [\mu(x)p(x, t)] + \frac{\partial^2}{\partial x^2} \left[\frac{\sigma^2(x)}{2} p(x, t) \right] \right\} \quad (5.54)$$

Thus, if $p(x, t)$ satisfies,

$$\underbrace{p(x, t+1) - p(x, t)}_{\approx \partial_t p(x, t)} = -\frac{\partial}{\partial x} [\mu(x)p(x, t)] + \frac{\partial^2}{\partial x^2} \left[\frac{\sigma^2(x)}{2} p(x, t) \right], \quad (5.55)$$

then $H(z, t)$ [and therefore $p(x, t)$] will satisfy the original recursion relation in Eq. (5.44).