

# CHAPTER 15

## Analyzing Continuous Stochastic Models—Diffusion in Time and Space

### 15.1 Introduction

In this chapter we present techniques for constructing and analyzing stochastic models with continuous state spaces. All of the models presented in Chapters 13 and 14 assume that the state of the system takes on discrete values. Some processes are more naturally modeled by allowing the state space to be continuous. For example, suppose we want to model the movement of an individual within its home range. There will likely be a great deal of stochasticity in movement, and at each point in time, the individual might be in any of a continuum of possible locations. Consequently, it would be appropriate to use a stochastic model with a continuous state space, representing the different possible spatial locations.

Aside from the fact that some processes are more naturally described by a continuous state space, some discrete models that are difficult to analyze can be well approximated using a stochastic model with a continuous state space. For example, the Wright-Fisher model for allele frequency change in a finite population (section 13.4) is difficult to analyze exactly, but it can be well approximated using a stochastic model with a continuous state space representing the allele frequency. Thus, techniques for analyzing models with a continuous state space often provide an alternative route for the approximate analysis of discrete-state models.

In section 15.2, we begin by describing different ways in which stochastic models with a continuous state space can be constructed. We focus on models referred to as *diffusion models*, which are based on the idea that variables disperse (or *diffuse*) away from their original location or state. In section 15.3, we then develop techniques allowing us to obtain general solutions, stationary distributions, probabilities of absorption, and waiting times for many diffusion models. Finally, in section 15.4 we illustrate how these mathematical techniques can be used to model the deterministic dynamics of populations in spatial settings.

### 15.2 Constructing Diffusion Models

In this section we use examples to illustrate different ways in which stochastic models with a continuous state space can be constructed. Interestingly, these different examples all generate the same type of equation for modeling diffusion.

#### Chapter Goals:

- To develop stochastic models with continuous state spaces based on diffusion
- To obtain general solutions, stationary distributions, probabilities of absorption, and waiting times for diffusion models

#### Chapter Concepts:

- Diffusion model
- Drift coefficient
- Diffusion coefficient
- Transition probability density function
- Brownian motion

The first example begins with a model having a discrete state space and then “takes the limit” as the state space becomes continuous. The second example proceeds directly to the construction of a model assuming a continuous state space. The third example considers how a continuous trait (e.g., height, growth rate) might evolve over time, using a diffusion model to track the long-term dynamics of a trait when mutations are limiting. Finally, the fourth example illustrates how a model having a discrete state space can be approximated using a continuous-state-space model. As with stochastic models having a discrete state space, models with a continuous state space can be formulated in either discrete or continuous time. It is often more natural to work with continuous time, however, and we do this in all of the examples below.

### 15.2.1 Modeling Individual Movement in a Continuous Spatial Habitat (Part 1)

Let us suppose that we are studying the movement of a species of stream-dwelling fish. Our ultimate goal is to understand how decisions that individuals make in terms of their movement either up- or downstream translate into patterns of population abundance at different locations. To begin, we will construct a stochastic model for the location  $x(t)$  of a single fish in discrete time (Figure 15.1) using the techniques from Chapter 14. We call  $X(t)$  the *state variable*, as it describes the state of the system at time  $t$ . Suppose that the stream can be divided into discrete locations that are equally spaced at a distance of  $\Delta x$  and that time proceeds in steps of length  $\Delta t$ .

As in section 13.3, we first specify the transition probabilities between the different spatial locations within a single time step. Let us suppose that the time step  $\Delta t$  is short enough that the fish can only move between two neighboring locations within a single time step (Figure 15.1). By analogy with the birth-death model (14.50), we can write the transition probabilities  $p_{yx}$  from spatial location  $x$  to  $y$  in one time step as

$$p_{yx} = \begin{cases} u & \text{for } y = x + \Delta x & \text{(upstream by one step),} \\ d & \text{for } y = x - \Delta x & \text{(downstream by one step),} \\ 1 - u - d & \text{for } y = x & \text{(no movement),} \\ 0 & \text{for } y \neq x - \Delta x, x, x + \Delta x & \text{(other movements),} \end{cases} \quad (15.1)$$

where  $u$  and  $d$  are the probabilities of moving up- and downstream, respectively. Let us label the site that is farthest downstream as 0, and successive sites by

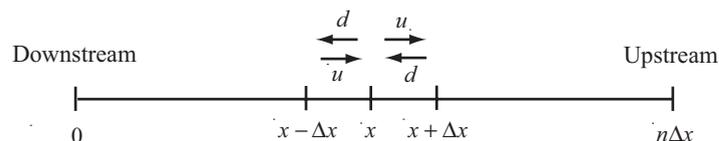


Figure 15.1: A schematic diagram of fish movement within a stream. The diffusion model is obtained by shrinking the distance,  $\Delta x$ , between sites to zero. The probabilities of moving upstream or downstream in a time step,  $\Delta t$ , are given by  $u$  and  $d$ , respectively.

their distance from this initial site as  $\Delta x, 2 \Delta x, \dots, n \Delta x$ , where  $n$  is the total number of sites in addition to the first one. The probability that the fish is in each of the  $n + 1$  sites at time  $t$  is given by the vector

$$\bar{x}(t) = \begin{pmatrix} P(X(t) = 0) \\ P(X(t) = \Delta x) \\ P(X(t) = 2 \Delta x) \\ \vdots \\ P(X(t) = n \Delta x) \end{pmatrix}. \quad (15.2)$$

Changes over time are then described by

$$\bar{x}(t + \Delta t) = \mathbf{M} \bar{x}(t) \quad (15.3)$$

(see equation (14.15)), where  $\mathbf{M}$  is the transition probability matrix whose elements are given by (15.1).

We can proceed to analyze model (15.3) using the techniques for birth-death processes in Chapter 14. For example, if we stop our observations once the fish leaves the region defined by the  $n + 1$  sites (i.e.,  $d = 0$  when  $x = 0$  and  $u = 0$  when  $x = n$ ), then we can use Box 14.1 to calculate the probability that the fish leaves the region upstream rather than downstream (Problem 15.1). Similarly, we could use Box 14.2 to calculate the expected time until the fish leaves the region.

Instead, we will use this model as a starting place for deriving the analogous model with a continuous state space. Carrying out the matrix multiplication in (15.3), the probability of being in a particular site  $x$  at time  $t + \Delta t$  is given by

$$P(X(t + \Delta t) = x) = P(X(t) = x - \Delta x) u + P(X(t) = x) (1 - u - d) + P(X(t) = x + \Delta x) d. \quad (15.4a)$$

To simplify the notation, define  $f(x, t)$  as the probability that the fish is at location  $x$  at time  $t$ ; that is,  $f(x, t) \equiv P(X(t) = x)$ . The above equation can then be written as

$$f(x, t + \Delta t) = f(x - \Delta x, t) u + f(x, t) (1 - u - d) + f(x + \Delta x, t) d. \quad (15.4b)$$

The key to deriving the continuous state space model is to consider what happens as the distance between spatial locations,  $\Delta x$ , gets smaller. As we do this, we must also shrink the time step,  $\Delta t$ . Otherwise the assumption that the fish can move only between neighboring locations in one time step would become unrealistic. Thus we want to consider the limit as both  $\Delta x$  and  $\Delta t$  go to zero. We start by using the Taylor series (Recipe P1.2) to rewrite (15.4b) as a polynomial function of  $\Delta x$ :

$$\begin{aligned} f(x, t + \Delta t) &= \left( f(x, t) - \frac{\partial f(x, t)}{\partial x} \Delta x + \frac{\partial^2 f(x, t)}{\partial x^2} \frac{1}{2} \Delta x^2 + O(\Delta x^3) \right) u \\ &\quad + f(x, t) (1 - u - d) \\ &\quad + \left( f(x, t) + \frac{\partial f(x, t)}{\partial x} \Delta x + \frac{\partial^2 f(x, t)}{\partial x^2} \frac{1}{2} \Delta x^2 + O(\Delta x^3) \right) d. \end{aligned} \quad (15.5a)$$

Gathering together similar terms involving  $\Delta x$ , we get

$$f(x, t + \Delta t) - f(x, t) = -\frac{\partial f(x, t)}{\partial x} (u - d)\Delta x + \frac{\partial^2 f(x, t)}{\partial x^2} \frac{1}{2} (u + d)\Delta x^2 + O(\Delta x^3). \quad (15.5b)$$

Equation (15.5b) gives the change in location within a time unit,  $\Delta t$ . If we divide by  $\Delta t$ , we get the rate of change in location:

$$\frac{f(x, t + \Delta t) - f(x, t)}{\Delta t} = -\frac{\partial f(x, t)}{\partial x} (u - d) \frac{\Delta x}{\Delta t} + \frac{\partial^2 f(x, t)}{\partial x^2} \frac{1}{2} (u + d) \frac{\Delta x^2}{\Delta t} + \frac{O(\Delta x^3)}{\Delta t}. \quad (15.5c)$$

We can now take the limit of (15.5c) as both  $\Delta x$  and  $\Delta t$  go to zero. By the definition of a derivative (Appendix 2), the limit of the left-hand side equals  $\partial f(x, t)/\partial t$ , which is a partial derivative because  $f$  is a function that depends on both location  $x$  and time  $t$ . To take the limit of the right-hand side, however, we must make some assumption about the relative rates at which  $\Delta x$  and  $\Delta t$  go to zero. By assumption, diffusion models choose these relative rates such that changes to the system depend on the first two leading terms in (15.5c),  $\partial f(x, t)/\partial x$  and  $\partial^2 f(x, t)/\partial x^2$ , but not on higher-order terms. Specifically, diffusion models assume that

$$\begin{aligned} \lim_{\Delta x, \Delta t \rightarrow 0} (u - d) \frac{\Delta x}{\Delta t} &= \mu, \\ \lim_{\Delta x, \Delta t \rightarrow 0} (u + d) \frac{\Delta x^2}{\Delta t} &= \sigma^2, \\ \lim_{\Delta x, \Delta t \rightarrow 0} \frac{O(\Delta x^3)}{\Delta t} &= 0, \end{aligned} \quad (15.6)$$

where  $\mu$  and  $\sigma^2$  are finite numbers (with  $\sigma^2 > 0$ ). Using these assumptions, the probability density function  $f(x, t)$  that the fish is in location  $x$  at time  $t$  satisfies

$$\frac{\partial f(x, t)}{\partial t} = -\mu \frac{\partial f(x, t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 f(x, t)}{\partial x^2} \quad (15.7)$$

A *diffusion model* describes changes over time in the state of a system when the state can take on a continuum of values and can change by only a very small amount in a small amount of time.

The *drift coefficient* in a diffusion model describes the expected (or mean) amount of movement per unit time.

The *diffusion coefficient* describes variance in the movement of a system per unit time.

The function  $f(x, t)$  now represents a continuous probability distribution specifying the probability that the fish is in an infinitesimal region of the continuous state space located at point  $x$  (see section P3.4). How this spatial probability distribution changes over time is described by the *partial differential equation* (15.7). Equation (15.7) can be analyzed to make predictions about the location of the fish using the techniques that we will introduce in section 15.3.

Before proceeding to the next example, it is worth making a couple of general remarks about the terms in (15.7). The quantities  $\mu$  and  $\sigma^2$  are referred to as the infinitesimal *mean* and *variance* (or the *drift* and *diffusion* coefficients) of the stochastic process. To appreciate where this terminology comes from, we calculate the mean or expected distance moved,  $E[\Delta X]$ , by the fish in a single

time step of the discrete model (15.1). The fish moves a distance  $\Delta x$  upstream with probability  $u$ , and distance  $-\Delta x$  downstream with probability  $d$ . It remains in the same location with probability  $1 - u - d$ . Therefore, from Definition P3.2, the expected distance moved is  $E[\Delta X] = \Delta x u - \Delta x d + 0(1 - u - d) = (u - d)\Delta x$ . Consequently  $(u - d)\Delta x/\Delta t$  is the expected movement *per unit time* in the discrete model. In the limit of small  $\Delta x$  and  $\Delta t$ , this value becomes  $\mu$  according to (15.6), which gives the expected movement per unit time in the continuous model.

Next, let's calculate the variance in the distance moved in a single time step of the discrete model (15.1):  $\text{Var}[\Delta X] = E[(\Delta X)^2] - E[\Delta X]^2$  (equation (P3.2)). The average displacement squared is  $E[(\Delta x)^2] = (\Delta x)^2 u + (-\Delta x)^2 d + 0(1 - u - d) = (u + d)\Delta x^2$ , and therefore the variance in the movement is  $(u + d)\Delta x^2 - ((u - d)\Delta x)^2$ . If we now divide this by  $\Delta t$ , we obtain the variance in movement per unit time:  $(u + d)(\Delta x^2)/(\Delta t) - (u - d)(\Delta x/\Delta t)\Delta x$ . Taking the limit as both  $\Delta x$  and  $\Delta t$  go to zero and using conditions (15.6) then gives  $\lim_{\Delta x, \Delta t \rightarrow 0} ((u + d)(\Delta x^2)/(\Delta t) - (u - d)(\Delta x/\Delta t)\Delta x) = \sigma^2 - \mu^2$  or simply  $\sigma^2$ . This reveals that  $\sigma^2$  represents the variance in movement per unit time in the continuous model (15.7).

Equation (15.7) is known as a *diffusion equation with drift* and also as the *forward Kolmogorov equation* (or, sometimes, an *advection-diffusion equation*). The term diffusion refers specifically to the random, undirected movement described by the variance in movement per unit time,  $\sigma^2$ . If  $\sigma^2$  were zero, then the system would always move by an amount equal to the expected rate of change,  $\mu$ , and the model would be deterministic. Random, undirected movement is also called *Brownian motion*, named after the Scottish botanist Robert Brown (1773–1858), who studied fertilization in the flower *Clarkia pulchella*. He noticed that pollen grains suspended in water exhibited erratic movements over time. Similar movements are exhibited by inorganic particles and were described mathematically by Albert Einstein in 1905 as a result of collisions between the observed particles and unobserved liquid (or gas) molecules.

Later in this chapter we will use equations such as (15.7) to model allele frequency change in finite populations. This sometimes causes confusion due to an unfortunate conflict in terminology. In genetics, the random fluctuations in allele frequency due to chance events are described as “random genetic drift.” In mathematics, however, “drift” refers to the tendency  $\mu$  to move in a particular direction, in the sense of drifting down a river. As we shall see, in neutral models of random genetic drift, there is no (mathematical) drift, i.e., the allele frequencies do not systematically rise or systematically fall over time and  $\mu = 0$ . Thus, “random genetic drift” would be better called “random genetic diffusion,” but the term is now too widely used to change.

### 15.2.2 Modeling Individual Movement in a Continuous Spatial Habitat (Part 2)

The example above illustrates an important conceptual connection between a discrete-time stochastic model in discrete state space (the birth-death model) and a continuous-time stochastic model in continuous state space (diffusion with drift). It is often the case, however, that stochastic models in continuous

state space are constructed directly rather than deriving them from models with a discrete state space. The next example reexamines the model of fish movement using this approach.

The *transition probability density function* specifies the probability density of the system moving between any two values in any given amount of time.

Let us work directly with the continuous probability distribution that describes the probability of the fish being in an infinitesimally small spatial region at location  $x$  at time  $t$ . Over any period of time, we are interested in knowing the *transition probability density function*,  $\phi(x, t | x_0, t_0)$ , that specifies the probability density of moving to location  $x$  at time  $t$ , given that the fish is at location  $x_0$  at time  $t_0$ .

The next step is to describe how this probability density function changes over time. For generality we allow the movement of the fish to depend on its current location (e.g., perhaps it moves downstream most often when it is very far upstream and vice versa), but we do not allow the movement behavior to depend on time (that is, we focus on *time-homogeneous* models). Although the transition probability density function could, in principle, allow any amount of change in a short period of time, it is difficult to make progress with this level of generality. Consequently, we focus on models for which only small amounts of change can occur in small amounts of time. Under these conditions, the function  $\phi(x, t | x_0, t_0)$  can be shown to satisfy the *forward Kolmogorov equation*:

$$\frac{\partial \phi(x, t | x_0, t_0)}{\partial t} = -\frac{\partial(\mu(x) \phi(x, t | x_0, t_0))}{\partial x} + \frac{1}{2} \frac{\partial^2(\sigma^2(x) \phi(x, t | x_0, t_0))}{\partial x^2}. \quad (15.8)$$

A derivation of equation (15.8) is provided by Allen (2003, pp. 376–377) and by Ewens (1979, pp. 116–117).

Equation (15.8) is a partial differential equation having a similar form to that of equation (15.7). In fact, equation (15.8) is again a model of diffusion with drift, except that we now allow the drift and diffusion coefficients to depend on  $x$  (spatial location in this example). In particular, the drift coefficient  $\mu(x)$  is the expected distance moved by the fish per unit time for the continuous stochastic process, given that it is currently at location  $x$ , defined as

$$\mu(x) = \lim_{\Delta t \rightarrow 0} \frac{E[X(t + \Delta t) - X(t) | X(t) = x]}{\Delta t} \quad (15.9a)$$

Similarly, the diffusion coefficient  $\sigma^2(x)$  is the variance in the distance moved per unit time, given that it is currently at location  $x$ , defined as

$$\sigma^2(x) = \lim_{\Delta t \rightarrow 0} \frac{E[(X(t + \Delta t) - X(t))^2 | X(t) = x]}{\Delta t}. \quad (15.9b)$$

Technically, to get a variance we must subtract the mean squared from (15.9b), but in the limit this term is zero:

$$\lim_{\Delta t \rightarrow 0} \frac{E[X(t + \Delta t) - X(t) | X(t)]^2}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{(\mu(x)\Delta t)^2}{\Delta t} = 0. \quad (15.10)$$

There is, however, a difference in notation between (15.7) and (15.8), which stems from the fact that  $\phi(x, t | x_0, t_0)$  specifies explicitly how the probability density at time  $t$  depends on the state of the system at the previous time  $t_0$ . In

contrast, we used  $f(x,t)$  in (15.7) to describe the probability density at time  $t$  without specifying the initial conditions. Besides being more explicit, the notation  $\phi(x,t|x_0,t_0)$  allows us to derive a second partial differential equation that is extremely useful for determining the behavior of the stochastic process (Box 15.1). This equation is referred to as the *backward Kolmogorov equation*:

$$\frac{\partial \phi(x,t|x_0,t_0)}{\partial t_0} = -\mu(x_0) \frac{\partial \phi(x,t|x_0,t_0)}{\partial x_0} - \frac{1}{2} \sigma^2(x_0) \frac{\partial^2 \phi(x,t|x_0,t_0)}{\partial x_0^2}. \quad (15.11)$$

Comparing (15.8) and (15.11), we see that the derivatives of  $\phi$  are now taken with respect to the initial variables,  $x_0$  and  $t_0$ , rather than with respect to  $x$  and  $t$  as in the forward equation. In section 15.3, we will use both the forward and backward Kolmogorov equations to derive a number of key results from diffusion models, but first we present two more examples illustrating the breadth of problems that can be described by diffusion models.

### 15.2.3 Modeling the Long-Term Evolutionary Dynamics of a Trait when Mutations Are Limiting

In Chapter 12 we introduced evolutionary invasion analysis. At the heart of this technique is an assumption that evolution is mutation limited. Specifically, it is assumed that there is a continuum of different possible alleles, coding for a continuum of different possible trait values that an individual might express. The key assumption is that mutations among these alleles are rare, so that the population reaches an equilibrium while it contains only a single allele. Then a new mutation arises and either replaces the former resident allele or dies out. If it replaces the resident allele, then a new trait value is attained in the population.

In Chapter 12 we inferred the direction of evolution under this process by determining which alleles can invade which populations. But to make predictions about the trait values that are likely to be observed over time, we must construct an explicit model for the stochastic appearance and fixation of mutations altering the trait of interest. A modeling framework referred to as adaptive dynamics has been developed to do exactly this (Dieckmann and Law 1996; Box 12.5), and here we develop an analogous approach based on the diffusion equation.

We begin by defining  $\phi(x,t|x_0,t_0)$  as the continuous probability density function describing the probability that the population is in state  $x$  at time  $t$  (i.e., the allele coding for trait value  $x$  is resident in the population at time  $t$ ) given the initial trait value  $x_0$  at time  $t_0$ . A diffusion model can be used to describe the dynamics of  $\phi(x,t|x_0,t_0)$  if we assume that the resident trait value does not change very much in a small amount of time. This is analogous to the assumption in sections 15.2.1 and 15.2.2 that a fish makes only small movements in small amounts of time. Assuming that large evolutionary jumps do not occur, the probability distribution  $\phi(x,t|x_0,t_0)$  will satisfy the diffusion equations (15.8) and (15.11). The diffusion coefficient  $\sigma^2(x)$  now represents the random, undirected evolutionary change in trait space (owing to mutation),

### Box 15.1: Deriving the Backward Kolmogorov Equation

The Kolmogorov equations (15.8) and (15.11) are differential equations that are satisfied by the probability distribution  $\phi(x, t | x_0, t_0)$  describing the probability density that the system is at position  $x$  at time  $t$ , given that it was at position  $x_0$  at time  $t_0$ . These equations have been derived in a number of ways, but here we focus only on the backward equation (15.11). Throughout this box, we assume without proof that the probability distribution  $\phi$  is a continuous function whose derivatives (to arbitrary order) are finite. (Within this seemingly innocuous statement is hidden a great deal of complexity. Mathematically rigorous derivations of a diffusion equation must demonstrate that this assumption holds.)

Before we begin, we first establish an important equation that  $\phi(x, t | x_0, t_0)$  must satisfy. Consider our example of fish movement, and imagine a fish that starts in location  $x_0$  at time  $t_0$  and ends up at location  $x$  at time  $t$ . We can obtain an expression for  $\phi(x, t | x_0, t_0)$  by considering all of the potential locations,  $x_1$ , of the fish at some intermediate point in time,  $t_1$ . In particular,  $\phi(x, t | x_0, t_0)$  can be expressed as the probability of moving from  $x_0$  to  $x_1$  between times  $t_0$  and  $t_1$ , and then moving from  $x_1$  to  $x$  between times  $t_1$  and  $t$ , evaluated over all possible intermediate states,  $x_1$ . We can write this logical statement mathematically as

$$\phi(x, t | x_0, t_0) = \int \phi(x, t | x_1, t_1) \phi(x_1, t_1 | x_0, t_0) dx_1, \quad (15.1.1)$$

where this integral (and those that follow) is evaluated over the range of possible values of the random variable. Equation (15.1.1) is known as the Chapman-Kolmogorov equation, and it will come in handy below.

Let us now derive the backward equation (15.11). Consider an intermediate point in time,  $t_1$ , between the present time  $t$  and the initial time point  $t_0$  but one that is very close to  $t_0$ . In this case we can write  $t_1 = t_0 + \Delta t$ , where  $\Delta t$  is very small. The crux of the derivation revolves around the assumption that the change in the random variable  $X(t)$  over the short time period  $\Delta t$  is small enough that we can use a Taylor series with respect to this change (Primer 1). In particular, after a small amount of time  $\Delta t$  elapses, the value of  $X$  is assumed to change by a small amount  $\Delta x$ . Thus we can write the value of  $X$  at time  $t_1$  as  $x_1 = x_0 + \Delta x$ .

Our goal is to derive an expression for the derivative,  $\partial\phi/\partial t_0$ . We start by using the definition for the derivative from Appendix A2:

$$\frac{\partial\phi(x, t | x_0, t_0)}{\partial t_0} \equiv \lim_{\Delta t \rightarrow 0} \frac{\phi(x, t | x_0, t_0 + \Delta t) - \phi(x, t | x_0, t_0)}{\Delta t} \quad (15.1.2)$$

To obtain the desired expression, we must obtain an expression for the ratio  $(\phi(x, t | x_0, t_0 + \Delta t) - \phi(x, t | x_0, t_0))/\Delta t$ . First, we can replace  $\phi(x, t | x_0, t_0)$  in this ratio with the Chapman-Kolmogorov equation (15.1.1). In addition, we know that  $\int \phi(x_1, t_1 | x_0, t_0) dx_1 = 1$  because the fish must be located somewhere at time  $t_1$ . Consequently, we are free to replace  $\phi(x, t | x_0, t_0 + \Delta t)$  in (15.1.2) with  $\phi(x, t | x_0, t_0 + \Delta t) \int \phi(x_1, t_1 | x_0, t_0) dx_1$ . Making these replacements we have

$$\frac{\phi(x, t | x_0, t_0 + \Delta t) \int \phi(x_1, t_1 | x_0, t_0) dx_1 - \int \phi(x, t | x_1, t_1) \phi(x_1, t_1 | x_0, t_0) dx_1}{\Delta t}. \quad (15.1.3a)$$

(continued)

**Box 15.1** (*continued*)

By factoring and using  $t_1 = t_0 + \Delta t$ , we get

$$\frac{\int \{\phi(x, t | x_0, t_1) - \phi(x, t | x_1, t_1)\} \phi(x_1, t_1 | x_0, t_0) dx_1}{\Delta t}. \quad (15.1.3b)$$

Next, we use the fact that  $x_1 = x_0 + \Delta x$  and take the Taylor series (Recipe P1.2) of the term within the curly braces with respect to  $x$ , near the point  $x_1 = x_0$  (i.e.,  $\Delta x = 0$ ):

$$\frac{\int \left\{ -\Delta x \frac{\partial \phi(x, t | x_0, t_1)}{\partial x_0} - \frac{\Delta x^2}{2} \frac{\partial^2 \phi(x, t | x_0, t_1)}{\partial x_0^2} - O(\Delta x^3) \right\} \phi(x_1, t_1 | x_0, t_0) dx_1}{\Delta t} \quad (15.1.4a)$$

At this point, we drop the higher-order terms  $O(\Delta x^3)$ . Doing so is equivalent to assuming that process does not make large “jumps” in small time intervals (see p. 327 in Allen 2003). Replacing  $\Delta x$ , with  $x_1 - x_0$  and factoring out terms that do not depend on  $x_1$  leaves

$$\frac{\frac{\partial \phi(x, t | x_0, t_1)}{\partial x_0} \int (x_1 - x_0) \phi(x_1, t_1 | x_0, t_0) dx_1}{\Delta t} - \frac{\frac{\partial^2 \phi(x, t | x_0, t_1)}{\partial x_0^2} \int (x_1 - x_0)^2 \phi(x_1, t_1 | x_0, t_0) dx_1}{2 \Delta t}. \quad (15.1.4b)$$

Finally, we replace  $t_1$  with  $t_0 + \Delta t$  and take the limit as  $\Delta t \rightarrow 0$ , allowing us to write (15.1.2) as

$$\frac{\partial \phi(x, t | x_0, t_0)}{\partial t_0} = -\mu(x_0) \frac{\partial \phi(x, t | x_0, t_0)}{\partial x_0} - \frac{1}{2} \sigma^2(x_0) \frac{\partial^2 \phi(x, t | x_0, t_0)}{\partial x_0^2}, \quad (15.1.5)$$

where

$$\begin{aligned} \mu(x_0) &= \lim_{\Delta t \rightarrow 0} \frac{\int (x_1 - x_0) \phi(x_1, t_0 + \Delta t | x_0, t_0) dx_1}{\Delta t}, \\ \sigma^2(x_0) &= \lim_{\Delta t \rightarrow 0} \frac{\int (x_1 - x_0)^2 \phi(x_1, t_0 + \Delta t | x_0, t_0) dx_1}{\Delta t}. \end{aligned} \quad (15.1.6)$$

Equations (15.1.6) can be more easily interpreted in terms of the expected rate of change. By definition, the expected value  $E[g(X)]$  of a function  $g(X)$  is given by multiplying  $g(X)$  by the probability density function for the random variable  $X$  and integrating over all possible values of the random variable (see Table P3.1). Thus, the expected change in the random variable over a time step  $\Delta t$ , raised to the power  $k$ , is defined as

$$E[(X(t_0 + \Delta t) - X(t_0))^k | X(t_0) = x_0] = \int (x_1 - x_0)^k \phi(x_1, t_0 + \Delta t | x_0, t_0) dx_1. \quad (15.1.7)$$

*(continued)*

**Box 15.1 (continued)**

Consequently, the functions  $\mu(x_0)$  and  $\sigma^2(x_0)$  in (15.1.6) are the same drift and diffusion coefficients used in equations (15.9). Equation (15.1.5), or the equivalent (15.11), is referred to as the backward Kolmogorov equation because its derivatives depend on the original position and time, not on the current position and time as in the forward equation.

while the drift coefficient  $\mu(x)$  represents directional evolutionary change in trait space (owing to natural selection).

We can derive expressions for the drift and diffusion coefficients of evolutionary change in trait space,  $\mu(x)$  and  $\sigma^2(x)$ , based on the underlying evolutionary processes. From equations (15.1.6) of Box 15.1, these two quantities are

$$\begin{aligned}\mu(x) &= \lim_{\Delta t \rightarrow 0} \frac{\int (x_m - x) \phi(x_m, t_0 + \Delta t | x, t_0) dx_m}{\Delta t}, \\ \sigma^2(x) &= \lim_{\Delta t \rightarrow 0} \frac{\int (x_m - x)^2 \phi(x_m, t_0 + \Delta t | x, t_0) dx_m}{\Delta t},\end{aligned}\tag{15.12}$$

where  $\phi(x_m, t_0 + \Delta t | x, t_0)$  is the probability that the population moves from trait  $x$  to trait  $x_m$  during the time interval  $\Delta t$ . Next, we specify how the transitions  $\phi(x_m, t_0 + \Delta t | x, t_0)$  are related to the underlying processes of mutation and natural selection.

The probability that a population moves to state  $x_m$  from state  $x$  in the time interval  $\Delta t$  equals the probability that a mutation of type  $x_m$  occurs in that time interval, which we denote by  $M(x_m, x; \Delta t)$ , times the probability that this new mutation ultimately replaces the resident allele (i.e., the probability of fixation), which we denote by  $U(x_m, x)$  (the time scale is assumed to be so long that the fixation or loss of the new allele is nearly instantaneous). The above drift and diffusion coefficients can then be written as

$$\begin{aligned}\mu(x) &= \lim_{\Delta t \rightarrow 0} \frac{\int (x_m - x) M(x_m, x; \Delta t) U(x_m, x) dx_m}{\Delta t}, \\ \sigma^2(x) &= \lim_{\Delta t \rightarrow 0} \frac{\int (x_m - x)^2 M(x_m, x; \Delta t) U(x_m, x) dx_m}{\Delta t}.\end{aligned}\tag{15.13}$$

The diffusion equations (15.8) and (15.11), along with the drift and diffusion parameters (15.13), describe how the probability distribution for the resident trait value changes through time as a result of mutation and selection. This gives us information about the probable states of a population over time,

which we shall explore further in section 15.3.2. The expected trait value over time can also be derived for this diffusion model using the forward Kolmogorov equation, as described in Supplementary Material 15.1.

#### 15.2.4 Diffusion Models as Approximations to Discrete-State-Space Models

In some models, a biological phenomenon of interest will inherently have a discrete state space (e.g., the number of copies of a particular allele in a finite population, or the number of islands occupied by a plant species). Even when a discrete state space is more natural, many stochastic models can be analyzed by approximating the process using a continuous state space model such as the diffusion model with drift. In essence, the approximation treats both the state variable and time as continuous, distilling the stochastic process down to two quantities: the expected rate of change in the variable and the amount of variability around this expectation.

Let us take a relatively simple example to illustrate. The Moran model for allele frequency change in a finite haploid population was introduced in section 13.5 of Chapter 13 and analyzed in section 14.4 of Chapter 14. This is a stochastic model in which the state space is inherently discrete: the random variable  $X(t)$  can take on any of the discrete values from zero to  $N$ . Suppose that there is no natural selection acting, in which case the expected change in  $X(t)$  is zero regardless of its current value. Although the expected change is zero, the value of  $X$  will nevertheless change by chance. In section 13.5, we calculated the variance of the change in  $X$  per time unit as  $2 p (1 - p)$ , where  $p = X/N$  is the frequency of the allele.

As a first step toward using a diffusion model to approximate the Moran process, we model the dynamics of the variable  $p$ , assuming that the drift and diffusion coefficients are given by  $\mu(p) = 0$  and  $\sigma^2(p) = \sigma^2(X/N) = 2 p (1 - p)/N^2$ , where we have used the fact from Table P3.1 that  $\sigma^2(c Y) = c^2 \sigma^2(Y)$ . (In Box 15.2, we describe how to derive these coefficients more formally; see Problem 15.2.) In this case, the forward diffusion equation (15.8) simplifies to

$$\frac{\partial \phi(p, t | p_0, t_0)}{\partial t} = \frac{1}{2} \frac{\partial^2 \left( \frac{2p(1-p)}{N^2} \phi(p, t | p_0, t_0) \right)}{\partial p^2} \quad (15.14)$$

In Figure 15.2, we use *Mathematica* to solve the partial differential equation (15.14) numerically and then compare the results to exact iterations of the Moran model. Even though the population size consists of only ten haploid individuals, the diffusion approximation provides an amazingly accurate description of the change in allele frequency over time.

To some extent, the fit of the diffusion in this case is a bit lucky, as we have not formally derived the drift and diffusion coefficients nor have we shown that the diffusion approximation is reasonable for this model. Fortunately, there is a procedure for determining whether a diffusion approximation is valid for a discrete-state-space model (Box 15.2). Not all discrete processes can be

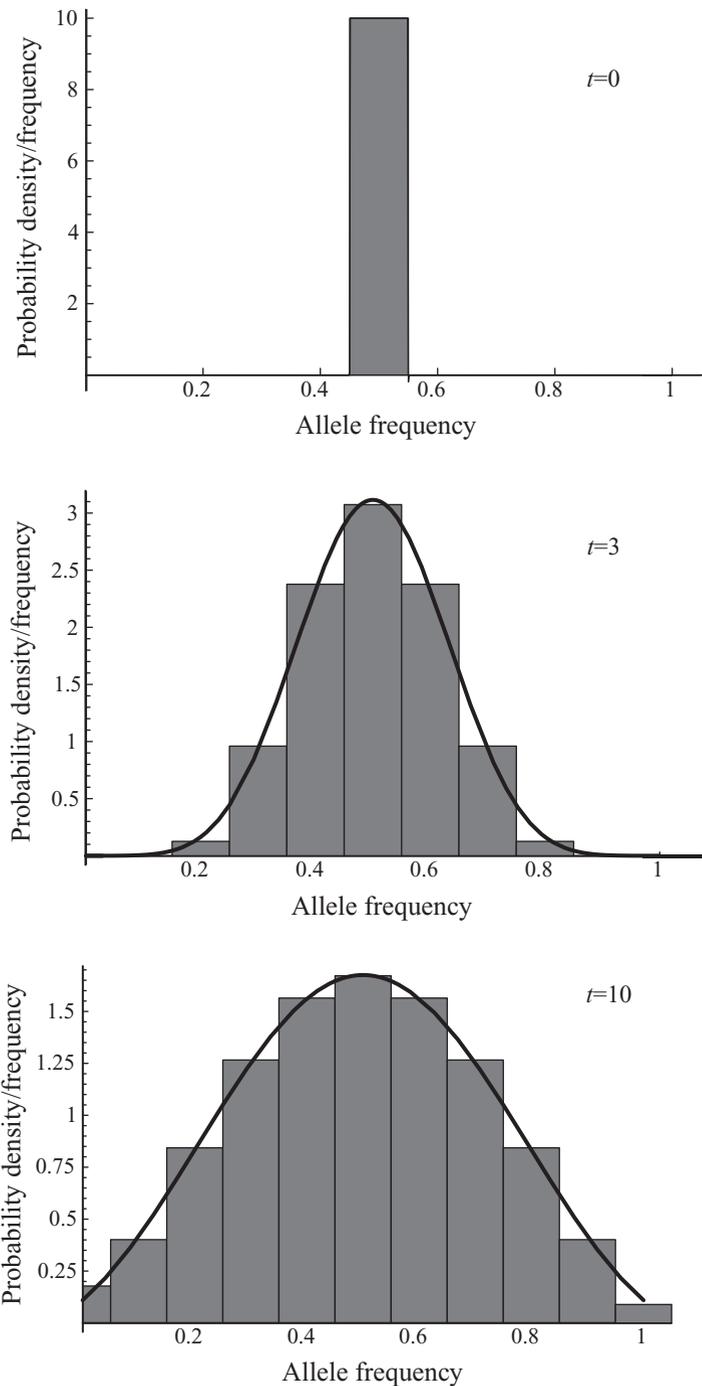


Figure 15.2: Allele frequency distributions in the Moran model. The distribution of allele frequencies is shown at three different points in time for a neutral allele in a finite haploid population. Bars illustrate the predicted frequency distribution from iterations of the exact matrix model using equation (13.9), and the curve illustrates the probability density obtained by numerically solving the partial differential equation (15.14) using *Mathematica*. Initial allele frequency is  $p = 0.5$ , and population size is  $N = 10$ .

adequately approximated in this way, but when they can, the techniques of Box 15.2 provide a way of choosing the drift and diffusion coefficients as well as the appropriate scales along which to measure the state variable and time. These techniques determine whether or not a diffusion model is likely to yield a reasonable approximation for a discrete-state-space model.

**Box 15.2: Deriving Diffusion Approximations**

In a diffusion analysis, the random variable and time are both treated as continuous variables, yet diffusion models are often used to approximate stochastic models with discrete variables and discrete time steps. In this box we illustrate how to go about deriving and justifying such an approximation. The key lies in choosing an appropriate transformation of the original model. Even if the original discrete-time, discrete-state-space model might not be well approximated by a diffusion model, perhaps some transformation of the model will be. If so, the diffusion approximation is not an approximation for the original model *per se* but rather, an approximation for the transformed model. Thus, to make predictions in terms of the original model's variables using a diffusion approximation, you must then transform the results of the diffusion model back into the original variables.

Consider a discrete-time discrete-variable stochastic model with transition probabilities  $p_{ji}$  determining the probability of moving from  $i$  to  $j$  in a single time step. We denote the random variable governed by this process as  $Y(t)$ . By summing up over all possible transitions that might occur in one time step, we can calculate the moments of the change in the variable  $Y$  (describing the mean change, the variance in this change, etc.) as

$$\begin{aligned} \text{First moment} \quad E[Y(t+1) - Y(t) \mid Y(t) = i] &= \sum_{j=0}^m (j - i)p_{ji}, \\ \text{Second moment} \quad E[(Y(t+1) - Y(t))^2 \mid Y(t) = i] &= \sum_{j=0}^m (j - i)^2 p_{ji}, \\ \text{Third moment} \quad E[(Y(t+1) - Y(t))^3 \mid Y(t) = i] &= \sum_{j=0}^m (j - i)^3 p_{ji}, \end{aligned} \quad (15.2.1)$$

where  $m$  is the maximum possible value of the random variable.

The next step is to transform the original state variable  $Y$  and the original time variable  $t$  into new state and time variables  $X$  and  $T$ , chosen such that, for large enough  $m$ , the third moment is essentially zero whereas the first moment remains finite, and the second moment remains finite and positive. A transformation of the state variable that often works is

$$X = \frac{Y}{m^\beta}, \quad (15.2.2)$$

Thus, while the original variable  $Y$  takes on discrete values between zero and  $m$ , the transformed variable  $X$  takes on discrete values between zero and  $m^{1-\beta}$ , where  $\beta$  is a constant that will be chosen shortly. Similarly, a transformation of the time variable that often proves useful is

$$T = \frac{t}{m^\alpha}, \quad (15.2.3)$$

where  $\alpha$  is a constant that will also be chosen shortly. Consequently, one unit of time in the transformed model corresponds to  $m^\alpha$  units of time in the original model.

(continued)

**Box 15.2 (continued)**

Let us take a step back for a moment to digest what we have done. The original random variable  $Y$  changes from one time step to the next, where the length of this time step might potentially be quite large depending upon what we are modeling. Furthermore, the size of the changes in  $Y$  that occur over these time steps might also be quite large. The transformed model, however, has different properties. It, too, is a discrete-time, discrete-state-space model, but each discrete step occupies a much smaller fraction of a unit in time. In particular, each time step of the original model corresponds to only  $1/m^\alpha$  units of time in the transformed model. Therefore, as long as the size of the state space,  $m$ , is reasonably large, the time steps of the transformed model will appear very small. Additionally, in each of these small time steps, the transformed variable  $X$  changes only by the small amount  $\Delta Y/m^\beta$ . Therefore, we might expect this transformed model to be better approximated by a diffusion model because its variables change in a more continuous fashion.

We now determine the moments of change in the transformed model, given the moments of change (15.2.1) in the original model. We begin by transforming the state space to get the moments of change in the transformed variable  $X$  but keeping the original time variable  $t$ . Across one of the original time steps, the first moment of change in the transformed variable  $X$  starting from position  $X$  (corresponding to  $i$  in the original state space) is, from (15.2.2),

$$E[X(t + 1) - X(t) | X(t) = x] = E\left[\frac{Y(t + 1)}{m^\beta} - \frac{Y(t)}{m^\beta} \middle| \frac{Y(t)}{m^\beta} = \frac{i}{m^\beta}\right]. \quad (15.2.4a)$$

Using (15.2.1), this equals

$$\begin{aligned} E[X(t + 1) - X(t) | X(t) = x] &= \frac{1}{m^\beta} E[Y(t + 1) - Y(t) | Y(t) = i] \\ &= \frac{1}{m^\beta} \sum_{j=0}^m (j - i) p_{ji}. \end{aligned} \quad (15.2.4b)$$

Equation (15.2.4b) describes the amount of change in one time step in the original time scale. To obtain the *rate of change* in the original time scale we must divide by the length of the time step. Because we have defined the time steps to be of length one in the original model, this means that we must divide by one to get the rate of change in the original model. To complete the transformation we must now calculate the rate of change measured in the new time scale. Starting at time point  $T$  and using (15.2.3), the change in the transformed time variable that occurs over one time step in the original time scale is  $\Delta T = (t + 1)/m^\alpha - (t)/m^\alpha = 1/m^\alpha$ . Therefore, the rate of change of the first moment in the fully transformed model equals equation (15.2.4b) divided by  $\Delta T$ . Setting  $\Delta T = 1/m^\alpha$  on the right, we get

$$\text{First moment} \quad \frac{E[X(T + \Delta T) - X(T) | X(T) = x]}{\Delta T} = \frac{m^\alpha}{m^\beta} \sum_{j=0}^m (j - i) p_{ji}. \quad (15.2.5a)$$

(continued)

**Box 15.2 (continued)**

Similar calculations give the rate of change of the second and third moments as

$$\text{Second moment} \quad \frac{E[X(T + \Delta T) - X(T)]^2 | X(T) = x]}{\Delta T} = \frac{m^\alpha}{m^{2\beta}} \sum_{j=0}^m (j - i)^2 p_{ji} \quad (15.2.5b)$$

$$\text{Third moment} \quad \frac{E[X(T + \Delta T) - X(T)]^3 | X(T) = x]}{\Delta T} = \frac{m^\alpha}{m^{3\beta}} \sum_{j=0}^m (j - i)^3 p_{ji}. \quad (15.2.5c)$$

We now have two descriptions of the biological process of interest: (i) our original model, whose moments are given by (15.2.1), and (ii) our transformed model, whose moments are given by (15.2.5). To approximate a model using a diffusion equation, we must demonstrate that: (a) the rate of change in the mean is a finite number (including, potentially, zero); (b) the rate of change in the variance is a positive and finite number; and (c) the rate of change of any one of the higher-order moments is zero (see pp. 157–165 in Karlin and Taylor 1981 and pp. 327–328 in Allen 2003), although, typically, this is proven for the third or fourth moment (Karlin and Taylor 1981).

Usually the original description of the biological process will not satisfy these three requirements exactly. For example, changes in the original random variable  $Y$  might be large, causing nonzero rates of change of the higher-order moments. The same might be true with the transformed model, except that we are now free to choose  $\alpha$  and  $\beta$ . The goal is to choose them in such a way that rules (a)–(c) are approximately satisfied. By “approximately” we mean that, as the size of the original state space,  $m$ , gets larger and larger (so that the discrete time steps in the transformed model, along with the changes in  $X$ , get smaller and smaller), conditions (a)–(c) are better met. Mathematically, in the limit as  $m$  goes to infinity, conditions (a)–(c) should hold exactly. Under these conditions, a diffusion model should yield a reasonable approximation for the transformed model so long as the actual value of  $m$  is not too small.

To summarize, as long as the state space,  $m$ , is not too small, a diffusion approximation will provide a reasonable description of the dynamics provided that we can choose values of  $\alpha$  and  $\beta$  such that:

$$\begin{aligned} \text{First moment} \quad \mu(x) &= \lim_{m \rightarrow \infty} \frac{m^\alpha}{m^\beta} \sum_{j=0}^m (j - i) p_{ji}, \\ \text{Second moment} \quad \sigma^2(x) &= \lim_{m \rightarrow \infty} \frac{m^\alpha}{m^{2\beta}} \sum_{j=0}^m (j - i)^2 p_{ji}, \\ \text{Third moment} \quad 0 &= \lim_{m \rightarrow \infty} \frac{m^\alpha}{m^{3\beta}} \sum_{j=0}^m (j - i)^3 p_{ji}, \end{aligned} \quad (15.2.6)$$

where  $\mu(x)$  is finite and  $\sigma^2(x)$  is finite and positive. These coefficients may then be used in the Kolmogorov equations (15.8) and (15.11) as a continuous-time, continuous-state-space approximation to the transformed model.

Given a discrete stochastic model, the first step in deriving a diffusion approximation is to obtain the moments (15.2.1) based on the transition probabilities for the original model. These

(continued)

**Box 15.2 (continued)**

moments are then plugged in for the sums in (15.2.6) and appropriate values of  $\alpha$  and  $\beta$  are chosen. If the restrictions in (15.2.6) cannot be met, consider whether there is a scaling of the parameters (in addition to the transformation of the variables) that will satisfy the restrictions (e.g., multiplying the parameters by  $m$ ). If the restrictions are still not met, consider whether there is some nondecreasing function  $g(m)$  other than  $m^\alpha$  that can be used to rescale the variables. If it is not possible to transform the model to satisfy (15.2.6), then the biological process might not be accurately described by a diffusion approximation. Finally, it is important to remember that you should reverse the transformations when you wish to interpret the results of a diffusion analysis in terms of the original variables and parameters.

Before leaving the subject, it should be mentioned that the diffusion approximation is often used without performing the transformation described in this box and ensuring that the approximation is valid. That is, the first and second moments for the original model (15.2.1) are naively taken to be the drift and diffusion coefficients (as we did in section 15.2.4). This approach sometimes works, yielding results that match simulations. But only by carrying out the above procedure can you be assured that the diffusion approximation is valid (see Problem 15.8).

### 15.3 Analyzing the Diffusion Equation with Drift

As with the other dynamical models that we have considered in this book, it would be ideal to derive a general solution for the diffusion equation with drift. Occasionally this is possible, but typically it isn't. Nevertheless, when a general solution is not available, we can still obtain many useful results analytically, just as we have done in previous chapters. In this section we illustrate some of the more important results for diffusion models. As it is often easier to appreciate a result when you know where it comes from, we provide derivations of the stationary distribution, the probability of absorption, and expected time to absorption in sections 15.3.2, 15.3.3, and Supplementary Material 15.2, respectively. Having derived these results, Recipes 15.1–15.3 summarize the solutions and provide a quick way to obtain results from any diffusion model by plugging in the appropriate drift and diffusion coefficients. These recipes involve integrals that can be calculated analytically or, when this is not possible, numerically. Thus, the main advantage of diffusion theory is that it can be used to reduce the analysis of a complex stochastic model to the evaluation of a handful of integrals.

*Brownian motion* describes random movement through space, with constant drift and diffusion coefficients.

#### 15.3.1 General Solutions to the Diffusion Model

Although it is not possible to derive a general solution to the diffusion model for arbitrary drift and diffusion coefficients,  $\mu(x)$  and  $\sigma^2(x)$ , it is possible to do so for certain special cases. One such example is constant Brownian motion, which is characterized by  $\mu(x) = \mu$  and  $\sigma^2(x) = \sigma^2$  for all  $x$ . In the classic

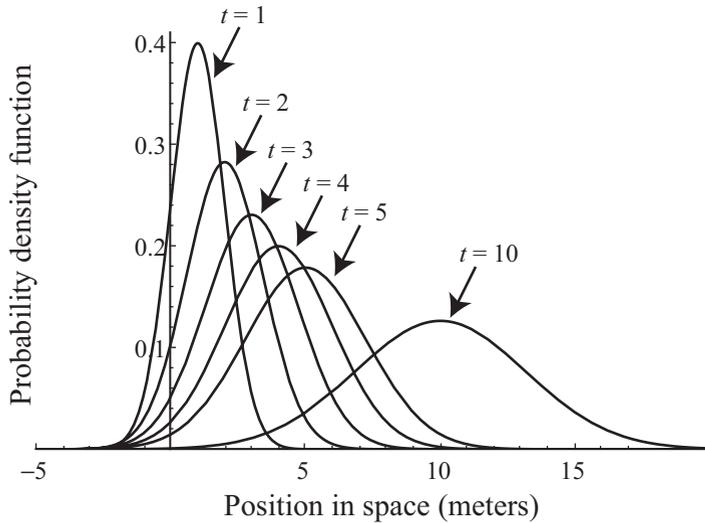


Figure 15.3: Diffusion over space with Brownian motion. Assuming that particles are initially located at  $x_0 = 0$  at time  $t_0 = 0$ , equation (15.15) describes how the spatial distribution of particles spreads over time under Brownian motion. Parameters were set to  $\mu = 1$  meter per unit time and  $\sigma^2 = 1$  meter<sup>2</sup> per unit time. The probability density of particles becomes normally distributed with a breadth that increases as a square-root function of time.

Brownian motion model,  $x$  represents location in one-dimensional space, and the particles are assumed to start at a particular position  $x_0$  at time  $t_0$ . This is exactly the case for our model of fish movement in equation (15.7). With constant drift and diffusion parameters, a solution for the probability density function for the location of the fish at time  $t$  (using either the backward or forward Kolmogorov equation) is

$$\phi(x, t | x_0, t_0) = \frac{e^{-(x - (x_0 + (t - t_0)\mu))^2 / (2(t - t_0)\sigma^2)}}{\sqrt{2\pi(t - t_0)\sigma^2}} \quad (15.15)$$

(see Problem 15.3).

The solution (15.15) is the probability density function for a normal distribution (see Definition P3.14a) with mean of  $x_0 + (t - t_0)\mu$  and variance  $(t - t_0)\sigma^2$ . From this we conclude that the expected location of the fish moves away from its starting location,  $x_0$ , at a rate that depends on the drift parameter  $\mu$ . According to equations (15.6), the drift parameter is proportional to the difference between upstream movement,  $u$ , and downstream movement,  $d$ . When prevailing currents cause  $d > u$ , the expected location of the fish is further and further downstream as time passes. From Equation (15.15) we can also see that the variance in the location of the fish increases linearly with time, at a rate proportional to the diffusion coefficient,  $\sigma^2$ . Consequently, the standard deviation, which measures the breadth of the normal curve, rises with the square root of time (Figure 15.3).

### 15.3.2 The Stationary Distribution of a Diffusion Model

While it is fantastic to obtain a general solution from a diffusion model, it is often not possible to solve the partial differential equations (15.8) or (15.11). Nevertheless, we can seek the stationary distribution of a diffusion model, just as we did for the discrete state space models in Chapter 14. By definition, a

stationary distribution,  $\pi(x)$ , does not change over time:  $\partial\pi(x)/\partial t = 0$ . If the system is at a stationary distribution at time  $t_0$ , then the probability density at any future time  $t$  will also be at the stationary distribution and must satisfy  $\pi(x) = \int \pi(x_0) \phi(x, t | x_0, t_0) dx_0$ , when integrated over all possible initial states,  $x_0$ . Using this fact, we can multiply both sides of the forward Kolmogorov equation (15.8) by  $\pi(x_0)$  and integrate over initial states  $x_0$  to get

$$\begin{aligned} \int \pi(x_0) \frac{\partial \phi(x, t | x_0, t_0)}{\partial t} dx_0 &= - \int \pi(x_0) \frac{\partial (\mu(x) \phi(x, t | x_0, t_0))}{\partial x} dx_0 \\ &+ \frac{1}{2} \int \pi(x_0) \frac{\partial^2 (\sigma^2(x) \phi(x, t | x_0, t_0))}{\partial x^2} dx_0. \end{aligned}$$

Because  $\pi(x_0)$  does not depend on time, we can interchange the order of integration and differentiation to get

$$\begin{aligned} \frac{\partial \int \pi(x_0) \phi(x, t | x_0, t_0) dx_0}{\partial t} &= - \frac{\partial \int \pi(x_0) \mu(x) \phi(x, t | x_0, t_0) dx_0}{\partial x} \\ &+ \frac{1}{2} \frac{\partial^2 \int \pi(x_0) \sigma^2(x) \phi(x, t | x_0, t_0) dx_0}{\partial x^2} \end{aligned}$$

Plugging  $\pi(x) = \int \pi(x_0) \phi(x, t | x_0, t_0) dx_0$  into each term and recalling that  $\partial\pi(x)/\partial t = 0$  leaves

$$0 = - \frac{d(\mu(x) \pi(x))}{dx} + \frac{1}{2} \frac{d^2(\sigma^2(x) \pi(x))}{dx^2}. \quad (15.16)$$

Because the state of the system at the stationary distribution does not depend on time, equation (15.16) is now an ordinary differential equation involving derivatives with respect to  $x$  only, rather than a partial differential equation. Equation (15.16) must be satisfied by any potential stationary distribution,  $\pi(x)$ .

To solve equation (15.16) for the stationary distribution,  $\pi(x)$ , first integrate both sides with respect to  $x$ :

$$c_0 = -\mu(x) \pi(x) + \frac{1}{2} \frac{d(\sigma^2(x) \pi(x))}{dx}, \quad (15.17a)$$

where  $c_0$  is a constant of integration. At this point, a change of notation makes things clearer. If we define the function  $n(x) = \sigma^2(x) \pi(x)$  and rearrange, (15.17a) becomes

$$\frac{dn(x)}{dx} = \frac{2\mu(x)}{\sigma^2(x)} n(x) + c_1, \quad (15.17b)$$

where  $c_1 = 2c_0$ . Equation (15.17b) is a more familiar ordinary differential equation, of the sort described in Box 6.2. Specifically, (15.17b) is an example of a linear differential equation (6.2.1), and therefore its solution is

$$n(x) = e^{A(x)} \left( c_1 \int e^{-A(x)} dx + c_2 \right) \quad (15.17c)$$

where

$$A(x) = \int \frac{2\mu(x)}{\sigma^2(x)} dx.$$

Using the fact that  $n(x) = \sigma^2(x) \pi(x)$ , we now have an explicit equation for the stationary distribution:

### Recipe 15.1

**The Stationary Distribution of the Diffusion Equation** The stationary distribution of the diffusion equation with drift parameter  $\mu(x)$  and diffusion parameter  $\sigma^2(x)$  is

$$\pi(x) = \frac{e^{A(x)} \left( c_1 \int e^{-A(x)} dx + c_2 \right)}{\sigma^2(x)} \quad (15.18)$$

where  $A(x) = \int 2\mu(x)/\sigma^2(x) dx$ . The constants of integration,  $c_1$  and  $c_2$ , must be chosen such that  $\pi(x)$  integrates to one, is never negative, and satisfies any other required conditions at the endpoints of the allowable values of  $x$ . Only if there are such choices of constants does the stationary distribution exist.

It is always possible to check whether or not the stationary distribution that you derive is correct by substituting it into the forward Kolmogorov equation. If you have not made any errors, then carrying out the differentiations on the right-hand side of equation (15.8) should result in zero (by the definition of a stationary distribution).

### Example: The Stationary Distribution for a Trait when Mutations Are Limiting

Here we apply Recipe 15.1 to the model of mutation-limited evolution with drift and diffusion coefficients given by (15.13). To make things more concrete, let's implement this recipe for the model (12.16) from Chapter 12 involving the evolution of daphnia feeding strategies, where the population size of daphnia was described by a deterministic recursion equation. The growth factor of a mutant daphnia strain with feeding strategy  $s_m$  in a population with resident feeding strategy  $s$  was given by equation (12.18):

$$\lambda = 1 + r - r \frac{\alpha(s_m, s) K(s)}{K(s_m)}.$$

where  $\alpha(s_m, s)$  is the competition coefficient and  $K(s)$  is the carrying capacity.

To construct a stochastic model for this question we make slightly different assumptions than those of model (12.16). In particular, we assume that the population of daphnia is fixed at size  $N$  and undergoes a birth-death process described by the Moran model. These assumptions allow us to use the results of Chapter 14 to describe the probability that a mutant allele reaches fixation in the population.

First, let us specify the mutant distribution  $M(s_m, s; \Delta t)$ , which describes the probability distribution of mutants with strategy  $s_m$  arising within a population whose feeding strategy is  $s$  over a time interval  $\Delta t$ . We suppose that the probability that one of the  $N$  individuals in the population gives birth to a mutant individual in time interval  $\Delta t$  is given by  $N \eta \Delta t$ , where  $\eta$  is the per capita rate at which mutant offspring are produced. Given that a mutant offspring is produced, we then suppose that its phenotype is normally distributed with a mean equal to the current trait value  $s$  and a variance of  $\nu$  (see Definition P3.14a). Therefore,

$$M(s_m, s; \Delta t) = N \eta \Delta t \frac{e^{-(s_m - s)^2 / (2\nu)}}{\sqrt{2\pi\nu}}, \quad (15.19)$$

Next, we use the birth-death model to describe the fixation probability of a mutant allele arising within the population of daphnia,  $U(s_m, s)$ . We assume that the per capita birth rate is constant at  $b = r = 1$ , and that the per capita death rate is  $d(s_m, s) = \alpha(s_m, s) K(s) / K(s_m)$ . Consequently, when the mutant type has the same feeding strategy as the resident type, the birth rate equals the death rate, and the mutation is completely neutral. Furthermore, we assume that the amount of resources follows a bell-shaped distribution (Definition P1.6) with a peak at  $s = 0$  using the function  $K(s) = K e^{-s^2/a_k}$ , and that the competition coefficient is  $\alpha(s_m, s) = e^{-(s_m - s)^2/a_c}$  as in Chapter 12.

The above birth-death model implicitly assumes that the death rate does not change as the number of mutant individuals in the population increases. In reality this is not the case because eventually the mutant becomes so abundant that it will begin to experience interactions with other mutant daphnia. This can be remedied by making the death rate function  $d(s_m, s)$  depend on the frequency of mutant individuals. We ignore this complication here, under the assumption that the fixation probability of the mutant is determined while the mutant is still rare. The probability of fixation of the mutant type in a population of  $N$  daphnia equals one minus the probability of loss, given by equation (14.56) for a birth-death model:

$$U(s_m, s) = \frac{1 - d(s_m, s)}{1 - d(s_m, s)^N}. \quad (15.20)$$

To obtain tractable solutions for the drift and diffusion coefficients in (15.13), we use a linear approximation to (15.20), under the assumption that the difference between the mutant strategy and the resident strategy is not too large.

(Identical results are obtained if we use a more accurate quadratic approximation.) Using Recipe P1.2 to calculate the first two terms of the Taylor series of (15.20) with respect to  $s_m$ , we get

$$U(s_m, s) = \frac{1}{N} - \frac{s(N-1)}{N a_K} (s_m - s). \quad (15.21)$$

(Evaluating the terms in the Taylor series in the limit as  $s_m$  goes to  $s$  requires L'Hôpital's rule, as described in Appendix 2.)

Plugging (15.19) and (15.21) into expressions (15.13) gives

$$\begin{aligned} \mu(s) &= \eta \int (s_m - s) \frac{e^{-(s_m - s)^2 / (2\nu)}}{\sqrt{2\pi\nu}} \left( 1 - \frac{s(N-1)}{a_K} (s_m - s) \right) ds_m \\ &= -\eta \frac{s(N-1)}{a_K} \nu, \end{aligned} \quad (15.22a)$$

$$\begin{aligned} \sigma^2(s) &= \eta \int (s_m - s)^2 \frac{e^{-(s_m - s)^2 / (2\nu)}}{\sqrt{2\pi\nu}} \left( 1 - \frac{s(N-1)}{a_K} (s_m - s) \right) ds_m \\ &= \eta \nu \end{aligned} \quad (15.22b)$$

where we have used the fact that  $\int (s_m - s)^i (e^{-(s_m - s)^2 / (2\nu)} / \sqrt{2\pi\nu}) ds_m$  is the  $i$ th central moment of the normal distribution and equals 0 for  $i = 1$  (the first central moment),  $\nu$  for  $i = 2$  (the second central moment), and 0 for  $i = 3$  (the third central moment) (Appendix 5, equations (A5.10)).

Equations (15.22) give the drift and diffusion coefficients, but we must also ensure that the rate of change of any one of the higher order moments is zero before we can validly use a diffusion model (see pp. 157–165 in Karlin and Taylor (1981) and pp. 327–328 in Allen (2003)). In Problem 15.4, you are asked to demonstrate that, by rescaling time, and measuring it in units of the mutational variance,  $\nu$ , the higher order moments vanish as  $\nu$  goes to zero. Furthermore, the infinitesimal mean and variance then become  $\mu(s) = -\eta (s(N-1)/a_K)$  and  $\sigma^2(s) = \eta$ . Therefore, on this new time scale, the diffusion equation with these rescaled moments should provide a suitable model for mutation-limited evolution provided that the mutational variance,  $\nu$ , is not too large. This is analogous to the principle we employed in Box 15.2, where we saw that we could use a diffusion equation to approximate a discrete stochastic model so long as the state space was large enough.

We are now ready to follow Recipe 15.1 to find the stationary distribution. From Recipe 15.1 we have  $A(s) = \int 2 \mu(s) / \sigma^2(s) ds = \int -2 s (N-1) / a_K ds$ , giving  $A(s) = -2 s^2 (N-1) / a_K$  (the constant of integration is already included in (15.18)). Substituting into (15.18), we obtain

$$\pi(s) = \frac{e^{-(2(N-1)s^2/a_K)} \left( c_1 \int e^{(2(N-1)s^2/a_K)} ds + c_2 \right)}{\eta} \quad (15.23)$$

The final step is to choose the constants of integration in (15.23). This must be done such that the probability density is never negative and integrates to one. Because the potential values of  $s$  are anywhere from  $-\infty$  to  $+\infty$ , we must ensure that the probability density decays to zero as  $s$  gets very large or very small. The term inside the integral gets very large as  $s$  goes to  $-\infty$  or to  $+\infty$ , and therefore we must set  $c_1 = 0$  to ensure that (15.23) is finite. The remaining terms in equation (15.23),  $e^{-2s^2(N-1)/a_K c_2/\eta}$ , integrate to  $\sqrt{\pi a_K/(2(N-1))} c_2/\eta$ , which we set equal to one and solve for the constant,  $c_2$ . The stationary distribution thus equals

$$\pi(s) = \frac{e^{-s^2/2V}}{\sqrt{2\pi V}} \quad (15.24)$$

where  $V = a_K/(4(N-1))$ . Interestingly, the probability density (15.24) is a normal distribution with a mean centered at zero and a variance of  $a_K/(4(N-1))$ .

Once the stationary distribution is reached, the population continues to change stochastically from one resident trait value to another. But if we observe the population at some point in time, then (15.24) describes the probability density that the observed state will be  $s$ . From the form of (15.24), we can see that the population is most likely to have a feeding strategy yielding the greatest carrying capacity (i.e.,  $s = 0$ ). The feeding strategies will vary widely around this optimum when the resource distribution is wide (large  $a_K$ ) and when the daphnia population size is small (more genetic drift). Conversely, the feeding strategies will cluster tightly around the optimum in very large populations.

In Chapter 12, we found that evolutionary branching occurred when the resource distribution was wider than the competition function ( $a_K > a_\alpha$ ), in which case a polymorphism was expected. Yet no such diversification occurs in the above stochastic model of mutation-limited evolution. What causes this discrepancy? We constructed the diffusion model above to describe the state of the population under the assumption that the population is always monomorphic. By assumption, when mutations arise, they either disappear or fix and immediately replace the parental strain. The possibility that the population becomes polymorphic is therefore not allowed. Thus, this diffusion approach is limited to describing the probability distribution of populations as long as they remain, for most periods of time, dominated by single alleles.

### Example: The Wright-Fisher model

Next, we consider the Wright-Fisher model of allele frequency change with mutation and selection in a haploid population. This model was introduced in section 13.4 as an example of a discrete-time, discrete-state stochastic model. The discrete states of the model represent the number of copies of an allele within a population, ranging from zero up to a maximum of  $N$ . Unless the population size is small, this model is difficult to analyze using the techniques of Chapter 14, so let us approximate the model with a diffusion equation using Box 15.2. In Box 15.3, we show that the Wright-Fisher model can be approximated by a diffusion if we measure time in units of  $N$  generations and keep

**Box 15.3: The Drift and Diffusion Parameters in the Wright-Fisher Model**

Here we derive a diffusion approximation for the Wright-Fisher model of allele frequency change in a haploid population using the methods of Box 15.2. Our goal is to show that there is a valid diffusion approximation for the Wright-Fisher model as the population size gets large. First we consider the case of no mutation or selection.

In Chapter 13, we introduced the Wright-Fisher model and pointed out that its transition probabilities are binomially distributed. Specifically, using  $Y(t)$  to denote the number of copies of allele  $A$  at time  $t$ , the probability of a transition from  $i$  copies to  $j$  copies is

$$\begin{aligned} p_{ji} &= P(Y(t+1) = j \mid Y(t) = i) \\ &= \binom{N}{j} (p)^j (1-p)^{N-j}, \end{aligned}$$

where  $N$  is the size of the population and  $p$  is the allele frequency ( $p = i/N$ ) when there is no selection or mutation. Using the mean of the binomial (Table 3.2), equation (15.2.1) becomes

$$\begin{aligned} \text{First moment } E[Y(t+1) - Y(t) \mid Y(t) = i] &= E[Y(t+1) \mid Y(t) = i] - E[Y(t) \mid Y(t) = i] \\ &= Np - i, \end{aligned}$$

which equals  $N(i/N) - i = 0$ . This reveals that there is no directional tendency in allele frequency, as we would expect in the absence of selection and mutation.

Because  $E[Y(t+1) \mid Y(t) = i] = i$ , we can rewrite the second moment in (15.2.1),  $E[(Y(t+1) - Y(t))^2 \mid Y(t) = i]$  as  $E[(Y(t+1) - E[Y(t+1) \mid Y(t) = i])^2 \mid Y(t) = i]$ , which is the formula for the variance of  $Y(t+1)$ . For a binomial distribution, this variance is given by equation (Table 3.2)

$$\text{Second moment } E[(Y(t+1) - E[Y(t+1) \mid Y(t) = i])^2 \mid Y(t) = i] = N \left( \frac{i}{N} \right) \left( 1 - \frac{i}{N} \right).$$

Similarly, the third moment in (15.2.1) for the Wright-Fisher model is given by the third central moment of the binomial distribution. From the calculations of Appendix 5, this moment equals

$$\text{Third moment } E[(Y(t+1) - E[Y(t+1) \mid Y(t) = i])^3 \mid Y(t) = i] = N \left( \frac{i}{N} \right) \left( 1 - \frac{i}{N} \right) \left( 1 - 2 \frac{i}{N} \right).$$

In terms of these original variables, the 3<sup>rd</sup> moment will not go to zero in the limit as the population size gets larger. Can we transform the model using equations (15.2.2) and (15.2.3) so that the diffusion approximation holds? To find out, we must find the moments in terms of the transformed variables by plugging in the above expressions into equation (15.2.6):

$$\text{First moment} \quad \mu(x) = 0,$$

$$\text{Second moment} \quad \sigma^2(x) = \lim_{m \rightarrow \infty} \frac{m^\alpha}{m^{2\beta}} N \left( \frac{i}{N} \right) \left( 1 - \frac{i}{N} \right),$$

$$\text{Third moment} \quad 0 = \lim_{m \rightarrow \infty} \frac{m^\alpha}{m^{3\beta}} N \left( \frac{i}{N} \right) \left( 1 - \frac{i}{N} \right) \left( 1 - 2 \frac{i}{N} \right).$$

(continued)

**Box 15.3 (continued)**

In the Wright-Fisher model, the system size is  $N + 1$ , because the original variable  $Y(t)$  can take on any value from 0 to  $N$ . Although we could set  $m = N + 1$ , it is inconvenient to keep track of the  $+1$  term. Because any transformation that satisfies the above moment equations will suffice, we set  $m = N$ . Next, we rewrite the  $i$ 's (which are particular values of the random variable  $Y$ ) as  $x$ 's (which are particular values of the transformed variable  $X$ ) using equation (15.2.2), giving  $i = x N^\beta$ . Making these substitutions, the new moments become

$$\begin{aligned} \text{First moment} & \quad \mu(x) = 0, \\ \text{Second moment} & \quad \sigma^2(x) = \lim_{N \rightarrow \infty} \frac{N^\alpha}{N^{2\beta}} N \left( \frac{x N^\beta}{N} \right) \left( 1 - \frac{x N^\beta}{N} \right), \\ \text{Third moment} & \quad 0 = \lim_{N \rightarrow \infty} \frac{N^\alpha}{N^{3\beta}} N \left( \frac{x N^\beta}{N} \right) \left( 1 - \frac{x N^\beta}{N} \right) \left( 1 - 2 \frac{x N^\beta}{N} \right). \end{aligned}$$

All that remains is to simplify these expressions and choose the values of  $\alpha$  and  $\beta$  so that the appropriate restrictions are met. The second moment expands to become

$$\sigma^2(x) = \lim_{N \rightarrow \infty} \left( \frac{N^\alpha x}{N^\beta} - \frac{N^\alpha x^2}{N} \right), \quad (15.3.1)$$

and the third moment expands to become

$$0 = \lim_{N \rightarrow \infty} \left( \frac{N^\alpha x}{N^{2\beta}} - 3 \frac{N^\alpha x^2}{N^{\beta+1}} + 2 \frac{N^\alpha x^3}{N^2} \right), \quad (15.3.2)$$

For the variance to be positive and finite as  $N$  becomes infinitely large,  $\alpha$  must equal  $\beta$ . If we let  $\alpha = \beta = 1$ , the transformed random variable,  $x = i/N$ , measures the allele frequency,  $p$ , which is a very natural scale. With this choice, the third moment does approach 0 as  $N$  becomes large. Having met the requirements for the moments, the neutral Wright-Fisher model can be approximated by a diffusion model in the transformed state space where we keep track of the allele frequency  $p = i/N$ , and measure time in units of  $N$  generations. With these transformed variables, the drift and diffusion coefficients are  $\mu(p) = 0$ ,  $\sigma^2(p) = p(1-p)$ .

(Although  $\alpha = \beta = 1$  is a natural choice, other choices can obey the requirements of a diffusion approximation. For example,  $\alpha = \beta = 1/2$  can be used to provide a diffusion approximation that more accurately describes the fine fluctuations in allele frequencies while the allele is rare; see Karlin and Taylor 1981, pp. 180–182.)

Selection and mutation can be incorporated into the model by assuming that these processes act deterministically on the large number of propagules produced by the population. For example, the allele frequency after selection followed by mutation in a haploid population would be

$$p' = (1 - v_1) \frac{(1 + s)p}{(1 + s)p + (1 - p)} + v_2 \frac{(1 - p)}{(1 + s)p + (1 - p)} \quad (15.3.3)$$

(continued)

**Box 15.3 (continued)**

where  $s$  is the selective advantage of allele  $A$ ,  $\nu_1$  is the mutation rate from  $A$  to  $a$ , and  $\nu_2$  is the mutation rate from  $a$  to  $A$ . Because we want the diffusion approximation to be consistent with the neutral model when selection and mutation become weak, we will continue to use the transformation with  $m = N$  and  $\alpha = \beta = 1$  (we will show in a moment that the variance remains positive and finite under this choice, as required). Now, the number of copies of allele  $A$  changes by  $E[Y(t + 1) - Y(t) \mid Y(t) = i] = Np' - i = N(p' - p)$ . Plugging this result into (15.2.6), we get

$$\mu(p) = \lim_{N \rightarrow \infty} N(p' - p). \quad (15.3.4)$$

This drift parameter measures the expected change in the transformed variable (the allele frequency) per unit of transformed time (corresponding to  $N$  generations). For the drift parameter to be finite in the limit, the forces causing a change in allele frequency,  $(p' - p)$ , must be weak, on the order of  $1/N$ . To ensure that this restriction is met, let us define  $\psi = sN$ ,  $\theta_1 = \nu_1 N$ , and  $\theta_2 = \nu_2 N$ . In this case, (15.3.3) becomes

$$p' = \left(1 - \frac{\theta_1}{N}\right) \frac{\left(1 + \frac{\psi}{N}\right)p}{\left(1 + \frac{\psi}{N}\right)p + (1 - p)} + \frac{\theta_2}{N} \frac{(1 - p)}{\left(1 + \frac{\psi}{N}\right)p + (1 - p)}. \quad (15.3.5)$$

Factoring  $N(p' - p)$  and taking the limit as  $N$  goes to infinity leaves us with the drift parameter

$$\mu(p) = \lim_{N \rightarrow \infty} N(p' - p) = \psi p(1 - p) - \theta_1 p + \theta_2(1 - p). \quad (15.3.6)$$

The drift parameter is finite only if the selection and mutation rates are small, such that  $\psi = sN$ ,  $\theta_1 = \nu_1 N$ , and  $\theta_2 = \nu_2 N$  remain finite even in the limit as the population size grows to infinity.

Under the above assumptions, the diffusion parameter for the Wright-Fisher model remains  $\sigma^2(p) = p(1 - p)$  even with selection and mutation. To demonstrate this fact, let us find the relationship between  $E[(Y(t + 1) - Y(t))^2 \mid Y(t) = i]$  and the variance of the binomial,  $E[(Y(t + 1) - E[Y(t + 1)])^2 \mid Y(t) = i]$ , which we know equals  $Np'(1 - p')$ . Expanding both expectations using the rules given in Table P3.1, we find that

$$\begin{aligned} E[(Y(t + 1) - Y(t))^2 \mid Y(t) = i] &= E[(Y(t + 1) - E[Y(t + 1)])^2 \mid Y(t) = i] \\ &\quad + E[Y(t + 1) \mid Y(t) = i]^2 - 2E[Y(t + 1) \mid Y(t) = i]i + i^2. \end{aligned} \quad (15.3.7)$$

Because sampling follows a binomial distribution, the expected number of  $A$  alleles after sampling is  $E[Y(t + 1) \mid Y(t) = i] = Np'$ , allowing us to rewrite (15.3.7) as

$$E[(Y(t + 1) - Y(t))^2 \mid Y(t) = i] = Np'(1 - p') + (Np')^2 - 2Np'i + i^2. \quad (15.3.8)$$

Plugging (15.3.5) into (15.3.8) and the result into  $\sigma^2(p) = \lim_{N \rightarrow \infty} (E[(Y(t + 1) - Y(t))^2 \mid Y(t) = i]/N)$ , we again get  $\sigma^2(p) = p(1 - p)$ . Following a similar procedure, the third moment can be shown to equal zero in the limit.

track of the frequency  $p$  of allele  $A$ . With these variables, we find that the drift coefficient is  $\mu(p) = -N\nu_1 p + N\nu_2(1-p) + Nsp(1-p)$  and the diffusion coefficient is  $\sigma^2(p) = p(1-p)$ , where  $s$  is the selective advantage of allele  $A$ ,  $\nu_1$  is the mutation rate from  $A$  to  $a$ , and  $\nu_2$  is the mutation rate from  $a$  to  $A$ .

With these drift and diffusion coefficients, we can follow Recipe 15.1 to find the stationary distribution. Evaluating  $A(p)$  first, we have  $A(p) = 2N\nu_1 \ln(1-p) + 2N\nu_2 \ln(p) + 2Nsp$  and  $e^{A(p)} = (1-p)^{2N\nu_1} (p)^{2N\nu_2} e^{2Nsp}$ . Substituting these into expression (15.18), we must next choose the constants of integration,  $c_1$  and  $c_2$ . It is always worth trying to identify a stationary distribution by first setting  $c_1 = 0$ , as this eliminates the integral in (15.18). If this choice fails to identify the stationary distribution of interest, other choices of  $c_1$  can then be made. (The choice of  $c_1 = 0$  in the Wright-Fisher model can also be formally motivated, see p. 222 of Karlin and Taylor 1981.) Setting  $c_1$  to zero, we immediately find the stationary distribution from (15.18):

$$\pi(p) = c_2 (1-p)^{2N\nu_1-1} (p)^{2N\nu_2-1} e^{2Nsp}$$

where

$$c_2 = 1 / \left( \int_{p=0}^1 (1-p)^{2N\nu_1-1} (p)^{2N\nu_2-1} e^{2Nsp} dp \right). \quad (15.25)$$

The stationary distribution (15.25) describing allele frequencies in the presence of selection and mutation is known as Wright's distribution in honor of its discoverer, the evolutionary biologist Sewall Wright (see Crow and Kimura 1970 for the equivalent distribution in a diploid population). It is illustrated for various parameters in Figure 15.4. In the absence of selection ( $s = 0$ ), equation (15.25) represents the probability density function for a beta distribution (see definition P3.16).

### 15.3.3 Probability of Absorption in a Diffusion Model

Let us now derive the probability that the stochastic processes reaches a particular point,  $b$ . We assume that the point of interest,  $b$ , is the maximum (or minimum) state that can be reached by the stochastic process and that once this state is reached, the stochastic process remains there (i.e.,  $x = b$  is an "absorbing boundary"). While the forward equation is particularly helpful in deriving the stationary distribution, the backward equation is more useful for deriving the absorption probability.

We define  $u(b, t | x_0, t_0)$  as the probability of reaching  $b$  by time  $t$  from initial position  $x_0$  at  $t_0$ . Because we have assumed that  $b$  is an absorbing boundary,  $u(b, t | x_0, t_0)$  equals the probability that the process makes a transition to  $b$  over this time period, i.e.,  $\phi(b, t | x_0, t_0)$ . Thus, we can replace  $\phi$  with  $u$  in the backward Kolmogorov equation (15.11):

$$\frac{\partial u(b, t | x_0, t_0)}{\partial t_0} = -\mu(x_0) \frac{\partial u(b, t | x_0, t_0)}{\partial x_0} - \frac{1}{2} \sigma^2(x_0) \frac{\partial^2 u(b, t | x_0, t_0)}{\partial x_0^2}, \quad (15.26)$$

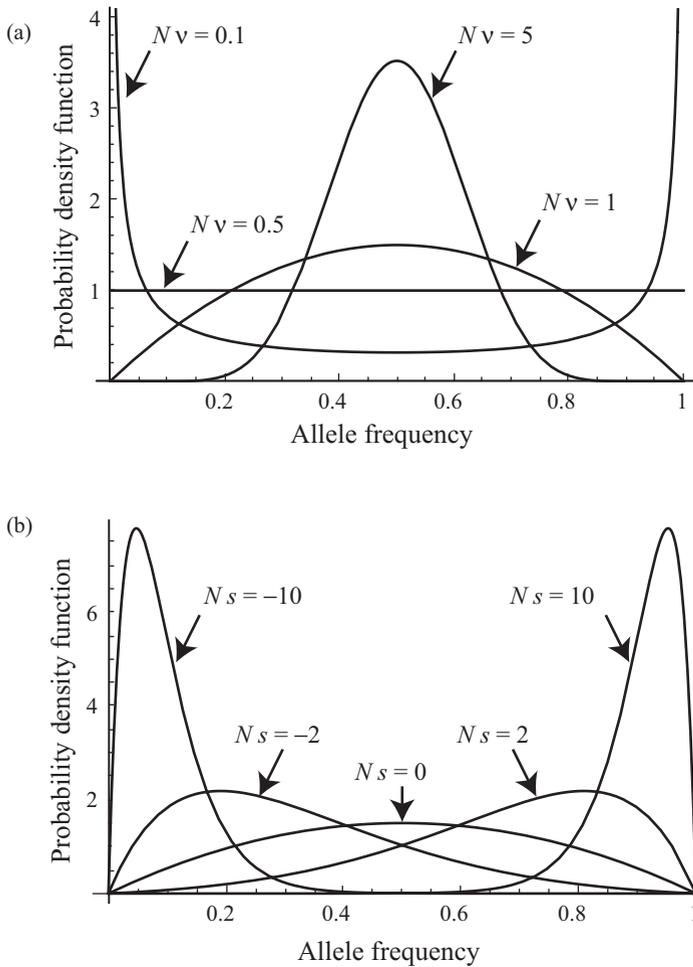


Figure 15.4: Wright's distribution. Wright's distribution describes the frequency of allele  $A$  based on a diffusion approximation of the Wright-Fisher model. The stationary probability density function (15.25) is plotted as a function of the allele frequency, assuming equal forward and backward mutation rates ( $\nu_1 = \nu_2 = \nu$ ). (a) In the absence of selection ( $Ns = 0$ ), the lower the mutation rate, the more likely the allele is nearly lost or fixed within the population. (b) In the presence of selection with a mutation rate of  $N\nu = 1$ , the more strongly selection favors allele  $A$ , the more frequent the  $A$  allele is likely to be. *Mathematica* was used for numerical integration. Notice that an increase in the population size is equivalent to an increase in both mutation and selection.

which considers the target point  $b$  to be held constant and the initial points to be varied,  $u(b, t | x_0, t_0)$ .

After a long amount of time has passed,  $u(b, t | x_0, t_0)$  typically approaches a constant value that depends only on its initial position. Let's define  $u(b | x_0) = \lim_{t \rightarrow \infty} u(b, t | x_0, t_0)$  as the probability of ultimately reaching  $b$  from initial position  $x_0$ . Because  $u(b | x_0)$  does not depend on time,  $\partial u(b | x_0) / \partial t_0 = 0$ , and the backward equation becomes an ordinary differential equation:

$$0 = -\mu(x_0) \frac{du(b | x_0)}{dx_0} - \frac{1}{2} \sigma^2(x_0) \frac{d^2u(b | x_0)}{dx_0^2}. \quad (15.27a)$$

To solve (15.27a), it is again helpful to change notation to make the calculations more transparent. In particular, if we define  $y \equiv du(b | x_0) / dx_0$  then (15.27a) can be written as

$$0 = -\mu(x_0) y - \frac{1}{2} \sigma^2(x_0) \frac{dy}{dx_0}. \quad (15.27b)$$

Equation (15.27b) is now an ordinary differential equation that can be solved using a separation of variables (Recipe 6.2). Placing terms involving  $y$  on the left and those involving  $x_0$  on the right, we can integrate both sides to get

$$\begin{aligned}\int \frac{1}{y} dy &= - \int \frac{2\mu(x_0)}{\sigma^2(x_0)} dx_0 \\ &= -A(x_0)\end{aligned}\quad (15.28)$$

where again  $A(x_0) = \int 2\mu(x_0)/\sigma^2(x_0) dx_0$ . Performing the integration on the left-hand side, we have

$$\ln(y) = -A(x_0) + c_0, \quad (15.29)$$

where  $c_0$  is a constant of integration. We can then exponentiate both sides, replace  $y$  with  $du(b|x_0)/dx_0$ , and rename the constant  $e^{c_0}$  as  $c_1$  to obtain the differential equation

$$\frac{du(b|x_0)}{dx_0} = e^{-A(x_0)} c_1. \quad (15.30)$$

Performing another separation of variables, the solution is

$$\begin{aligned}u(b|x_0) &= \int e^{-A(x_0)} c_1 dx_0 + c_2 \\ &= c_1 S(x_0) + c_2,\end{aligned}\quad (15.31)$$

where  $c_2$  is another constant of integration and

$$S(x) = \int e^{-A(x)} dx. \quad (15.32)$$

To solve for the constants of integration, we assume that the random variable  $x_0$  ranges from  $a$  to  $b$  and that both  $a$  and  $b$  are absorbing boundaries. If we start the system at the absorbing boundary  $a$ , it will never reach state  $b$ . This implies that  $u(b|a) = 0$  and, from equation (15.31), that  $c_2 = -c_1 S(a)$ . Substituting this requirement into equation (15.31) gives

$$u(b|x_0) = c_1 (S(x_0) - S(a)). \quad (15.33)$$

Conversely, if we start in state  $b$  then the probability of hitting state  $b$  is one; i.e.,  $u(b|b) = 1$ . Using this fact in equation (15.33), we can solve for  $c_1 = 1/(S(b) - S(a))$ . Thus the probability that the system eventually reaches the point  $b$  is

$$u(b|x_0) = \frac{S(x_0) - S(a)}{S(b) - S(a)}. \quad (15.34)$$

Because the difference between two indefinite integrals,  $S(x) - S(y)$ , equals the definite integral evaluated from  $y$  to  $x$ , we can write equation (15.34) in two equivalent forms:

### Recipe 15.2

#### The Probability of Absorption in a Diffusion Model

The probability that a diffusion process that begins in state  $x_0$  ultimately reaches the absorbing state  $b$  (the upper limit to the range) rather than the absorbing state  $a$  (the lower limit) is

$$u(b | x_0) = \frac{S(x_0) - S(a)}{S(b) - S(a)}, \quad (15.35a)$$

$$= \frac{\int_{x=a}^{x_0} e^{-A(x)} dx}{\int_{x=a}^b e^{-A(x)} dx}, \quad (15.35b)$$

where  $A(x) = \int 2\mu(x)/\sigma^2(x) dx$  and  $S(x) = \int e^{-A(x)} dx$ . By the complement rule (Rule P3.2), the probability that the system ultimately hits  $a$  instead is

$$u(a | x_0) = 1 - u(b | x_0) = \frac{S(b) - S(x_0)}{S(b) - S(a)} \quad (15.36a)$$

$$= \frac{\int_{x=x_0}^b e^{-A(x)} dx}{\int_{x=a}^b e^{-A(x)} dx}. \quad (15.36b)$$

Recipe 15.2 is extremely useful and can be used for any diffusion model with absorbing states. Even when the integrals cannot be evaluated explicitly, equations (15.35) and (15.36) can be numerically integrated to obtain the absorption probabilities.

#### Example: Modeling Individual Movement in a Continuous Spatial Habitat

Let us return to our fish example and suppose that the drift and diffusion coefficients are the constants  $\mu$  and  $\sigma^2$ , where the drift term is positive (reflecting the fact that the fish tends to move upstream more often than downstream). Suppose you are monitoring only a portion of the stream, from a location labeled 0 downstream to a location labeled  $b$  upstream. Once the fish leaves this part of the stream, you stop monitoring. What is the probability that the fish leaves the region at the downstream end (point 0) instead of the upstream end?

We can answer this question using Recipe 15.2. First we must calculate  $A(x) = \int (2\mu/\sigma^2) dx$ . Because the drift and diffusion coefficients are constants,  $A(x) = 2\mu x/\sigma^2$ . We can save some time if we evaluate  $S(x) = \int e^{-A(x)} dx$  next, because we can then use equation (15.36a) without having to perform any further integrations. Carrying out this integration, we have  $S(x) = -(\sigma^2/2\mu) e^{-2\mu x/\sigma^2}$ . Consequently, the probability that the fish leaves at the downstream end of the monitored region is

$$\begin{aligned} u(0 | x_0) &= \frac{S(b) - S(x_0)}{S(b) - S(0)} \\ &= \frac{-\frac{\sigma^2}{2\mu} e^{-2\mu b/\sigma^2} + \frac{\sigma^2}{2\mu} e^{-2\mu x_0/\sigma^2}}{-\frac{\sigma^2}{2\mu} e^{-2\mu b/\sigma^2} + \frac{\sigma^2}{2\mu}} \\ &= \frac{-e^{-2\mu b/\sigma^2} + e^{-2\mu x_0/\sigma^2}}{-e^{-2\mu b/\sigma^2} + 1}. \end{aligned} \quad (15.37)$$

Equation (15.37) answers our question, but we can simplify it even further if we suppose that the fish's rate of movement is slow relative to length of the stream. In this case we might expect the drift coefficient to be small, and we can expand (15.37) as a Taylor series in  $\mu$ :

$$u(0 | x_0) = \frac{b - x_0}{b} - \frac{x_0(b - x_0)}{b \sigma^2} \mu. \quad (15.38)$$

Equation (15.38) reveals that, in the absence of any directional movement ( $\mu = 0$ ) the probability of leaving the region at the downstream end is given by where the fish starts, measured as a proportion of the distance from the original position to the downstream end:  $(b - x_0)/b$ . If the fish has a small tendency to move upstream, however, this exit probability is decreased by an amount  $x_0(b - x_0)/(b \sigma^2) \mu$ . The larger the variance in movement, the smaller this directional effect is, because the fish is likely to exit sooner as a result of random diffusive movement, before substantial directional movement has taken place.

While framed in terms of the movement of the fish, equations (15.37) and (15.38) apply to any model of diffusion with constant drift and diffusion coefficients (i.e., to any Brownian motion model) along a single axis. Generalizations to movement in more than one dimension are given by Karlin and Taylor (1981).

### Example: Fixation in the Wright-Fisher Model

As a second example, let us calculate the probability that an allele becomes fixed within a haploid population using the Wright-Fisher model with selection but no mutation ( $\nu_1 = 0, \nu_2 = 0$ ). From Box 15.3, the drift and diffusion parameters are  $\mu(p) = \psi p(1 - p)$  and  $\sigma^2(p) = p(1 - p)$ , where  $p$  represents the allele frequency,  $N$  is the size of the haploid population, and  $\psi = Ns$  (see Problem 15.7 for the equivalent result in diploids). With these drift and diffusion

parameters,  $A(p) = 2\psi p$  and  $S(p) = \int e^{-2\psi p} dp = -1/(2\psi) e^{-2\psi p}$ . According to equation (15.35a), the probability of fixation of an allele whose initial frequency is  $p_0$  is then

$$\begin{aligned} u(1|p_0) &= \frac{S(p_0) - S(0)}{S(1) - S(0)} \\ &= \frac{-e^{-2Ns p_0} + 1}{-e^{-2Ns} + 1} \end{aligned} \quad (15.39)$$

The diffusion approximation for the probability of fixation (15.39) is a classic result in evolutionary biology (Kimura 1957, 1962). In Figure 15.5, we compare the diffusion approximation (15.39) to exact numerical results for the probability of fixation, which can be obtained using (14.29) for small to moderately sized populations. Even though the diffusion approximation technically assumes that the population size is large and that selection is weak (see Box 15.3), the diffusion provides a remarkably good approximation to the fixation probability even for very small populations.

In the absence of selection and mutation, the fixation probability can be found from (15.39) by taking the limit as  $s$  goes to zero using L'Hôpital's rule (A2.30), in which case  $u(1|p_0) = p_0$ . This shows that the fixation probability equals the initial allele frequency, as we might expect when all alleles are equally fit. It is instructive, however, to rederive the fixation probability by noting that  $A(p)$  is zero whenever  $\mu(p)$  is zero, so that  $S(p)$  becomes  $p$ . The fixation probability is then easy to solve using equation (15.35a) and again gives  $u(1|p_0) = (p_0 - 0)/(1 - 0) = p_0$ . This makes it clear that the fixation probability will equal  $p_0$  whenever  $\mu(p)$  is zero, regardless of the amount of variation caused by chance events during reproduction (as measured by the diffusion coefficient).

### 15.3.4 Time to Absorption in a Diffusion Model

While a stochastic model eventually becomes absorbed whenever absorbing states exist and can be reached, whether absorption occurs slowly or quickly

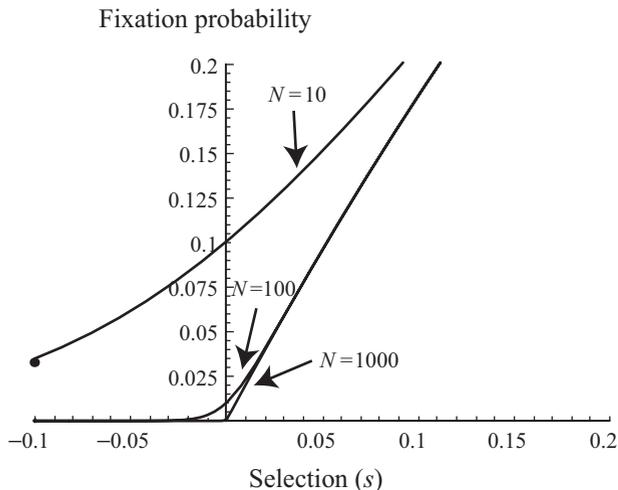


Figure 15.5: The probability of fixation in the Wright-Fisher model. Starting from a single allele ( $p_0 = 1/N$ ), the probability that the allele ultimately fixes within a population is shown as a function of its selection coefficient,  $s$ . The diffusion approximation using (15.39) (curves) is compared to exact numerical results using (14.29) for  $N = 10$  and  $N = 100$ .

cannot be determined from the absorption probability,  $u(b|x_0)$ . To determine the relevance of the absorbing states for any biological model, we need to know the time frame over which absorption occurs. Does the fixation of an allele or the extinction of a population tend to occur over the course of a few years or millions of years? We derive the expected waiting time to absorption in Supplementary Material 15.2 (it is similar in flavor to the derivation of the absorption probability) and summarize the results here.

### Recipe 15.3

#### Waiting Time until Absorption in a Diffusion Model

Suppose that there are two absorbing state labeled  $a$  and  $b$ . If the system begins in state  $x_0$ , the expected waiting time until absorption in state  $b$ , conditioned upon reaching state  $b$ , is

$$\bar{t}_b(x_0) = 2(S(b) - S(a)) \left( \int_{y=x_0}^b \frac{u(b|y)(1 - u(b|y))}{e^{-A(y)} \sigma^2(y)} dy + \frac{(1 - u(b|x_0))}{u(b|x_0)} \int_{y=a}^{x_0} \frac{u(b|y)^2}{e^{-A(y)} \sigma^2(y)} dy \right). \quad (15.40)$$

Similarly, the expected waiting time until absorption in state  $a$ , conditioned upon reaching state  $a$ , is

$$\bar{t}_a(x_0) = 2(S(b) - S(a)) \left( \frac{(1 - u(a|x_0))}{u(a|x_0)} \int_{y=x_0}^b \frac{u(a|y)^2}{e^{-A(y)} \sigma^2(y)} dy + \int_{y=a}^{x_0} \frac{u(a|y)(1 - u(a|y))}{e^{-A(y)} \sigma^2(y)} dy \right). \quad (15.41)$$

The expected time until absorption in any state,  $\bar{t}(x_0)$ , is the average of the above two equations, each weighted by the probability of absorption in the corresponding state:  $\bar{t}(x_0) = u(b|x_0) \bar{t}_b(x_0) + u(a|x_0) \bar{t}_a(x_0)$  (see Rule P3.9). Using the fact that the fixation probabilities sum to one,  $u(a|x_0) + u(b|x_0) = 1$ , the expected waiting time until absorption in either state is given by

$$\bar{t}(x_0) = 2(S(b) - S(a)) \left( u(b|x_0) \int_{y=x_0}^b \frac{(1 - u(b|y))}{e^{-A(y)} \sigma^2(y)} dy + (1 - u(b|x_0)) \int_{y=a}^{x_0} \frac{u(b|y)}{e^{-A(y)} \sigma^2(y)} dy \right). \quad (15.42)$$

If there is only one absorbing state in the model, the expected waiting time until absorption is

$$\bar{t}(x_0) = 2 \int_{y=a}^{x_0} e^{-A(z)} \int_{y=z}^b \frac{e^{A(y)}}{\sigma^2(y)} dy dz, \quad (15.43)$$

where we assume that all processes are eventually absorbed in state  $a$  and that state  $b$  is a (potentially infinite) upper boundary that is either unattainable or rapidly left (see Karlin and Taylor 1981; Lande 1993).

**Example: Expected Time to Absorption in the Wright-Fisher Model**

Let us apply Recipe 15.3 to the Wright-Fisher model of allele frequency change. We focus on the case where there is no selection or mutation, in which case there are two absorbing states at  $a = 0$  (allele loss) and  $b = 1$  (allele fixation). The drift and diffusion coefficients are given by  $\mu(p) = 0$  and  $\sigma^2(p) = p(1 - p)$  (see Box 15.3). The first step is to calculate the various quantities needed in the waiting time; specifically, we have  $A(p) = 0$ ,  $S(p) = p$ , and  $u(1 | p_0) = p_0$ . Plugging these results into (15.40) and integrating (see Problem 15.5), we find that the average time until fixation starting from an initial allele frequency  $p_0$  is

$$\bar{t}_1(p_0) = -2 \frac{1 - p_0}{p_0} \ln(1 - p_0). \quad (15.44)$$

The unit of time in (15.44) is measured in whatever time scale is required for the diffusion approximation to hold. In Box 15.3, the appropriate scale for the Wright-Fisher model measured time in units of  $N$  generations in a haploid population. Thus, we must multiply (15.44) by  $N$  to obtain the fixation time in generations. (This result also applies to a diploid population if we replace  $N$  with  $2N$ .) The result is plotted in Figure 15.6 as a function of the initial allele frequency. The fixation time is on the order of  $N$  generations and drops only when the initial allele frequency is very near one. This reflects the fact that allele frequencies vary so much under random genetic drift that the initial allele frequency is not very predictive of the exact time at which fixation occurs.

Of particular interest is the time until fixation for an allele that appears in a single copy ( $p_0 = 1/N$ ). Taking the Taylor series of (15.44) with respect to  $p_0$  using Recipe P1.3, we find that the mean time until fixation is, to leading order,  $2N$  generations in a haploid population ( $4N$  in a diploid population). Interestingly, this result is consistent with the time that it takes for a population to coalesce down to a single individual (see equation (13.17) with  $n$  large). This parallel makes sense; looking either forward in time or backward in time, the fixation time and the coalescence time measure how long it takes, on average, until a single ancestor gives rise to every member of a population. The waiting time

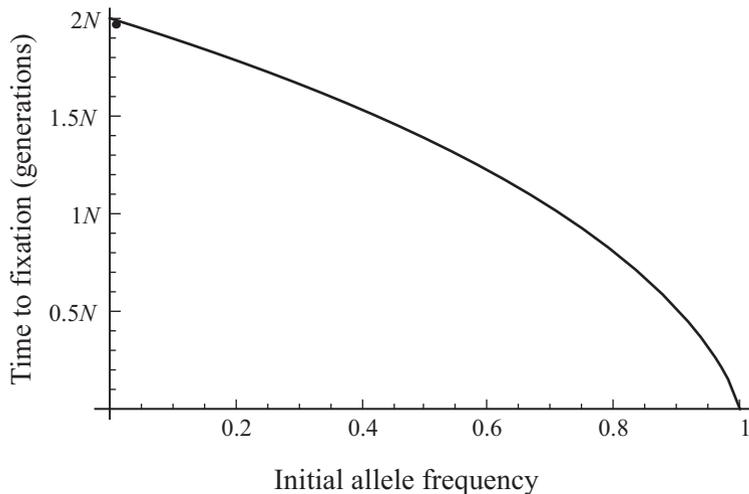


Figure 15.6: Fixation time for neutral alleles in the Wright-Fisher model. The expected time to fixation of a neutral allele is plotted as a function of the initial allele frequency. The waiting time is averaged across only those cases in which the neutral allele ultimately fixes. The solid curve is based on the diffusion approximation (15.44), while the dots give the exact waiting time for a population of size  $N = 100$  using equation (14.48).

until fixation is slightly longer than the coalescent time because the fixation time includes any periods during which the stochastic process leaves and later returns to the state with only a single ancestor ( $p = 1/N$ ) whereas the coalescent time excludes these periods.

In the presence of selection in the Wright-Fisher model, we previously found the probability of fixation is given by (15.39) and  $A(p) = 2\psi p$ , where  $\psi = Ns$ . Plugging these values into (15.40) results in an integral that cannot be evaluated for general values of the parameters. Nevertheless, we can integrate the equation numerically using software such as *Mathematica* (Figure 15.7). One interesting finding is that the time until fixation is the same whether the selection coefficient  $s$  is positive or negative (Ewens 1979). At first, this result makes no sense. Clearly, if an allele is favorable, it should fix faster when selection is stronger. What is less obvious is that if an allele is disfavored, it also tends to fix faster when selection against it is stronger. The reason is that there must be large and rapid chance increases in the allele frequency for a disfavored allele to rise to fixation in the face of strong selection against it. Such chance increases are unlikely, and a disfavored allele has a low fixation probability. But given that a disfavored allele does fix, it does so over the same average time frame as a favored allele experiencing selection of the same magnitude.

A similar result occurs in the birth-death model of a population at risk of extinction. The average time until a population becomes established (i.e.,

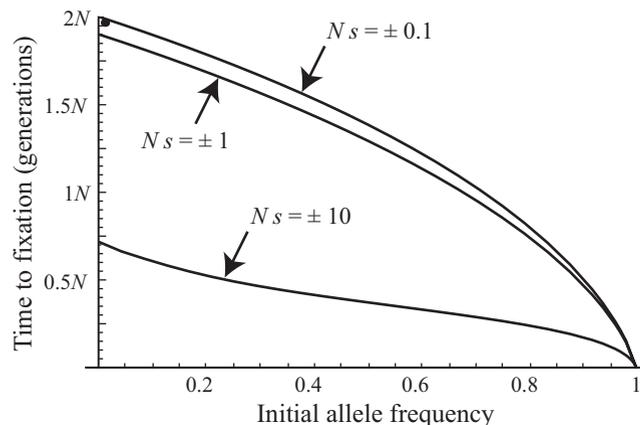


Figure 15.7: Fixation time for selected alleles in the Wright-Fisher model. The expected time to fixation of a selected allele is plotted as a function of the initial allele frequency of  $A$  ( $p_0$ ). The waiting time is averaged across those cases in which the  $A$  allele ultimately fixes. The solid curve is based on numerical integration of the diffusion approximation (15.40), and the dots give the exact waiting time for a population of size  $N = 100$  using equation (14.48). Parameters used were  $Ns = -10, -1, -0.1, 0.1, 1,$  and  $10$ . Surprisingly, the waiting time for a deleterious allele  $A$  to fix cannot be distinguished from the waiting time for a beneficial allele, given the same magnitude of selection. (For the exact numerical results, the fitnesses were set to  $W_o = 1, W_A = 1 + s$  when  $A$  was beneficial and to  $W_o = 1 + s, W_A = 1$  when  $A$  was deleterious. This choice keeps the strength of selection equivalent in the two cases.)

reaches a threshold size  $m$ ) is the same whether the birth rate is greater than the death rate or vice versa (Figure 15.8a). The same is true for the average time until extinction (Figure 15.8b). Of course, if deaths are more common than births, the most likely fate of the population is extinction (see Problem 15.6). But, assuming that the population has risen from some initial size to  $m$ , there is no way to know for sure whether the population truly had a higher birth rate than death rate or whether there happened, by some small chance, to be more births than deaths during the monitoring period despite the fact that the expected birth rate is less than the death rate.

The above results depend critically on the assumption that there are two absorbing boundaries. For example, compare Figure 15.8 to Figure 15.9, which is based on the same birth-death model, except that we now impose a “hard carrying capacity” on the population at  $K$ . Specifically, when  $i = K$ , the probability of a birth drops to zero. This model has only one absorbing boundary: extinction of the population. As shown in Figure 15.9, the average time until extinction based on (15.43) now rises dramatically when the birth rate exceeds the death rate.

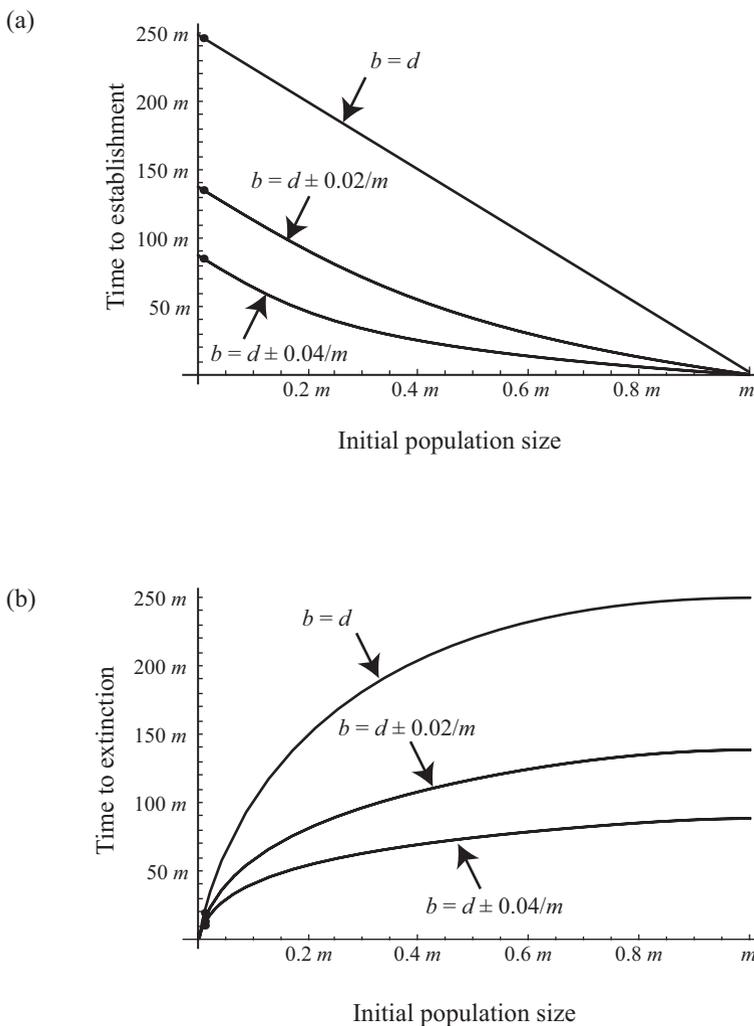
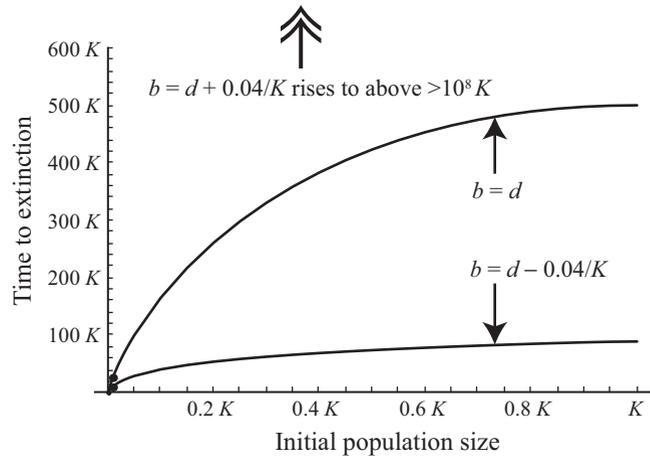


Figure 15.8: Expected waiting time until extinction or recovery. The expected waiting time until an at-risk population becomes (a) established or (b) extinct is plotted as a function of the initial population size. Population dynamics were assumed to follow a simple birth-death process (section 14.4). Using the drift and diffusion coefficients from Problem 15.6, the solid curves are based on the diffusion approximations (15.40) and (15.41), and the dots give the average waiting time using the exact equation (14.48) with  $m = 100$ . The horizontal axis represents the initial population size, measured as a fraction of the threshold size,  $m$ . The vertical axis represents the waiting time, scaled by  $m$ . The total probability of an event per individual per time step was held constant at  $b + d = 0.004$ , so that a single individual undergoes, on average, one birth or death every 250 time steps, with  $b = d$  (highest curves),  $b = d \pm 0.02/m$  (middle curves), and  $b = d \pm 0.04/m$  (lowest curves). The waiting time depends on whether births and deaths differ in frequency but not on which is more frequent.

Figure 15.9: The average waiting time until extinction of a population with a hard carrying capacity of  $K$ . Population dynamics were assumed to follow the simple birth-death process (section 14.4) with a hard carrying capacity of  $K$ . The solid curves are based on the diffusion approximation (15.43) with drift and diffusion coefficients given in Problem 15.6, and the dots give the average waiting time for  $K = 100$  using the exact equation (14.38). Again,  $b + d = 0.004$ , with  $b = d + 0.04/K$  (highest curve rises to above  $10^8 K$  time steps and is off the scale of the figure),  $b = d$  (middle curve), and  $b = d - 0.04/K$  (lowest curve).



## 15.4 Modeling Populations in Space Using the Diffusion Equation

In section 15.2.1, we introduced a stochastic model of fish movement and showed that a diffusion model could be used to describe the probability that an individual fish is found at position  $x$  at time  $t$ . What if we were tracking an entire population of fish, each of which moves stochastically over space? Provided that the movement of each individual is independent of all others and provided that there is a large number of individuals in the population, then  $\phi(x, t | x_0, t_0)$  can be thought of as describing the *fraction* of a population currently found at position  $x$  (e.g., the distribution of a population of fish over space). The requirement that the population be large follows from the frequency interpretation of a probability (Primer 3); we can interpret the probability of an event as the fraction of times it occurs in a very large number of trials. Interestingly, under this interpretation, the stochastic equations describing individual movement provide a *deterministic* description of the entire population over space.

### Example: Sperm Movement in a Reproductive Tract

The reproductive tract of females in many species can be viewed as a one-dimensional structure through which sperm travel to reach and fertilize an egg. Typically there are many millions of sperm transferred to a female during a mating event, and therefore it is probably not unreasonable to suppose that the number at any given location is always very large. Furthermore, it seems reasonable to model the movement of each spermatozoa as being stochastic, using a diffusion model with drift.

Let us suppose that the drift rate and the variance are constant throughout the reproductive tract, and denote them by  $\mu$  and  $\sigma^2$ . Thus,  $\mu$  represents the average direction of movement of the individual sperm at a particular location, and  $\sigma^2$  represents the variance in the movement of an individual. We index each location along the reproductive tract by the distance to the egg and suppose that the egg is located at  $x = b$ . We will view both ends of the reproductive

tract as being absorbing boundaries with the rationale that sperm that exit the tract at  $x = 0$  are lost, and sperm that reach the egg ( $x = b$ ) attempt to fertilize the egg. We can then use Recipe 15.2 to calculate the probability that any given sperm ends up at the egg (i.e., absorption at  $x = b$ ). Alternatively, given the assumption that each of the millions of sperm move independently, this Recipe can also be interpreted as giving us the proportion of the sperm that end up at the egg.

In fact, the calculations of  $A(x)$  and  $S(x)$  are identical to those for the model of fish movement:  $A(x) = (2\mu/\sigma^2) \int 1 dx = 2\mu x/\sigma^2$  and  $S(x) = -(\sigma^2/2\mu) e^{-2\mu x/\sigma^2}$ . Therefore, from equation (15.35a), the proportion of the sperm that make it to the egg is

$$\begin{aligned} u(b | x_0) &= \frac{S(x_0) - S(0)}{S(b) - S(0)} \\ &= \frac{-e^{-2\mu x_0/\sigma^2} + 1}{-e^{-2\mu b/\sigma^2} + 1} \end{aligned} \quad (15.45)$$

where  $x_0$  is the initial location of all the sperm, which we assume to be nearer 0 than  $b$ . Equation (15.45) reveals that, unless there is a large amount of directionality in movement (i.e., large  $\mu$ ), a very small proportion of the sperm actually make it to the egg (Figure 15.10). This might be one reason why such large numbers of sperm are transferred during mating. Indeed, it has been hypothesized that the very long and intricate reproductive tracts of some animals have evolved to select vigorous sperm with the greatest motility (Parker 1970).

#### Example: The Drift Paradox

Let us return to our model of movement in a stream, but now we consider a population of aquatic invertebrates (e.g., the aquatic stage of mayflies). Because invertebrates are typically quite small, it is reasonable to suppose that the population size at any given location in the stream is quite large. Thus, if each individual moves independently and stochastically, we can interpret a model of diffusion with drift as giving the dynamics of the number of individuals in each location.

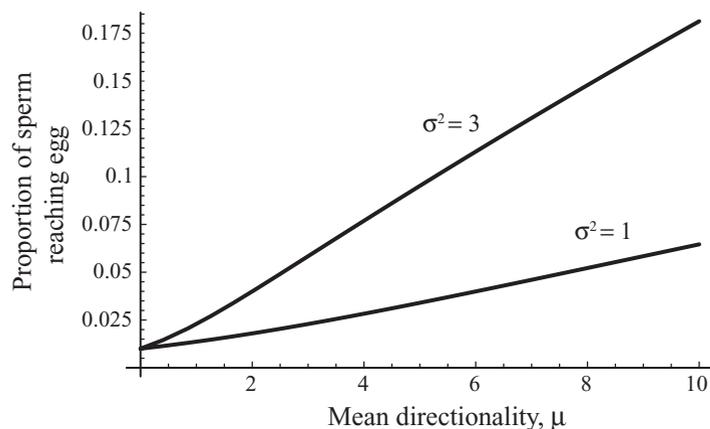


Figure 15.10: Proportion of sperm reaching an egg. It was assumed that sperm start out 1/100 of the way along the reproductive tract ( $x_0 = 0.01$ ,  $b = 1$ ) and travel according to a diffusion model with drift coefficient  $\mu$  and diffusion coefficient  $\sigma$ .

Let us keep the model as general as possible and suppose that the average direction of movement of individuals at location  $x$ , as well as the variance in their movement, depends on the location in the stream. Assuming that movement occurs continuously over time and space, the population size at each location, denoted by  $n(x)$ , obeys the diffusion equation with drift (15.8); that is,

$$\frac{\partial n(x, t)}{\partial t} = -\frac{\partial(\mu(x) n(x, t))}{\partial x} + \frac{1}{2} \frac{\partial^2(\sigma^2(x) n(x, t))}{\partial x^2} \quad (15.46)$$

To complete the model we need to specify what happens to the population at the boundaries  $x = b$  (taken to be the upper reaches of the stream) and  $x = 0$  (taken to be where the stream empties into a lake or ocean). It is reasonable to suppose that there is a reflecting boundary at the top of the stream (i.e., individuals cannot go past this point but they are not “absorbed” there either) and that there is an absorbing boundary at  $x = 0$  (the invertebrates exit the stream).

What does the above model predict for the dynamics of the invertebrate population? Our results from section 15.3.3 on the probability of absorption reveal that the population of invertebrates eventually go extinct in the stream because there is only one absorbing state. In other words, eventually 100% of the population of invertebrates will be “absorbed” at state  $x = 0$ . This finding is referred to as the “drift paradox” (Speirs and Gurney 2001). Clearly populations of invertebrates exist in streams, so how are they able to persist?

There are several resolutions to this paradox, but two possibilities stand out. The first is that, even if the invertebrate population eventually disappears, the time to extinction might be exceedingly long. The second is that we have neglected an important feature of the biology of invertebrate populations: reproduction! To remain faithful to the stochastic underpinnings of the original model, we must model not only how each individual moves stochastically over time, but also how they interact and reproduce with other individuals and how they die.

Constructing a complete stochastic model for the above processes is a difficult enterprise, and simplifications are often made. One of the most common is to interpret the stochastic model of movement as a deterministic description of a large population, to which we add further deterministic components to describe other processes of interest.

As an example, reproduction can be included in model (15.46) by supposing that the birth rate of new invertebrates at location  $x$  is given by some function,  $f(n(x, t))$ :

$$\frac{\partial n(x, t)}{\partial t} = f(n(x, t)) - \frac{\partial(\mu(x) n(x, t))}{\partial x} + \frac{1}{2} \frac{\partial^2(\sigma^2(x) n(x, t))}{\partial x^2}. \quad (15.47)$$

Ignoring the last two terms describing movement,  $\partial n(x, t)/\partial t = f(n(x, t))$  is akin to the deterministic models of population growth considered in Chapter 3 (e.g., exponential or logistic growth). This seemingly small addition, however, brings

with it major changes in behavior. In all stochastic models of movement considered so far, individuals were neither created nor destroyed (although they might “pile up” at a boundary). In contrast, equations such as (15.47) now allow the population to grow or shrink over time. The analysis of spatial diffusion equations with varying population sizes is considerably more difficult and beyond the scope of this book (interested readers should consult Britton 1986; Edelman-Keshet 1988; Kot 2001; Okubo and Levin 2001).

Interestingly, analyses of the drift paradox have shown that reproduction does not guarantee the persistence of aquatic invertebrates unless the rate of reproduction is sufficiently high (Speirs and Gurney 2001). Alternative explanations for the persistence of stream dwelling populations have also been explored using diffusion models, including the importance of zones where drift and diffusion rates are minimal, such as along the edge of a river (Pachepsky et al. 2004).

#### **Example: Hybrid Zones**

In Supplementary Material 15.3, we illustrate how a diffusion approach can be extended to explore the dynamics of evolutionary change in a spatial setting. We consider a model of how the frequencies of two types vary across a “hybrid zone,” which is a region in space where two types meet and produce less fit hybrids. The model again assumes that the population is large at every point in space, so that a diffusion equation can be used to describe the effects of dispersal on the population distribution over space. Selection is then incorporated as an additional deterministic process, much as population growth was added to equation (15.47). In the case of a hybrid zone, it is possible to make substantial analytical progress, and we describe results on the shape of the hybrid zone when a balance is reached between dispersal and selection.

## **15.5 Concluding Message**

In this chapter, we have introduced various methods for constructing and analyzing stochastic models with continuous state spaces, focusing on a diffusion model with drift. Diffusion models have played an important role in many areas of science, including biology. For example, they have been used to describe the diffusion of retroviruses used in gene therapy (Chuck et al. 1996), the diffusion of water through normal and injured brain tissue (Melhem 2002), and changes in lizard morphology over evolutionary time (Schluter et al. 1997). They have also been used to estimate the time until extinction of populations in the presence of demographic and/or environmental stochasticity (Lande 1993) and to assess the minimum population size needed to reduce the risk of extinction during a time period of interest (e.g., less than 1% extinction risk over 1000 years). Such analyses have played an important role in scientific assessments of the level of risk facing endangered species (Clegg 1995). From this small selection of applications alone, it is clear that diffusion methods form an important group of mathematical tools in biology.

## Problems

- Problem 15.1:** For the model of fish movement in section 15.2.1 derive the probability that the fish exits the monitored region at the downstream end, using the results for the birth-death model from Chapter 14. Compare your result to equation (15.37) for the continuous-space model.
- Problem 15.2:** Use the techniques of Box 15.2 to derive a diffusion approximation for the Moran model for a neutral allele. (a) Show that the choices  $\alpha = 2$  and  $\beta = 1$  are appropriate scaling parameters. This implies that one time unit in the diffusion approximation corresponds to  $N^2$  time units in the original model. (b) Calculate the drift and diffusion coefficients from (15.2.6). (c) Use your results from (b) in the forward Kolmogorov equation (15.8) to derive a diffusion approximation for the Moran model. (d) Show that your result is consistent with (15.14) by rescaling time.
- Problem 15.3:** For the Brownian motion model, verify that the general solution (15.15) satisfies (a) the forward Kolmogorov equation (15.8) and (b) the backward Kolmogorov equation (15.11).
- Problem 15.4:** (a) Following analogous calculations to those in equation (15.22), show that the rate of change of the  $n$ th moment for the adaptive dynamics model specified by equations (15.19) and (15.21) is given by  $\eta \mathcal{M}_n - (\eta(N-1)s/a_K) \mathcal{M}_{n+1}$ , where  $\mathcal{M}_n$  denotes the  $n$ th central moment of the mutational distribution, which is assumed to be normal:  $e^{-(s_m - s)^2/(2v)}/\sqrt{2\pi v}$ . (b) Use central moment generating functions (Appendix 5) to show that all odd moments of a normal distribution are zero. (c) Use central moment generating functions to show that the even moments of a normal distribution are proportional to the variance raised to a power that increases for higher and higher moments. (d) Use the results from (a) to (c) to argue that all higher moments of the adaptive dynamics model will therefore be negligible provided that the mutational variance  $v$  is very small.
- Problem 15.5:** In this problem, we examine the diffusion approximation to the Wright-Fisher model of allele frequency change in haploid population in the absence of selection and mutation. Use the drift and diffusion parameters  $\mu(p) = 0$  and  $\sigma^2(p) = p(1-p)$  from Box 15.3 to (a) derive the average time until fixation given that the allele is fixed,  $\bar{t}_1(p_0)$ , and confirm that it equals (15.44), (b) derive the average time until loss given that the allele is lost,  $\bar{t}_0(p_0)$ , and (c) derive the average time until loss or fixation, regardless of which occurs first,  $\bar{t}(p_0)$ . (d) Finally, take the Taylor Series with respect to  $p_0$  of your results to parts (a) – (c) assuming that the allele is initially rare, then set  $p_0 = 1/N$  to obtain the waiting times starting from a single mutation. In each case, specify the time scale. [Hint: You can check your answer to part (b) by thinking about the relationship between  $\bar{t}_1(p_0)$  and  $\bar{t}_0(1-p_0)$  in this neutral model.]
- Problem 15.6:** Derive a diffusion approximation to the birth-death model describing the number of individuals in a population that is at risk of extinction. Assume that the birth and death probabilities are given by  $b_i$  and  $d_i$ , where  $i$  is the population size. Also assume that the population is tracked until it reaches a minimum viable population size  $m$ . (a) Use the method outlined in Box 15.2 to derive the drift and diffusion parameters and to show that the third moment equals zero. [Hint: You can use the scale parameters  $\alpha = 1$  and  $\beta = 1$ , but you will need to assume that  $(b-d)$  is small in the same way that we assumed  $s$  was small in Box 15.3.] (b) From these

drift and diffusion parameters, show that  $A(x) = \int 2\mu(x)/\sigma^2(x) dx$  is a linear function of  $x = i/m$ . Use this fact and equation (15.35) to write down the probability that the population reaches the minimum viable population size before going extinct. (c) Plot your result from (b) and the exact result (14.29) for this birth-death model for your choice of parameters. How well does the diffusion approximation perform?

**Problem 15.7:** Derive a diffusion approximation for the diploid version of the Wright-Fisher model with selection in the absence of mutation, following the steps in Box 15.3. We assume that random mating produces  $N$  diploid individuals that experience viability and fertility selection such that the total frequency of the  $A$  allele among their gametes becomes

$$p' = \frac{(1+s)p^2 + p(1-p)(1+hs)}{(1+s)p^2 + 2p(1-p)(1+hs) + (1-p)^2}.$$

$2N$  gametes are then sampled at random to produce the next generation of diploids, following a binomial distribution with parameters  $2N$  and  $p'$ . (a) Show that the diffusion approximation is valid if we rescale selection using  $\psi = 2Ns$  and the scaling parameters  $\alpha = \beta = 1$ . Specifically, show that the drift and diffusion coefficients are finite. [EXTRA CHALLENGE: Show that the third moment is zero.] (b) Using this scaling, what are the drift and diffusion coefficients and how is time measured? (c) Use equation (15.35b) to write the probability of fixation for an allele  $A$  that is initially at frequency  $p_0$  (do not evaluate the integrals). (d) If allele  $A$  is beneficial, its fate is generally determined while it remains rare. In this case, show that we can approximate  $A(p) = \int 2\mu(p)/\sigma^2(p) dp$  by  $A(p) = \int 2h\psi dp$ . Having done so, carry out the integrals in (15.35b) and derive the probability of fixation. Compare your result to the fixation probability in the haploid model (15.39). [Note: In general, the integrals in (c) cannot be evaluated analytically. For alleles whose fates are not decided while rare, such as recessive beneficial alleles and deleterious alleles, the integrals in (c) must be evaluated numerically to determine the probability of fixation.]

**Problem 15.8:** Consider a birth-death model of population growth incorporating density dependence. Specifically, in a time step, a birth occurs with a density-dependent probability  $b i (1 - i/K)$ , and a death occurs with a density-independent probability  $d i$ , where  $i$  is the current population size and  $K$  is the maximum population size (the birth probability is zero at  $i = K$ ). (a) Calculate the first three moments describing the change in population size (15.2.1). (b) Assuming that a diffusion approximation is valid and using the first and second moments from (a) for the drift and diffusion coefficients, plot the average time to extinction for various values of  $d$  between 0.001 and 0.01, using  $i = 10$ ,  $K = 100$ , and  $b = 0.002$ . Because there is only one absorbing boundary (at  $i = 0$ ) in this model, use equation (15.43) and a mathematical software package to evaluate the integrals numerically. (c) Show that your results from (b) are similar to a numerical evaluation of the exact time to extinction using Recipe 14.3. (d) Show that the restrictions (15.2.6) required for the diffusion approximation to be valid can be met by choosing  $\alpha = \beta = 1/2$  (with  $m = K$ ), as long as the intrinsic birth rate minus the death rate is small, such that  $(b - d)K^{1/2} = \psi$ . (e) Repeat part (b) using the drift and diffusion coefficients from part (d). [Hint: The results should be the same if you rescale the variables appropriately.]

## Further Reading

For more information on the mathematical underpinnings of diffusion models, see

- Karlin, S., and H. M. Taylor. 1981. *A Second Course in Stochastic Processes*. Academic Press, New York.
- Allen, L. J. S. 2003. *An Introduction to Stochastic Processes with Applications to Biology*. Pearson Prentice Hall, Upper Saddle River, N.J.

For more information on diffusion models in evolution and ecology, see

- Crow, J. F., and M. Kimura. 1970. *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- Ewens, W. J. 1979. *Mathematical Population Genetics*. Springer-Verlag, Berlin.
- Lande, R., and B.-E. Sæther. 2003. *Stochastic Population Dynamics in Ecology and Conservation*. Oxford University Press, Oxford.
- Nisbet, R. M., and W.S.C. Gurney. 1982. *Modelling Fluctuating Populations*. Wiley, New York.
- Rice, S. 2004. *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sinauer Associates. Sunderland, Mass.

## References

- Allen, L. J. S. 2003. *An Introduction to Stochastic Processes with Applications to Biology*. Pearson Prentice Hall, Upper Saddle River, N.J.
- Britton, N. F. 1986. *Reaction-Diffusion Equations and Their Applications to Biology*. Academic Press, New York.
- Chuck, A. S., M. F. Clarke, and B. O. Palsson. 1996. Retroviral infection is limited by Brownian motion. *Hum. Gene Therapy* 7:1527–1534.
- Clegg, M. T. 1995. *Science and the Endangered Species Act*. National Academy Press, Washington D.C.
- Crow, J. F., and M. Kimura. 1970. *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- Dieckmann, U., and R. Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.* 34:579–612.
- Edelstein-Keshet, L. 1988. *Mathematical Models in Biology*. McGraw-Hill, New York.
- Ewens, W. J. 1979. *Mathematical Population Genetics*. Springer-Verlag, Berlin.
- Karlin, S., and H. M. Taylor. 1981. *A Second Course in Stochastic Processes*. Academic Press, New York.
- Kimura, M. 1957. Some problems of stochastic processes in genetics. *Ann. Math. Stat.* 28:882–901.
- Kimura, M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713–719.
- Kot, M. 2001. *Elements of Mathematical Ecology*. Cambridge University Press, Cambridge.
- Lande, R. 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* 142:911–927.
- Melhem, E. R. 2002. Time-course of apparent diffusion coefficient in neonatal brain injury: the first piece of the puzzle. *Neurology* 59:798–799.
- Okubo, A., and S. A. Levin. 2001. *Diffusion and Ecological Problems*. Springer-Verlag, Berlin.

- Pachepsky, E., F. Lutscher, R. M. Nisbet, and M. A. Lewis. 2004. Persistence, spread and the drift paradox. *Theor. Popul. Biol.* 67:61–73.
- Parker, G. 1970. Sperm competition and its evolutionary consequences in the insects. *Biol. Rev.* 45:525–556.
- Schluter, D., T. D. Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51:1699–1711.
- Speirs, D. C., and W.S.C. Gurney. 2001. Population persistence in rivers and estuaries. *Ecology* 85:1219–1237.
- Taylor, P., and T. Day. 1997. Evolutionary stability under the replicator and the gradient dynamics. *Evol. Ecol.* 11:579–590.