CHAPTER 8

Equilibria and Stability Analyses—Nonlinear Models with Multiple Variables

Chapter Goals:
- To construct dynamical models involving nonlinear combinations of multiple variables
- To find equilibria of these models
- To analyze the stability of these equilibria

Chapter Concepts:
- Multivariable equilibria
- Linearization near equilibria
- Stability analyses with multiple equilibria
- Linkage disequilibrium
- Approximate stability analyses

8.1 Introduction

Chapter 7 discussed methods for determining equilibria and their stability properties for models involving multiple variables, but it only considered cases where the dynamical equations were linear functions of the variables. Most interesting biological models are not linear, because any interaction among individuals requires a nonlinear model. For example, Phillips' (1996) model of within-human HIV dynamics involved interactions: virus particles contact and infect CD4+ cells (note the nonlinear $R\ V$ term in (2.4.1)). Similarly, the model of Blower et al. (2000) describing the spread of HIV involved interactions: unprotected sexual contact between infected and uninfected males (note the nonlinear $\lambda\ X$ terms in (2.5.1)). In this chapter,

we describe how the methods of Chapter 7 can be extended to nonlinear models.

In section 8.2 we present nonlinear models with multiple variables in continuous time. Section 8.3 then presents the discrete-time counterpart. Again, we develop the techniques in the context of two variable models for simplicity, and then present general recipes for handling models with an arbitrary number of variables. Finally, section 8.4 illustrates how perturbation techniques can be used in nonlinear models with multiple variables to obtain useful approximations when conducting stability analyses.

8.2 Nonlinear Multiple-Variable Models

We begin with an example to illustrate the process of finding equilibria in nonlinear models with multiple variables. Consider a two-variable model for the spread of a disease that tracks the dynamics of the number of susceptible and infected individuals in a population (denoted by S and I):

$$\begin{aligned} \frac{dS}{dt} &= \theta - dS - \beta SI + \gamma I, \\ \frac{dI}{dt} &= \beta SI - (d + \nu + \gamma) I. \end{aligned} \tag{8.1}$$

For simplicity, we have assumed that the host population is replenished by immigration at total rate, θ, and that recovered individuals immediately become susceptible again (i.e., there is no immunity). In equation (8.1), β denotes the transmission rate of the disease (i.e., $\beta = c\,a$ where c is the rate of contact between susceptible and infected hosts and a is the probability of transmission given that a contact occurs; Chapter 3), d denotes the per capita background mortality rate of the host, ν denotes the additional mortality that is caused by infection, and γ denotes the rate of clearance of disease through host defense mechanisms. Given these definitions, all parameters in the model are positive.

8.2.1 Finding Equilibria

To identify the equilibria $\hat{S}$, $\hat{I}$ of the model, we set dS/dt = 0 and dI/dt = 0. This gives two equilibrium conditions:

$$0 = \theta - d\hat{S} - \beta\hat{S}\,\hat{I} + \gamma\hat{I}, \tag{8.2a}$$

$$0 = \beta\hat{S}\,\hat{I} - (d + \nu + \gamma)\hat{I}. \tag{8.2b}$$

There can be several equilibria in a nonlinear model, unlike a linear model, and we would like to obtain explicit expressions for all of them.

In general, finding all equilibria can be a difficult task because the equations that these equilibria must satisfy might be complicated functions of the dynamic variables. One good place to start is to factor these equations. If we can do this, then we can look for values of the dynamic variables that make any of the factors zero (because the entire expression will then be zero). For the present model we can see that equation (8.2a) cannot be factored, but equation (8.2b) can, giving

$$0 = \hat{I}\{\beta\hat{S} - (d + \nu + \gamma)\}. \tag{8.2c}$$

This simplifies our task because it is much clearer from (8.2c) that dI/dt equals zero only if $\hat{I} = 0$ or $0 = \beta\hat{S} - (d + \nu + \gamma)$. To find an equilibrium of the model as a whole, however, we also require that the dynamic variable S is unchanging (i.e., that equation (8.2a) holds). Therefore, for each of the different ways in which the variable I can be unchanging, we must determine the conditions required for S to be unchanging as well. Specifically, when $\hat{I} = 0$, we must determine if there are conditions under which S is also constant. Similarly, when $0 = \beta\hat{S} - (d + \nu + \gamma)$, we must determine if there are conditions under which S is again constant.

We begin with the case where $\hat{I} = 0$. Substituting this into equation (8.2a), we see that the equation, $0 = \theta - d\hat{S}$ must hold for S to remain constant. This implies that $\hat{S} = \theta/d$. As a result, one equilibrium of this model is

$$\hat{S} = \frac{\theta}{d}, \quad \hat{I} = 0 \tag{8.3a}$$

Similarly, we can substitute $\hat{S} = (d + \nu + \gamma)/\beta$ into equation (8.2a) to obtain another equilibrium (try this)

$$\hat{S} = \frac{d + \nu + \gamma}{\beta}, \quad \hat{I} = \frac{\theta - \frac{d}{\beta}(d + \nu + \gamma)}{d + \nu} \tag{8.3b}$$

Equilibrium (8.3a) corresponds to the case where the disease is absent, and equilibrium (8.3b) corresponds to the case where the disease is present, which is often referred to as the *endemic* equilibrium.

Before proceeding, we should step back for a moment and consider how the above procedure works generally. Model (8.1) is based on a specific set of assumptions about how the population of susceptible hosts gets replenished (i.e., by immigration) as well as how susceptible hosts become infected. A general model involving two variables can be written as

$$\begin{aligned} \frac{dS}{dt} &= f(S, I), \\ \frac{dI}{dt} &= g(S, I), \end{aligned} \tag{8.4}$$

where $f(S,I)$ and $g(S,I)$ can be any functions of the variables S and I. To find the equilibria of this model, we again need to identify values of the variables $\hat{S}$ and $\hat{I}$ that result in no change in either variable. That is, we set $dS/dt = 0$ and $dI/dt = 0$ to get two equilibrium conditions, which we must solve for the two unknowns $\hat{S}$ and $\hat{I}$:

$$0 = f(\hat{S}, \hat{I}), \tag{8.5a}$$

$$0 = g(\hat{S}, \hat{I}). \tag{8.5b}$$

The equilibrium conditions (8.5) describe the null clines of the model (Figure 8.1). If (8.5a) holds, then S remains constant. If (8.5b) holds, then I remains constant. But unless both equilibrium conditions hold, changes to one variable will typically lead to changes in the other. Therefore, for both variables to remain constant, an equilibrium must simultaneously satisfy both equilibrium conditions. Graphically, this means that any equilibrium

must lie on null clines for every variable in a model (see filled circles in Figure 8.1).

At a *multivariable equilibrium*, all variables must remain unchanged. There can be multiple equilibria in nonlinear models.

Depending on the functions f and g, identifying the possible equilibria of a model can be straightforward, difficult, or even impossible. A good strategy is to factor each of the equilibrium conditions, identify which one is easiest to solve for its null clines, and then plug in these null clines, one by one, into the other equilibrium condition to see if it can be solved as well.

These procedures can be generalized for models involving any number of variables:

Definition 8.1: General Nonlinear Models in Continuous Time

A general, nonlinear, continuous-time model with n dynamic variables $x_1, \ldots, x_n$ can be written as

$$\begin{aligned}
\frac{dx_1}{dt} &= f_1(x_1, x_2, \ldots, x_n), \\
\frac{dx_2}{dt} &= f_2(x_1, x_2, \ldots, x_n), \\
&\vdots \\
\frac{dx_n}{dt} &= f_n(x_1, x_2, \ldots, x_n),
\end{aligned}$$

where $f_1, f_2, \ldots, f_n$ denote different functions specifying the rate of change of each variable.

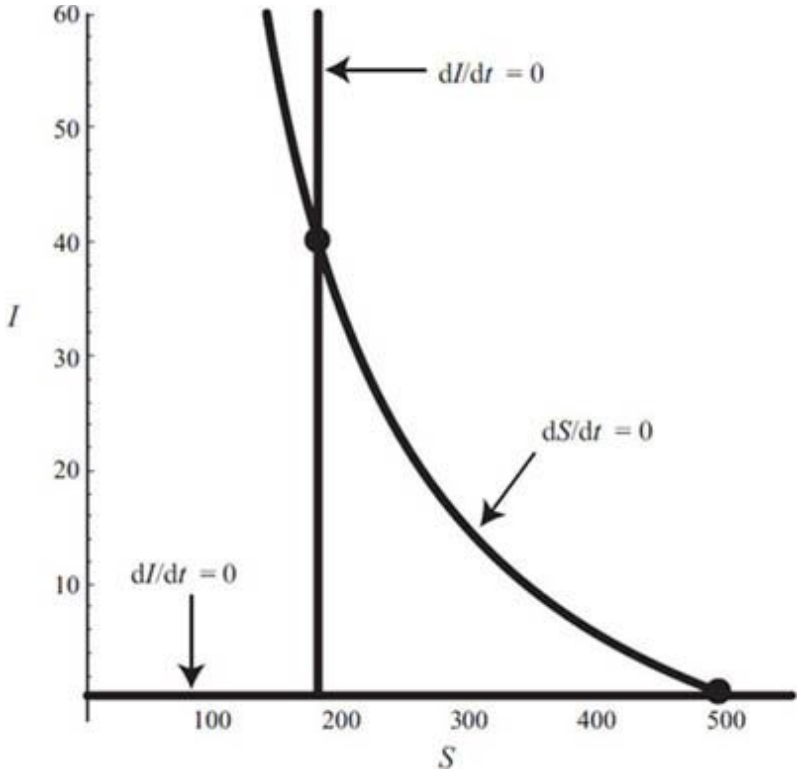

Figure 8.1: Null clines for the disease transmission model. For model (8.1), the equilibrium condition (8.2a) defines the null cline along which the number of susceptible individuals does not change (dS/dt = 0). The equilibrium condition (8.2b) defines two null clines along which the number of infected individuals does not change (dI/dt = 0). An equilibrium must lie on the null clines for both variables, which occurs at $\hat{S} = 180$, $\hat{I} = 40$ and at $\hat{S} = 500$, $\hat{I} = 0$ (filled circles). Parameter values are $v = 0.7$, $d = 0.1$, $\gamma = 0.1$, $\beta = 0.005$, $\theta = 50$.

Recipe 8.1 then describes how to find equilibria of models of the form in Definition 8.1:

Recipe 8.1: Equilibria of Nonlinear Multivariable Models in Continuous Time

Equilibria are found by determining the values of the variables that cause all of the variables to remain constant: $dx_1/dt = 0$, $dx_2/dt = 0$, . . . , $dx_n/dt = 0$. This results in n equations in n unknowns: $f_1(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = 0$, $f_2(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = 0$, . . . , $f_n(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = 0$. Any point that satisfies *all* of these conditions simultaneously is an equilibrium. To identify the equilibria:

Step 1: Factor each equation, $f_i(\hat{\chi}_1, \hat{\chi}_2, \ldots, \hat{\chi}_n) = 0$.

Step 2: Identify all possible solutions to one of these equations (start with the simplest one).

Step 3: Plug each possible solution into the remaining equations, and repeat the above steps until equilibrium values for all variables are identified.

For nonlinear models, there can be more than one equilibrium. Depending on the complexity of the model, it may or may not be possible to identify all of the equilibria explicitly.

8.2.2 Determining the Stability of the Equilibria

Once the equilibria have been identified, our next task is to determine the behavior of the model if we move the system slightly away from one of these points. In other words, we want to know when each equilibrium is *locally stable* (Definition 5.2). Stability conditions can be used to address very fundamental biological questions. For example, in our example model, can the disease spread and become established within a population? We can answer this question by asking if the equilibrium without the disease, $\hat{S} = \theta/d$, $\hat{I} = 0$, is unstable following the introduction of a small number of infected individuals.

To address this question, we imagine starting the model very close to an equilibrium. We define ε_S and ε_I to be the *displacements* of the starting values of the variables from the equilibrium $\hat{S}$, $\hat{I}$. That is, $\varepsilon_S = S - \hat{S}$ and $\varepsilon_I = I - \hat{I}$ where S and I are the numbers of susceptible and infected individuals, respectively. The terms ε_S and ε_I are dynamic variables that change over time as S and I change. These displacements will get larger or smaller depending on whether the system moves away from the equilibrium (an unstable equilibrium) or toward the equilibrium (a stable equilibrium). Thus, to determine whether the displacements grow or decay we need equations describing how the displacements change through time.

Differential equations for the displacements can be obtained by differentiating ε_S and ε_I with respect to time:

$$\begin{aligned}\frac{d\varepsilon_S}{dt} &= \frac{d}{dt}(S - \hat{S}) \\ &= \frac{dS}{dt} \\ &= \theta - dS - \beta S I + \gamma I\end{aligned} \tag{8.6a}$$

and

$$\begin{aligned}\frac{d\varepsilon_I}{dt} &= \frac{d}{dt}(I - \hat{I}) \\ &= \frac{dI}{dt} \\ &= \beta SI - (d + \nu + \gamma) I.\end{aligned} \tag{8.6b}$$

These results follow from the fact that $\hat{S}$ and $\hat{I}$ are constants and hence their derivatives with respect to time are zero. Consequently, the dynamics of the displacements are governed by the same equations that govern the dynamics of the variables S and I. We have not quite finished our derivation, however, because the right-hand side of each equation is written in terms of the original variables S and I. To have a self-contained model for the displacements ε_S and ε_I, we need to rewrite equations (8.6) in terms of ε_S and ε_I alone. This can be accomplished by using the fact that $S = \hat{S} + \varepsilon_S$ and $I = \hat{I} + \varepsilon_I$, giving

$$\begin{aligned}\frac{d\varepsilon_S}{dt} &= \theta - d(\hat{S} + \varepsilon_S) - \beta(\hat{S} + \varepsilon_S)(\hat{I} + \varepsilon_I) + \gamma(\hat{I} + \varepsilon_I) \\ &= \theta - d\hat{S} - d\varepsilon_S - \beta\hat{S}\hat{I} - \beta\hat{I}\varepsilon_S - \beta\hat{S}\varepsilon_I - \beta\varepsilon_S\varepsilon_I + \gamma\hat{I} + \gamma\varepsilon_I,\end{aligned} \tag{8.7a}$$

$$\frac{d\varepsilon_I}{dt} = \beta(\hat{S} + \varepsilon_S)(\hat{I} + \varepsilon_I) - (d + \nu + \gamma)(\hat{I} + \varepsilon_I)$$

$$= \beta\hat{S}\,\hat{I} + \beta\hat{I}\varepsilon_S + \beta\hat{S}\varepsilon_I + \beta\varepsilon_S\varepsilon_I - d\hat{I} - \nu\hat{I} \qquad (8.7b)$$

$$- \gamma\hat{I} - d\,\varepsilon_I - \nu\varepsilon_I - \gamma\varepsilon_I.$$

We are almost done, but we can now simplify (8.7) by using the fact that the displacements ε_S and ε_I are small. As a consequence, higher powers in these terms (e.g., ε_S^2, ε_I^2, and $\varepsilon_S\,\varepsilon_I$) will be extremely small and thus we can ignore them. Doing so, and grouping terms together that involve the displacements ε_S and ε_I, we get

$$\frac{d\varepsilon_S}{dt} = (\theta - d\hat{S} - \beta\hat{S}\hat{I} + \gamma\hat{I}) + (-d\varepsilon_S - \beta\hat{S}\varepsilon_I - \beta\hat{I}\varepsilon_S + \gamma\varepsilon_I), \qquad (8.8a)$$

$$\frac{d\varepsilon_I}{dt} = (\beta\hat{S}\hat{I} - d\hat{I} - \nu\hat{I} - \gamma\hat{I}) + (-d\varepsilon_I - \nu\varepsilon_I - \gamma\varepsilon_I + \beta\hat{S}\varepsilon_I + \beta\hat{I}\varepsilon_S). \qquad (8.8b)$$

The beauty of writing the equations in this way is that the first parenthetical term in each equation is identical to one of the equilibrium conditions, and therefore it must be zero. The equilibrium condition (8.2a) tells us that ($\theta - d\,\hat{S} - \beta\,\hat{S}\,\hat{I} + \gamma\,\hat{I}$) is zero, while the equilibrium condition (8.2b) tells us that ($\beta\,\hat{S}\,\hat{I} - d\,\hat{I} - \nu\,\hat{I} - \gamma\,\hat{I}$) is zero. As a result, the dynamics of the small displacements ε_S and ε_I near any equilibrium point of the original model are governed by the equations

$$\frac{d\varepsilon_S}{dt} = -d\varepsilon_S - \beta\hat{S}\varepsilon_I - \beta\hat{I}\varepsilon_S + \gamma\varepsilon_1, \qquad (8.9a)$$

$$\frac{d\varepsilon_I}{dt} = -d\,\varepsilon_I - \nu\varepsilon_I - \gamma\varepsilon_I + \beta\hat{S}\varepsilon_I + \beta\hat{I}\varepsilon_S. \qquad (8.9b)$$

The most important point to take away from the above calculations is that we started with a *nonlinear* model and we have arrived at a *linear* system of equations that governs the dynamics of the displacements from an equilibrium. We have been able to do this because we allow only small

displacements, and therefore we can approximate the nonlinear model with a linear one near the equilibrium by ignoring higher-powered terms in the ε's. This process is called a *local stability analysis* or *linearization* near an equilibrium. Given that we have a linear system of equations, we can then use the techniques from Chapter 7 to determine whether the equilibrium is stable or unstable. If the displacements get smaller with time then the equilibrium is locally stable.

> The stability of an equilibrium is determined by a local stability analysis, which is a *linearization* of the nonlinear model near the equilibrium of interest.

As in Chapter 7, we can write equations (8.9) in matrix form:

$$\begin{pmatrix} \frac{d\varepsilon_S}{dt} \\ \frac{d\varepsilon_I}{dt} \end{pmatrix} = \begin{pmatrix} -\beta\hat{I} - d & -\beta\hat{S} + \gamma \\ \beta\hat{I} & \beta\hat{S} - (d + \nu + \gamma) \end{pmatrix} \begin{pmatrix} \varepsilon_S \\ \varepsilon_I \end{pmatrix}. \tag{8.10}$$

Matrix equation (8.10) describes the dynamics of small displacements from any equilibrium of the original model. According to Rule 7.2 for continuous-time linear models, the equilibrium is stable provided that all eigenvalues have negative real parts. This ensures that the system moves toward the equilibrium along all eigenvectors.

To use the above results for determining local stability properties, we must first specify which equilibrium we are near. Let us begin with the equilibrium where the disease is absent. The stability of this equilibrium is determined by the eigenvalues of the matrix in (8.10). For the disease-absent equilibrium, $\hat{S} = \theta/d$, $\hat{I} = 0$, and this matrix therefore simplifies to

$$\begin{pmatrix} -d & -\beta\frac{\theta}{d} + \gamma \\ 0 & \beta\frac{\theta}{d} - (d + \nu + \gamma) \end{pmatrix}. \tag{8.11}$$

Matrix (8.11) is upper triangular, and therefore its eigenvalues can be read directly from the diagonal (Rule P2.26):

$$r_1 = \beta \frac{\theta}{d} - (d + \nu + \gamma) \quad \text{and} \quad r_2 = -d. \tag{8.12}$$

Because d is a positive constant, the stability of the disease-absent equilibrium is completely determined by the sign of the first eigenvalue $\beta (\theta/d) - (d + \nu + \gamma)$. If this is negative, then the equilibrium will be stable and the disease will not spread. If it is positive, then the disease will spread into the population.

The requirement for a disease to spread, $\beta (\theta/d) - (d + \nu + \gamma) > 0$, can be rewritten in terms of the number of susceptible individuals $\hat{S} = \theta/d$. Making this substitution and rearranging, we get the condition $\beta\hat{S}/(d + \nu + \gamma)>1$ for a disease to spread. The quantity on the left-hand side is sometimes called R_0, the reproductive number, and represents the expected number of new infections produced per infected host when a disease is introduced into a susceptible population. For R_0 to be greater than one, the population size in the absence of the disease, $\hat{S}$, must be large enough that an infected individual encounters enough susceptible individuals to ensure infection of at least one other member of the population before the infected individual dies or recovers.

We have just completed a local stability analysis of the $\hat{S} = \theta/d$, $\hat{I} = 0$ equilibrium, but this model has a second equilibrium: the endemic equilibrium $\hat{S} = (d + \nu + \gamma)/\beta$, $\hat{I} = (\theta - (d/\beta)(d + \nu + \gamma))/(d + \nu)$. Each equilibrium of a model has its own stability properties, and we must identify these properties by performing a local stability analysis for each equilibrium separately.

When there are *multiple equilibria*, the *stability* of each must be evaluated separately.

Before we begin, we should determine the conditions under which the second equilibrium is biologically feasible (i.e., biologically valid) by asking when $\hat{S}$ and $\hat{I}$ are both positive. Given that the parameters are assumed to be positive, $\hat{S}$ will always be positive, but $\hat{I}$ need not be. For $\hat{I}$ to be positive, we require that $\theta > (d/\beta)(d + \nu + \gamma)$, which can be reorganized as $\beta (\theta/d) - (d + \nu + \gamma) > 0$. This is exactly the same condition required for the disease-

absent equilibrium to be unstable. Therefore, the endemic equilibrium is biologically feasible only when the disease can spread when rare.

Once again, we use the linear version of the model (8.10) that describes the dynamics near an equilibrium for our stability analysis, but we now plug in the second equilibrium. After factoring, equation (8.10) becomes

$$\begin{pmatrix} \frac{d\varepsilon_S}{dt} \\ \frac{d\varepsilon_I}{dt} \end{pmatrix} = \begin{pmatrix} \frac{-\beta\theta + \gamma d}{d + \nu} & -\nu - d \\ \frac{\beta\theta - d(d + \nu + \gamma)}{d + \nu} & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_S \\ \varepsilon_I \end{pmatrix}. \tag{8.13}$$

Again, according to Rule 7.2, the equilibrium of (8.13) will be stable only if all of the eigenvalues of the matrix have negative real parts. We could calculate the eigenvalues of (8.13) directly and then examine their real parts, but it is easier to use the trace and determinant conditions in Rule P2.25 of Primer 2. According to Rule P2.25, for the real part of both eigenvalues to be negative, the determinant of a 2 × 2 matrix must be positive and its trace must be negative.

The determinant of the matrix in (8.13) is $\beta\,\theta - d\,(d + \nu + \gamma)$. The sign of this expression might not be obvious at first, but recall that we require that $\beta\,(\theta/d) - (d + \nu + \gamma) > 0$ for $\hat{I}$ to be positive. This implies that the determinant is positive whenever the endemic equilibrium is biologically feasible. The trace of (8.13) is $(-\beta\,\theta + \gamma\,d)/(d + \nu)$, and because all the parameters are positive, the sign of the trace depends on the sign of $(-\beta\,\theta + \gamma\,d)$. The fact that $\hat{I}$ must be positive implies that $\beta\,\theta$ must be greater than $d^2 + d\,\nu + d\,\gamma$ (from our condition for biological feasibility), which in turn implies that $\beta\,\theta$ is greater than $d\,\gamma$. Therefore, $(-\beta\,\theta + \gamma\,d)$ must be negative, meaning that the trace must be negative. Thus, according to Rule P2.25, the real parts of both eigenvalues are negative, and we conclude that the endemic equilibrium is locally stable whenever it is biologically feasible.

The above example illustrates how we can use the techniques of Chapter 7 for linear models to conduct a local stability analysis of our nonlinear epidemiological model. The approach rests on the assumption that our linear approximation adequately captures the dynamics of the nonlinear model near equilibria. Figure 8.2 shows that, indeed, this linearization provides a remarkably good approximation to the nonlinear model (8.1) near both equilibria. In fact, mathematicians have proven that such approximations

will typically work for local stability analyses of general nonlinear models as well. These general results are presented next.

8.2.3 The General Approach for Determining the Local Stability of Equilibria

In the above derivation, we approximated the dynamics near an equilibrium with linear equations by first writing out nonlinear equations (8.7) and then dropping higher-order terms in the displacement, like ε_S^2, ε_I^2, and $\varepsilon_S\varepsilon_I$. In other models, however, it might not be so obvious which terms should be kept and which can be dropped. For example, what should we do with a term like $(1 + \varepsilon_I)/(1 - \varepsilon_S)$? Fortunately, there is a much simpler route to reaching matrix (8.10b), both conceptually and computationally, which automatically drops the correct terms.

We describe the procedure using the more general two-variable model (8.4). Having found the equilibria of a model, we can again define the displacements ε_S and ε_I from one of the equilibria and derive equations for their dynamics. The same procedure used in the specific model can be followed to obtain the more general version of (8.7a) and (8.7b):

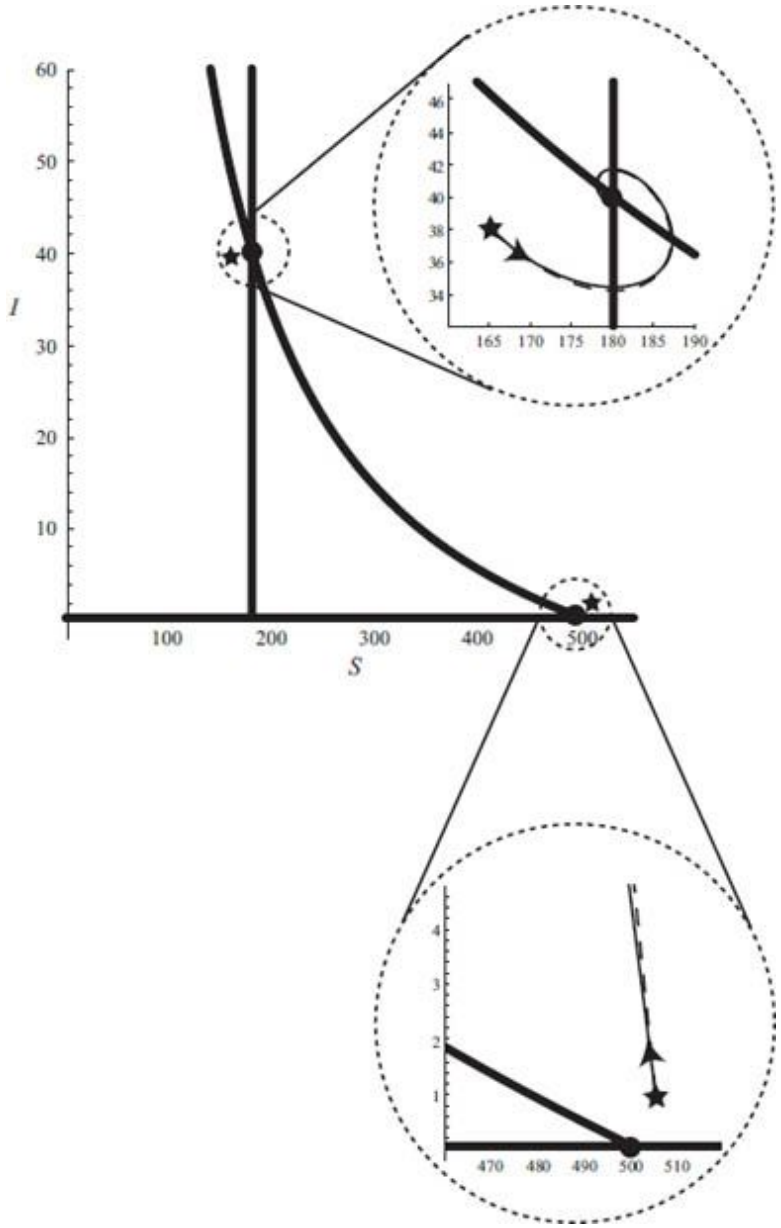

Figure 8.2: Dynamics near the equilibria of the disease transmission model. The inset figure below the graph shows a numerical solution to the differential equations using *Mathematica* (see Chapter 4) starting from an initial number of susceptible and infected individuals (stars) near the disease-absent equilibrium with $S(0) = 505$, $I(0) = 1$; the solid thin curve uses the exact differential equations (8.1), while the dashed curve uses the linear differential equations (8.9) that are approximately correct near the equilibrium. The inset figure to the right of the graph is similarly drawn, but starting near the endemic equilibrium with $S(0) = 165$, $I(0) = 38$. In the region of each equilibrium, the non-linear dynamics are well approximated by linear differential equations. Parameters as in Figure 8.1.

$$\frac{d\varepsilon_S}{dt} = f(\hat{S} + \varepsilon_S, \hat{I} + \varepsilon_I), \tag{8.14a}$$

$$\frac{d\varepsilon_I}{dt} = g(\hat{S} + \varepsilon_S, \hat{I} + \varepsilon_I). \tag{8.14b}$$

The next step is to come up with a linear approximation to these equations.

In the model of section 8.2.2, we first multiplied out the equations and then discarded terms involving higher powers of the ε's. We can accomplish the same task more generally by linearizing the functions f and g using the Taylor series for each function (Primer 1) near the equilibrium point $(\hat{S}, \hat{I})$. This is exactly what we did for one-variable nonlinear models in Chapter 5. Box 8.1 generalizes the Taylor series for functions of more than one variable.

Box 8.1: The Taylor Series Approximation with Multiple Variables

The Taylor series was introduced in Primer 1 as a method to rewrite functions as a series of power terms, $f(x) = b_0 + b_1x + b_2x^2 \cdots$. Although the functions we considered in Primer 1 involved only one variable, x, the Taylor series can also be used for functions involving multiple variables. The Taylor series of a function of d variables, $f(x_1, x_2, \ldots, x_d)$, around the point $x_i = a_i$ is given by

$$f(x_1, x_2, \ldots, x_d) = f(a_1, a_2, \ldots, a_d) + \sum_{j=1}^{d} \left(\frac{\partial f}{\partial x_j}\bigg|_{x_1=a_1, x_2=a_2, \ldots, x_d=a_d} \right) (x_j - a_j) + \text{residual}. \tag{8.1.1}$$

Before going any further, we must take a short mathematical excursion to discuss exactly what we mean by $\partial f/\partial x_j$.

Mathematical aside: Partial derivatives

The term $\partial f/\partial x_j$ is known as a "partial derivative" and is written with curly ∂'s rather than the d's normally used in more familiar derivatives (see Appendix 2). A partial derivative is a straightforward extension of the normal derivative to functions involving multiple

variables. Specifically, $\partial f/\partial x_j$ involves taking the normal derivative of the function f with respect to the variable x_j treating all of the other variables as constants.

For example, consider the function, $f(x,y) = x + e^{xy}$. If we treat y as if it were a constant, we get the partial derivative with respect to x: $\partial f/\partial x = 1 + ye^{xy}$. Similarly, if we treat x as a constant, we get the partial derivative with respect to y: $\partial f/\partial y = xe^{xy}$. As another example, consider the function, $f(x,y) = x^2 + 3xy + y^2$. If we treat y as a constant, we get the partial derivative with respect to x: $\partial f/\partial x = 2x + 3y$. If we treat x as a constant, we get the partial derivative with respect to y: $\partial f/\partial y = 3x + 2y$.

In the Taylor series (8.1.1), each $\partial f/\partial x_j$ term involves taking the partial derivative of $f(x_1, x_2, \ldots, x_d)$ with respect to only one variable, x_j. This partial derivative is then evaluated at the point $x_1 = a_1$, $x_2 = a_2$, . . . , $x_d = a_d$. Thus, once again, the derivative terms are no longer functions of the variables x_j.

The "residual" term in (8.1.1) contains second-order terms (double derivatives $\partial f^2/\partial x_j \partial x_k$ evaluated at $x_i = a_i$ multiplied by $(x_j - a_j)(x_k - a_k)/2$ for all variable pairs j and k), plus third-order terms, etc. As long as the higher-order derivatives are not too large, the residual term may be dropped, giving a linear approximation to the multivariable function for points near enough to $(a_1, a_2, \ldots, a_d)$.

For example, let us approximate the function $f(x,y) = e^x \sin(y)$ as a linear function of the variables x and y around the point $x = y = 0$. The constant term is $f(0,0) = e^0 \sin(0) = 0$. The linear term consists of two parts,

$$\left(\left.\frac{\partial f}{\partial x}\right|_{x=0,\,y=0}\right)(x - 0) \text{ and } \left(\left.\frac{\partial f}{\partial y}\right|_{x=0,\,y=0}\right)(y - 0).$$

The two partial derivatives that we need are $\partial f/\partial x = e^x \sin(y)$ and $\partial f/\partial y = e^x \cos(y)$, which become 0 and 1, respectively, when evaluated at $x = y = 0$. Plugging these terms into (8.1.1) gives the linear approximation $e^x \sin(y) \approx y$. This approximation is certainly not obvious, but we can

see how well it works when we plot $f(x,y)$ as a function of both x and y (Figure 8.1.1).

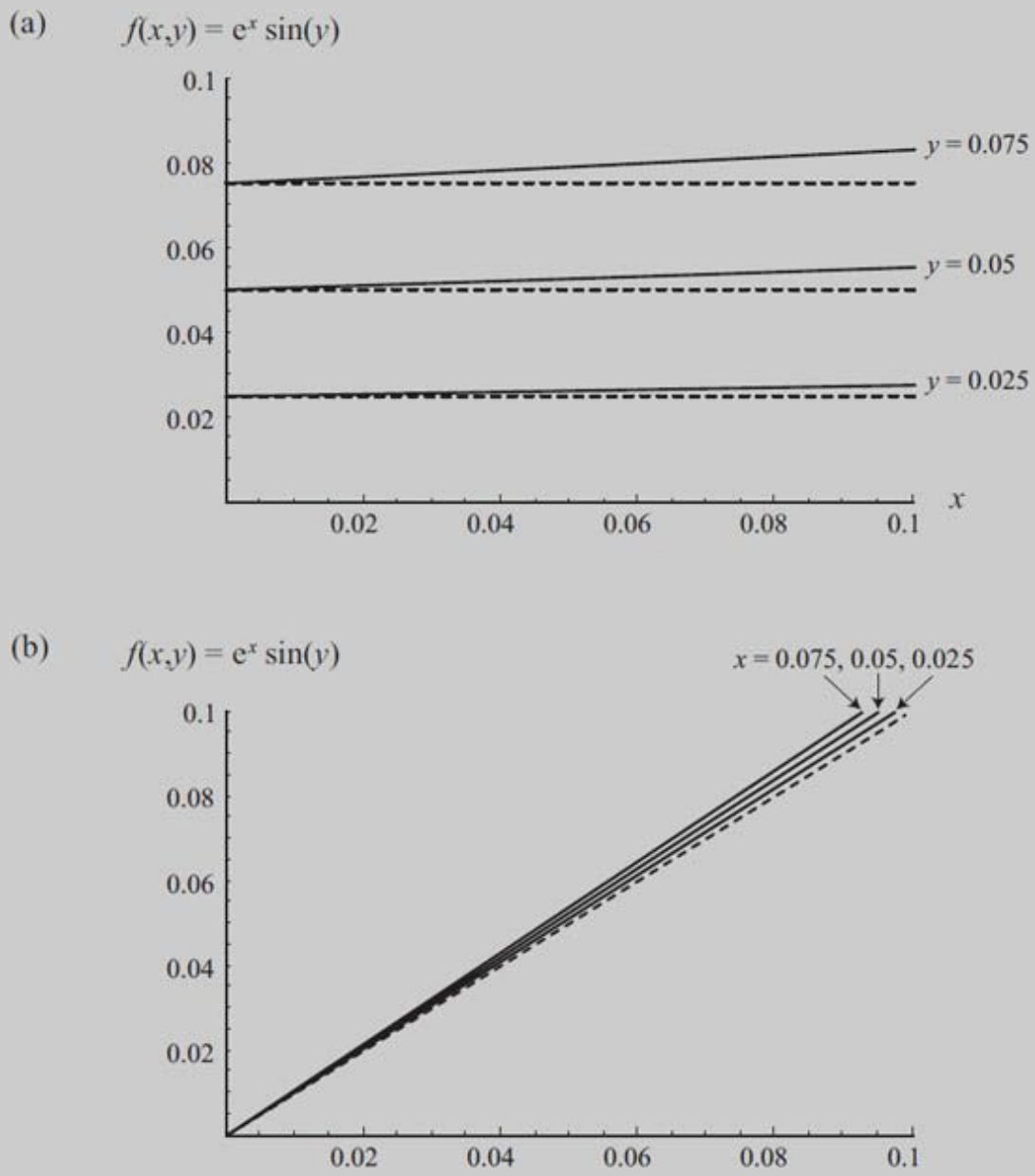

Figure 8.1.1: The Taylor series of a function of two variables. The function, $f(x,y) = e^x \sin(y)$, plotted as a function of x and y for various values of the other variable. For small values of x and y, $f(x,y)$ is (a) fairly insensitive to x but (b) almost proportional to y, as predicted by the linear Taylor series approximation, $f(x,y) \approx y$ (illustrated for comparison by dashed lines).

For more practice, use the Taylor series to derive the following linear approximations for the given functions when x and y are both

near zero:

$$f(x,y) = e^x e^y \approx 1 + x + y, \tag{8.1.2}$$

$$\frac{y}{1+x} \approx y, \tag{8.1.3}$$

$$\frac{e^y}{1+e^x} \approx \frac{1}{2} + \frac{y}{2} - \frac{x}{4}. \tag{8.1.4}$$

The Taylor series of the function f near the equilibrium $(\hat{S}, \hat{I})$ is

$$\frac{d\varepsilon_S}{dt} = f(\hat{S}, \hat{I}) + \left.\frac{\partial f}{\partial s}\right|_{S=\hat{S}, I=\hat{I}} (S - \hat{S}) + \left.\frac{\partial f}{\partial I}\right|_{S=\hat{S}, I=\hat{I}} (I - \hat{I}) \tag{8.15a}$$

$$+ \text{ higher-power terms}$$

$$= \left.\frac{\partial f}{\partial s}\right|_{S=\hat{S}, I=\hat{I}} \varepsilon_s + \left.\frac{\partial f}{\partial I}\right|_{S=\hat{S}, I=\hat{I}} \varepsilon + \text{higher-power terms}, \tag{8.15b}$$

where $(\partial f/\partial S)\Big|_{S=\hat{S}, I=\hat{I}}$ denotes the partial derivative of f with respect to the variable S, evaluated at the equilibrium point $S = \hat{S}$, $I = \hat{I}$ (see Box8.1). Expression (8.15b) follows from (8.15a) using the definitions of the deviations ε_S and ε_I, and from the fact that $f(\hat{S}, \hat{I}) = 0$ at equilibrium (from equation (8.5a)). Importantly, the derivatives $\partial f/\partial S$ and $\partial f/\partial I$ are evaluated at the equilibrium point $(\hat{S}, \hat{I})$ meaning that they are constants and not functions of S and I. Finally, the "higher-power terms" in expressions (8.15) are terms that involve higher-powers of ε_S and ε_I. Ignoring these, we arrive at the linear approximation

$$\frac{d\varepsilon_S}{dt} = \left.\frac{\partial f}{\partial s}\right|_{S=\hat{S}, I=\hat{I}} \varepsilon_s + \left.\frac{\partial g}{\partial I}\right|_{S=\hat{S}, I=\hat{I}} \varepsilon_I. \tag{8.16a}$$

The same calculations can be done for the function g, giving

$$\frac{d\varepsilon_I}{dt} = \left.\frac{\partial g}{\partial s}\right|_{S=\hat{S},\, I=\hat{I}} \varepsilon_s + \left.\frac{\partial g}{\partial I}\right|_{S=\hat{S},\, I=\hat{I}} \varepsilon_I. \tag{8.16b}$$

As a result, we can write the system of differential equations in matrix form as

$$\begin{pmatrix} \frac{\mathrm{d}\varepsilon_s}{\mathrm{d}t} \\ \frac{\mathrm{d}\varepsilon_I}{\mathrm{d}t} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial S} & \frac{\partial f}{\partial I} \\ \frac{\partial g}{\partial S} & \frac{\partial g}{\partial I} \end{pmatrix}_{S=\hat{S},\, I=\hat{I}} \begin{pmatrix} \varepsilon_s \\ \varepsilon_I \end{pmatrix}. \tag{8.17}$$

Fortunately, to perform a local stability analysis or linearization, we need not bother with the above derivation every time. Instead, we can just determine the matrix of derivatives in (8.17), which is known as the *Jacobian matrix*. For example, the matrix in (8.17) can be calculated for our epidemiological model using the functions $f(S,I) = \theta - d\,S - \beta\,S\,I + \gamma\,I$ and $g(S,I) = \beta\,S\,I - (d + \nu + \gamma)\,I$. Doing so reduces the Jacobian matrix in equation (8.17) to that of (8.10).

Definition 8.2: The Jacobian Matrix

Given n functions $f_1(x_1, x_2, \ldots, x_n), f_2(x_1, x_2, \ldots, x_n), \ldots, f_n(x_1, x_2, \ldots, x_n)$ describing the dynamics of n variables $x_1, \ldots, x_n$, the *Jacobian matrix* **J** is defined as

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}.$$

CAUTION: The order of the functions determines the order in which the derivatives must be taken. If the ith row uses the function describing the change over time in x_i, then the ith column must contain derivatives with respect to the variable x_i.

The Jacobian matrix, evaluated at an equilibrium, provides a linear approximation of the non-linear model near that equilibrium. In the vicinity of an equilibrium, the trajectories predicted using the full general model (8.4) will be very nearly the same as those of the linear approximation (Figure 8.2). The accuracy of the linearization allows us to use the techniques of Chapter 7 to determine stability, by using the linear model as a proxy for the more complicated nonlinear model. This approximation breaks down as the system moves farther away from the equilibrium, but it provides an excellent way to determine how the model behaves near any equilibrium point.

A local stability analysis (Recipe 8.2) must be performed separately on each equilibrium of interest.

Recipe 8.2

Local Stability of Equilibria for Nonlinear Multivariable Models in Continuous Time

To determine whether an equilibrium of interest is locally stable:

Step 1: Evaluate the Jacobian matrix (Definition 8.2) at the equilibrium of interest, $\mathbf{J}|_{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n}$. This matrix is often called the *local stability matrix*.

Step 2: Solve the characteristic polynomial $\text{Det}\,(\mathbf{J}|_{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n} - r\ \mathbf{I}) = 0$, which is an nth-degree polynomial.

Step 3: The n solutions ("roots") to this characteristic polynomial are the n eigenvalues $r_1, r_2, \ldots, r_n$.

Step 4: The real parts of all n eigenvalues must be negative for the equilibrium to be locally stable. Equivalently, the equilibrium is locally stable if the real part of the eigenvalue with the largest real part (the *leading eigenvalue*) is negative.

(If the real part of the leading eigenvalue is zero, the local stability analysis is inconclusive, and higher order terms must be considered.)

Although the real part completely determines stability, if the eigenvalues have complex parts, then the system will spiral around the equilibrium along some axes (for details, see Box 9.3).

Example: Predator-Prey Dynamics

Let us now get some practice by analyzing the predator-prey model introduced in equations (3.17) of Chapter 3. In these equations, the prey species is replenished by immigration rather than by reproduction, and the predator species has its birth rate solely determined by its rate of prey intake:

$$\begin{aligned} \frac{dn_1}{dt} &= \theta - acn_1n_2, \\ \frac{dn_2}{dt} &= \varepsilon acn_1n_2 - \delta n_2. \end{aligned} \tag{8.18}$$

In this model, θ is the prey immigration rate, c is the rate of contact between predator and prey, a is the probability that a predator consumes a prey given a contact, ε is the conversion efficiency of predators into prey, and δ is the per capita death rate of the predator. We used this model to explore how the population dynamics of a predator affects the population dynamics of a prey species and vice versa. Now we shall conduct a local stability analysis of its equilibria.

Following Recipe 8.1, the equilibria of this model are obtained by solving the equations that result from setting $dn_1/dt = 0$ and $dn_2/dt = 0$. This gives $0 = \theta - a\,c\,\hat{n}_1\,\hat{n}_2$ and $0 = \varepsilon\,a\,c\,\hat{n}_1\hat{n}_2 - \delta\,\hat{n}_2$. The second equation can be factored into $0 = \hat{n}_2(\varepsilon\,a\,c\,\hat{n}_1 - \delta)$, and if either factor is zero, then the population size of the predator remains constant. This occurs if either $\hat{n}_2 = 0$ or $\hat{n}_1 = \delta/(\varepsilon\,a\,c)$. We first consider the case where $\hat{n}_2 = 0$. Substituting this into the other equilibrium condition gives $0 = \theta$. This equilibrium condition can never be met because we have assumed that the immigration rate θ is

positive. Therefore, there is no possible equilibrium when $\hat{n}_2 = 0$. Biologically, if predators are absent (i.e., $\hat{n}_2 = 0$) then the prey population will grow indefinitely through immigration. Now consider the second possibility, in which $\hat{n}_1 = \delta/(\varepsilon\ a\ c)$. Substituting this into the equilibrium equation for $dn_1/dt = 0$ gives $0 = \theta - a\ c\ \delta/(\varepsilon\ a\ c)\ \hat{n}_2$, or $\hat{n}_2 = \theta\ \varepsilon/\delta$. Therefore, the only equilibrium is

$$\hat{n}_1 = \frac{\delta}{\varepsilon a c}, \quad \hat{n}_2 = \frac{\theta\varepsilon}{\delta}. \tag{8.19}$$

To determine the local stability of this equilibrium, we first calculate the Jacobian matrix (Definition 8.2)

$$\mathbf{J} = \begin{pmatrix} -ac\hat{n}_2 & -ac\hat{n}_1 \\ \varepsilon ac\hat{n}_2 & \varepsilon ac\hat{n}_1 - \delta \end{pmatrix}. \tag{8.20}$$

Evaluating (8.20) at the equilibrium (8.19) then gives

$$\mathbf{J} = \begin{pmatrix} -ac\varepsilon\theta/\delta & -\delta/\varepsilon \\ \varepsilon^2 ac\theta/\delta & 0 \end{pmatrix} \tag{8.21}$$

To determine if the equilibrium is stable (i.e., if the real parts of both eigenvalues are negative), we can use the determinant and trace conditions for a 2 × 2 matrix (Rule P2.25). The determinant of (8.21) is $\varepsilon\ a\ c\ \theta$, which is always positive because all parameters are positive. Furthermore, the trace of (8.21) is $-a\ c\ \varepsilon\ \theta/\delta$, which is always negative. Therefore, the real parts of both eigenvalues are negative according to Rule P2.25, and the equilibrium (8.19) is always locally stable.

Although the equilibrium is locally stable, we do not yet know whether the system approaches it smoothly or whether it displays cycles. To answer this question, we can calculate the eigenvalues of (8.21) explicitly. Using equations (P2.7) of Primer 2, we have

$$r = \frac{-\varepsilon ac\theta \pm \sqrt{\varepsilon ac\theta(\varepsilon ac\theta - 4\delta^2)}}{2\delta} \tag{8.22}$$

The eigenvalues will be real if $\varepsilon\ a\ c\ \theta - 4\ \delta^2 > 0$, which occurs if the rate of conversion of prey to predator offspring, multiplied by the prey immigration rate, is high relative to the predator's mortality rate. In this case, the equilibrium is reached smoothly without oscillations (Figure 8.3a). On the other hand, the eigenvalues will be complex if $\varepsilon\ a\ c\ \theta - 4\delta^2 < 0$, in which case, the equilibrium population size is reached in an oscillatory fashion (Figure 8.3b).

Finally, we can go one step further, by first writing the complex eigenvalues as $r = \alpha \pm \beta i$, where α and β are real and $i = \sqrt{-1}$. The results of Box 9.3 in the next chapter then show that dynamics of the variables are described by sinusoidal cycles with a period of $\tau = 2\pi/\beta$. Thus, when our predator-prey model has complex eigenvalues, the population sizes of each species will approach their equilibrium values while cycling, with a period $\tau = 4\pi\delta/\sqrt{\varepsilon a c \theta(4\delta^2 - \varepsilon a c \theta)}$. For the parameters used in Figure 8.3b ($\delta = 2$, $\theta = 10$, $\varepsilon = 1$, and $a\ c = 0.1$), this predicts a period of 6.5 generations, which is consistent with the observed cycles.

Example: Phillips' Model of HIV Dynamics

The above example illustrates a linear stability analysis for a two-variable model. It would be a bit misleading, however, for us to imply that stability analyses are always so straightforward. The calculations can sometimes be very tedious or even impossible for some models, particularly those having more than two dynamic variables. Thankfully some additional techniques are available, and here we introduce one known as the Routh-Hurwitz conditions (Box 8.2). Phillips' model of HIV infection from Chapter 1 is used as an example to illustrate the approach.

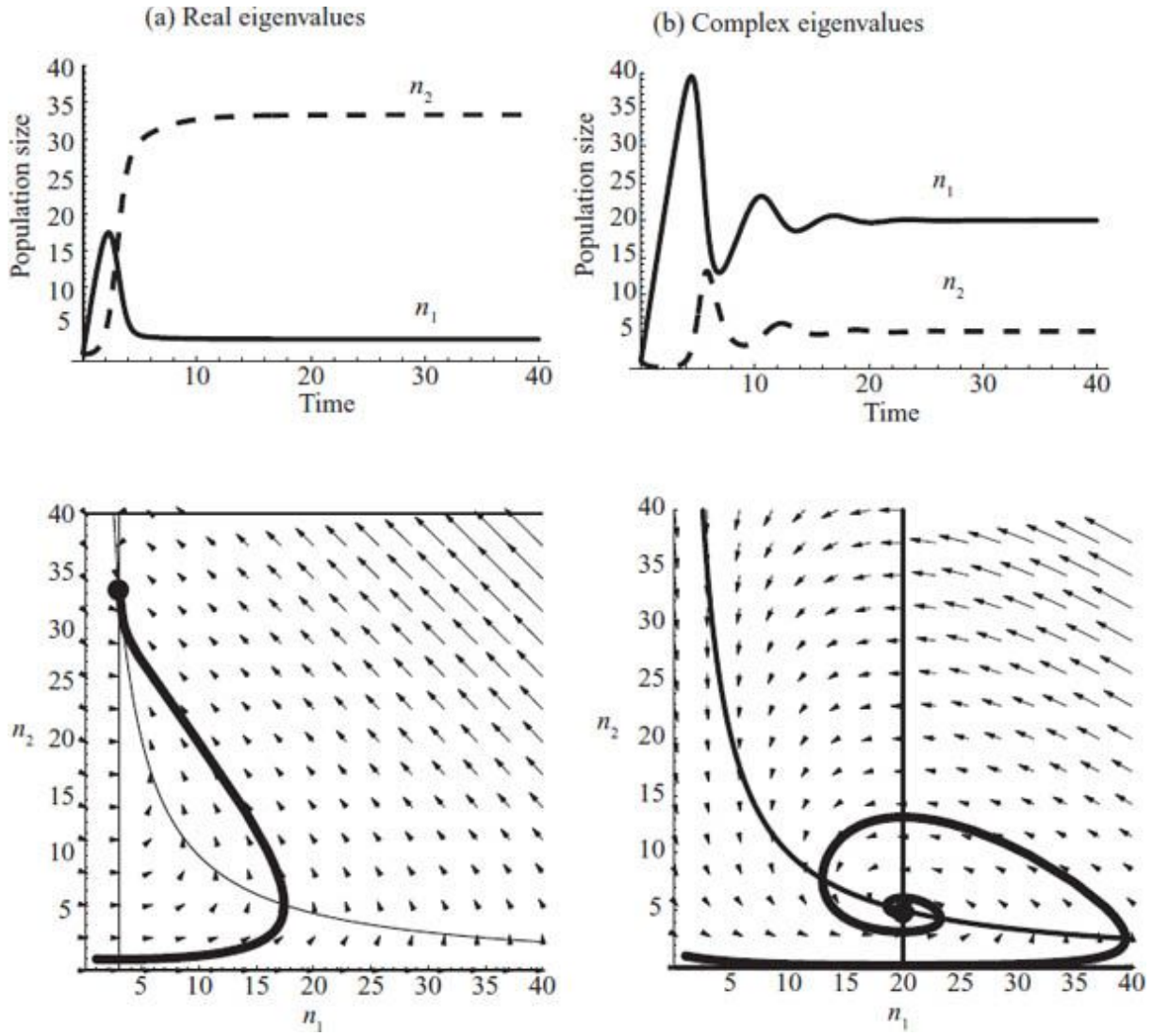

Figure 8.3: Examples of predator-prey dynamics. Dynamics over time (top) and phase plane diagrams (bottom) with (a) real eigenvalues ($\delta = 0.3$) or (b) complex eigenvalues ($\delta = 2$). In both cases, the dynamics are illustrated for 40 generations with $\theta = 10$, $\varepsilon = 1$, and $a\,c = 0.1$.

Box 8.2: Routh-Hurwitz Conditions for Local Stability

The Routh-Hurwitz conditions can be used to determine whether an equilibrium is stable without having to calculate the eigenvalues explicitly. Here we describe the Routh-Hurwitz conditions for models with different numbers of variables. We begin by assuming that the model is a continuous-time model, which is the most common context in which the Routh-Hurwitz conditions are used. Nevertheless, under certain circumstances, the Routh-Hurwitz conditions can be used to

determine stability in discrete-time models, as we describe at the end of this box.

Two-Dimensional Matrices: To begin, we consider a model with two variables. In this case, the Jacobian matrix is a 2 × 2:

$$\mathbf{J} = \begin{pmatrix} j_{11} & j_{12} \\ j_{21} & j_{22} \end{pmatrix}, \tag{8.2.1}$$

whose characteristic polynomial is a quadratic equation in r (see equations (7.11) in Chapter 7):

$$r^2 - (j_{11} + j_{22})\, r + (j_{11}\, j_{22} - j_{12}\, j_{21}) = 0. \tag{8.2.2a}$$

We can simplify the notation by defining the coefficients $a_1 = -(j_{11} + j_{22})$ and $a_2 = j_{11}\, j_{22} - j_{12}\, j_{21}$. Then we can rewrite (8.2.2a) as

$$r^2 + a_1 r + a_2 = 0 \tag{8.2.2b}$$

The eigenvalues of (8.2.1) are the values of r that satisfy equation (8.2.2b).

For a continuous-time model, the equilibrium is locally stable only if both eigenvalues have negative real parts. According to Rule P2.25, this will be true if the trace of (8.2.1) is negative ($j_{11} + j_{22} < 0$) and the determinant is positive ($j_{11}\, j_{22} - j_{12}\, j_{21} > 0$). We can rewrite these trace and determinant conditions in terms of the coefficients of the characteristic polynomial (8.2.2b):

$$\begin{aligned} a_1 &> 0, \\ a_2 &> 0. \end{aligned} \tag{8.2.3}$$

These are the Routh-Hurwitz conditions for a second-degree characteristic polynomial.

Three-Dimensional Matrices: Next we consider a 3 × 3 Jacobian matrix:

$$\mathbf{J} = \begin{pmatrix} j_{11} & j_{12} & j_{13} \\ j_{21} & j_{22} & j_{23} \\ j_{31} & j_{32} & j_{33} \end{pmatrix}. \tag{8.2.4}$$

The characteristic polynomial for (8.2.4) can be written in the form

$$r^3 + a_1 r^2 + a_2 r + a_3 = 0 \tag{8.2.5}$$

where the coefficients a_1, a_2, and a_3 are calculated from the elements of the matrix (8.2.4), just as the coefficients in (8.2.2b) were calculated from the elements of the matrix (8.2.1). The conditions on the coefficients for local stability are no longer easily identified as involving the trace or the determinant of (8.2.4). Nonetheless, it has been proven that the eigenvalues of (8.2.4) all have negative real parts if (and only if) the following three conditions are met (Edelstein-Keshet 1988):

$$\begin{aligned} a_1 &> 0, \\ a_3 &> 0, \\ a_1 a_2 &> a_3. \end{aligned} \tag{8.2.6}$$

If these conditions are met for the characteristic polynomial of a continuous-time model, then we are guaranteed that an equilibrium is locally stable. If one or more condition fails, however, the equilibrium will be unstable.

Four-Dimensional Matrices: For a 4×4 Jacobian matrix, the characteristic polynomial is a fourth-order polynomial, having the general form

$$r^4 + a_1 r^3 + a_2 r^2 + a_3 r + a_4 = 0. \tag{8.2.7}$$

The Routh-Hurwitz conditions on the coefficients of (8.2.7) that must be met for local stability are then (Edelstein-Keshet 1988)

$$\begin{aligned} a_1 &> 0, \\ a_3 &> 0, \\ a_4 &> 0, \\ a_1 a_2 a_3 &> a_3^2 + a_1^2 a_4. \end{aligned} \tag{8.2.8}$$

Again, all of these conditions must be met for an equilibrium to be locally stable in a continuous-time model.

n-Dimensional Matrices: There is a general recipe that can be followed for matrices of any size to obtain the Routh-Hurwitz conditions that must be met for local stability. As the dimension of the matrix gets large, however, the recipe becomes more and more difficult to apply (see Edelstein-Keshet 1988).

Consider the characteristic polynomial for an $n \times n$ matrix. It will be an nth-order polynomial, having the form

$$r^n + a_1 r^{n-1} + a_2 r^{n-2} + \cdots + a_n = 0, \tag{8.2.9}$$

where, again, the a_i's will be functions of the elements of the $n \times n$ matrix. To obtain the desired conditions on the coefficients of (8.2.9) we first need to construct the following n matrices:

$$\mathbf{H}_1 = (a_1), \tag{8.2.10a}$$

$$\mathbf{H}_2 = \begin{pmatrix} a_1 & 1 \\ a_3 & a_2 \end{pmatrix}, \tag{8.2.10b}$$

$$\mathbf{H}_3 = \begin{pmatrix} a_1 & 1 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{pmatrix}, \tag{8.2.10c}$$

$$\mathbf{H}_4 = \begin{pmatrix} a_1 & 1 & 0 & 0 \\ a_3 & a_2 & a_1 & 1 \\ a_5 & a_4 & a_3 & a_2 \\ a_7 & a_6 & a_5 & a_4 \end{pmatrix}, \tag{8.2.10d}$$

etc., where the i,j element in the $k \times k$ matrix $\mathbf{H}_k$ is given by

$$\begin{array}{ll} a_{2i-j} & \text{for } 0 < 2i - j < k, \\ 1 & \text{for } 2i = j, \\ 0 & \text{for } 2i < j \text{ or } 2i > k + j. \end{array} \tag{8.2.11}$$

An equilibrium will be stable in a continuous-time model (i.e., all eigenvalues will have negative real parts) if and only if the determinants of all n matrices are positive:

$$\text{Det}(\mathbf{H}_1) > 0, \text{Det}(\mathbf{H}_2) > 0, \ldots, \text{Det}(\mathbf{H}_n) > 0 \tag{8.2.12}$$

(Problem 8.13). In principle, the Routh-Hurwitz conditions can be used for Jacobian matrices of any size. In practice, even these conditions become cumbersome and difficult to use when the size of the Jacobian matrix gets large.

Discrete-Time Models: While the Routh-Hurwitz conditions can be directly applied to determine the stability of continuous-time models, they can also be used in special circumstances to assess the stability of equilibria in discrete-time models. We illustrate this point using a two-variable model, but the principle holds generally.

If the stability matrix for a discrete-time model with two variables is given by (8.2.1), its eigenvalues λ will solve the characteristic polynomial

$$\lambda^2 - (j_{11} + j_{22})\lambda + (j_{11} j_{22} - j_{12} j_{21}) = 0. \tag{8.2.13}$$

Stability of an equilibrium requires that each eigenvalue lies between -1 and $+1$. If all of the elements of the stability matrix are non-negative, however, then the Perron-Frobenius theorem (Appendix 3) guarantees that there is a real and positive eigenvalue larger in magnitude than (or equal to) all other eigenvalues. This eigenvalue will be the leading eigenvalue. For such nonnegative matrices, we need only focus on determining whether the leading eigenvalue falls above or below one. Equivalently, if we write λ as $1 + r$, we need only focus on determining whether r is positive or negative. While the Routh-Hurwitz conditions only tell us whether the real part of r is positive or negative, we are guaranteed by the Perron-Frobenius theorem that the

leading eigenvalue of a non-negative matrix is real, and hence its real part must be larger than or equal to the real parts of all remaining eigenvalues.

Making the substitution $\lambda = 1 + r$, we can rewrite (8.2.13) as

$$\begin{aligned}(1 + r)^2 - (j_{11} + j_{22})(1 + r) + (j_{11} j_{22} - j_{12} j_{21}) &= 0,\\ r^2 + (2 - j_{11} - j_{22})r + (1 - j_{11} - j_{22} + j_{11} j_{22} - j_{12} j_{21}) &= 0.\end{aligned} \tag{8.2.14}$$

Defining the coefficients $a_1 = (2 - j_{11} - j_{22})$ and $a_2 = 1 - j_{11} - j_{22} + j_{11} j_{22} - j_{12} j_{21}$, we are guaranteed that the solutions for r are negative and that the leading eigenvalue will be less than one in magnitude if the matrix satisfies the Perron-Frobenius theorem and the Routh-Hurwitz conditions (8.2.3) are met.

The same procedure can be followed for matrices of any size. First define λ as $1 + r$, then rewrite the characteristic polynomial in terms of r, and apply the Routh-Hurwitz conditions. Only if all of these conditions hold will the equilibrium be locally stable. This assumes that you know that the leading eigenvalue is real and positive, as is true for non-negative matrices (Appendix 3). These conditions come in handy for discrete-time models because non-negative matrices often arise in biological applications (e.g., the matrix (8.48b)).

We introduced Phillips' model for HIV dynamics within the body in Chapter 1 and derived equations for its four dynamic variables in Chapter 2 (Box 2.4). In Chapter 4 we explored the dynamical behavior of the model using graphical techniques. One of our findings was that there appeared to be a critical value of the transmission parameter β above which HIV infection takes hold and below which it is cleared from the body. This makes sense because there must be some level of transmission β ensuring that viruses infect new cells and replicate faster than they die off. But what is this critical value of β? We can identify this critical value in terms of the parameters of the model by performing a local stability analysis. These results then allow us to determine which factors most increase the critical value of β. Targeting these factors soon after exposure to HIV could boost the chances that the body is able to clear the infection.

The first step is to calculate the equilibria of the model (see equations 2.4.1). You can carry out these calculations yourself, but our main interest here is the possibility that there is an equilibrium at which $\hat{E} = 0$, $\hat{L} = 0$, and $\hat{V} = 0$ (i.e., the infection is absent from the body). Substituting these values into the system of equations (2.4.1) reveals that all variables remain constant at this point as long as the number of uninfected cells is $\hat{R} = \Gamma_T/\mu$.

We now want to determine the conditions on the transmission parameter β under which this equilibrium is stable, and HIV cannot become established within the body. To do so, we first calculate the Jacobian matrix using Definition 8.2. For Phillips' model, we obtain

$$\mathbf{J} = \begin{pmatrix} -V\beta - \mu & 0 & 0 & -R\beta \\ pV\beta & -\alpha - \mu & 0 & pR\beta \\ (1-p)V\beta & \alpha & -\delta & (1-p)R\beta \\ 0 & 0 & \pi & -\sigma \end{pmatrix}. \tag{8.23}$$

The above Jacobian matrix must now be evaluated at the equilibrium of interest, $\hat{R} = \Gamma_T/\mu$, $\hat{E} = 0$, $\hat{L} = 0$, and $\hat{V} = 0$, giving

$$\mathbf{J} = \begin{pmatrix} -\mu & 0 & 0 & -\Gamma\tau\beta/\mu \\ 0 & -\alpha - \mu & 0 & p\Gamma\tau\beta/\mu \\ 0 & \alpha & -\delta & (1-p)\Gamma\tau\beta/\mu \\ 0 & 0 & \pi & -\sigma \end{pmatrix}. \tag{8.24}$$

Next, we need to determine the eigenvalues of matrix (8.24). The eigenvalues are the roots r of the characteristic polynomial equation given by $\mathrm{Det}(\mathbf{J} - r\,\mathbf{I}) = 0$, which can be calculated using Rule P2.17 (we used *Mathematica*). For a 4×4 matrix, the characteristic polynomial is a fourth-degree polynomial in r and yields four roots (four eigenvalues). Unfortunately, identifying the eigenvalues from such fourth-order polynomial equations is difficult unless the polynomial factors. Fortunately, in this example, the characteristic polynomial *can* be factored into two terms:

$$(r + \mu)(r^3 + a_1 r^2 + a_2 r + a_3) = 0, \tag{8.25}$$

where

$$a_1 = \alpha + \delta + \mu + \sigma, \tag{8.26a}$$

$$a_2 = \mu\sigma + \alpha(\delta + \sigma) + \delta(\mu + \sigma) - \frac{(1 - p)\beta\Gamma\tau\pi}{\mu}, \tag{8.26b}$$

$$a_3 = \delta\sigma(\alpha + \mu) - \frac{(\alpha + \mu(1 - p))\beta\Gamma\tau\pi}{\mu}. \tag{8.26c}$$

One eigenvalue is easy to identify: $r = -\mu$. Because Phillips defined his parameters so that they are all positive (e.g., μ is the rate at which cells die or are eliminated from the body), this eigenvalue will always have a negative real part. But the other three eigenvalues are the roots of a cubic, which are harder to interpret. At times like these, the Routh-Hurwitz conditions (Box 8.2) come in very handy. These conditions are generalizations of the trace and determinant conditions (Rule P2.25) for 2 × 2 matrices, and tell us when the real parts of all eigenvalues are negative. Consequently, the Routh-Hurwitz conditions are often used to determine whether an equilibrium in a continuous-time model is locally stable.

To determine when the real parts of the remaining three eigenvalues are negative, we can use the Routh-Hurwitz conditions given by (8.2.6). The process of examining the three conditions required for stability (8.2.6) is somewhat tedious for this model, but we will go through it in detail to provide an illustration of how these conditions can be applied. Remember that we are trying to find the critical value of the transmission parameter β above which the equilibrium loses stability (and the virus takes hold in the body).

For the equilibrium without HIV to be stable requires four conditions to be met: $-\mu < 0$ (from the first eigenvalue), $a_1 > 0$, $a_3 > 0$, and $a_1 a_2 - a_3 > 0$ (from the Routh-Hurwitz conditions (8.2.6) for the remaining three eigenvalues). Biologically, it makes sense that the infection should not spread if β is low enough, in which case we expect each of these conditions to hold. Now imagine increasing β but holding all other parameters fixed. At some point, one of the conditions should fail, and the HIV-absent equilibrium should become unstable. We need to determine which of these conditions fails first.

We begin by examining the first Routh-Hurwitz condition: $a_1 > 0$. From (8.26a) we have $a_1 = \alpha + \delta + \mu + \sigma$, which does not involve β at all. In fact, under the assumption that all parameters are positive, a_1 is positive. Therefore, the first Routh-Hurwitz condition is always satisfied, and we can proceed to the next condition.

From (8.26c), the second Routh-Hurwitz condition, $a_3 > 0$, requires that

$$\delta \sigma (\alpha + \mu) > \frac{(\alpha + \mu (1 - p)) \beta \Gamma \tau \pi}{\mu} \tag{8.27a}$$

Rearranging (8.27a) as a condition on β, we find that this condition on stability is satisfied only for β values below

$$\beta_{a_3} = \frac{\delta \mu \sigma (\alpha + \mu)}{\Gamma \tau \pi (\alpha + \mu(1 - p))}. \tag{8.27b}$$

If β is greater than β_{a_3}, then the condition $a_3 > 0$ fails and the equilibrium with HIV absent will be unstable.

We now proceed to the third and final Routh-Hurwitz condition, $a_1 a_2 - a_3 > 0$. Factoring terms involving, and not involving, β we get

$$(\alpha + \delta + \mu)(\delta + \sigma)(\alpha + \mu + \sigma) - ((\delta + \sigma)(1 - p) - p \alpha) \frac{\beta \Gamma \tau \pi}{\mu} > 0. \tag{8.28a}$$

Whether or not (8.28a) holds is not obvious. To proceed, consider the two halves on the left-hand side of (8.28a). The first half is positive under the assumption that each parameter is positive. The second half is also positive if $(\delta + \sigma)(1 - p) - p \alpha$ is negative, in which case we are guaranteed that $a_1 a_2 - a_3 > 0$. If, however, $(\delta + \sigma)(1 - p) - p \alpha$ is positive, we have an additional requirement for stability: β must be less than

$$\beta_{a_1 a_2 - a_3} = \frac{\mu(\alpha + \delta + \mu)(\delta + \sigma)(\alpha + \mu + \sigma)}{\Gamma \tau \pi ((\delta + \sigma)(1 - p) - p\alpha)} \tag{8.28b}$$

If β is larger than $\beta_{a_1a_2-a_3}$, then the condition $a_1a_2 - a_3 > 0$ fails, and the equilibrium will be unstable.

Now we need to determine which cutoff, $\beta_{a_1a_2-a_3}$ or β_{a_3}, is smaller, because it is the smaller cutoff that determines when the HIV-absent equilibrium first becomes unstable. To do this, we can try to factor $\beta_{a_1a_2-a_3} - \beta_{a_3}$ to see if the result is clearly positive or negative. In this case, the result does factor:

$$\beta_{a_1a_2-a_3} - \beta_{a_3} =$$

$$\frac{\mu(\alpha + \delta + \mu + \sigma)(\alpha\delta\sigma p + (\alpha\delta + \delta\mu + \alpha\sigma + \mu\sigma)(\alpha + \mu(1 - p)))}{\Gamma\tau\pi(\alpha + \mu(1 - p))((\delta + \sigma)(1 - p) - p\alpha)} \quad (8.29)$$

The numerator in (8.29) is positive because all of the parameters are positive and because p is a proportion. The denominator is also positive as long as $(\delta + \sigma)(1 - p) - p\alpha$ is positive. In this case, $\beta_{a_1a_2-a_3}$ will be greater than β_{a_3}. If $(\delta + \sigma)(1 - p) - p\alpha$ is negative then $a_1a_2 - a_3 > 0$ is necessarily satisfied, and we are left with the single condition β_{a_3}. Thus, either $\beta_{a_1a_2-a_3}$ is larger than β_{a_3} or there is only one critical value β_{a_3}. In either case, β_{a_3} is the value of β at which stability is lost.

We must admit that it took some effort to write (8.29) in a way that was clearly positive. When you suspect that an equation is positive, the first thing to do is to check numerically that it is positive over a range of values for each parameter. There is no point in trying to prove an equation is positive if it isn't! A helpful technique is to identify terms that are positive but that include negative terms within them. For example, $(1 - p)$ is positive if p is a proportion. You can then try to rewrite negative parts of your equation (e.g., involving $-p$) in terms of these positive quantities. Determining the sign of terms is often the most time-consuming part of a stability analysis; success requires persistence and a lot of trial and error.

In summary, the Routh-Hurwitz conditions have shown us that HIV can take hold in the body only if the transmission parameter is greater than β_{a_3}; i.e.,

$$\beta > \frac{\delta\mu\sigma\,(\alpha + \mu)}{\Gamma\pi\tau\,(\alpha + (1 - p)\mu)}. \tag{8.31}$$

In our numerical analysis, we found that the HIV-absent equilibrium first became unstable around $\beta = 0.000034$ per susceptible cell per day. Using (8.31), we now have a more precise estimate of the critical value of the transmission parameter: $\beta = 0.0000331$.

More importantly, we now have a general formula that specifies how large the transmission parameter must be in order for HIV to establish itself within the body. With such a general condition in hand, we can determine how each parameter affects the critical transmission rate (Figure 8.4). For example, increasing the daily death rate of viruses in the bloodstream (σ) makes it harder for the virus to invade (a higher critical value of the transmission rate is required). More surprising, HIV is less likely to establish within the body if CD4+ cells die at a higher rate (higher δ and μ) or are produced at a lower rate (lower Γ and τ). Although a rigorous immune system and a high CD4+ count are necessary for long-term survival, in the short term, a healthy CD4+ population provides a large reservoir of potential cells for HIV to attack, and it therefore makes it easier for HIV to invade. This suggests a counterintuitive drug strategy; drugs that target and remove CD4+ cells might help the body ward off HIV in the first few days after exposure.

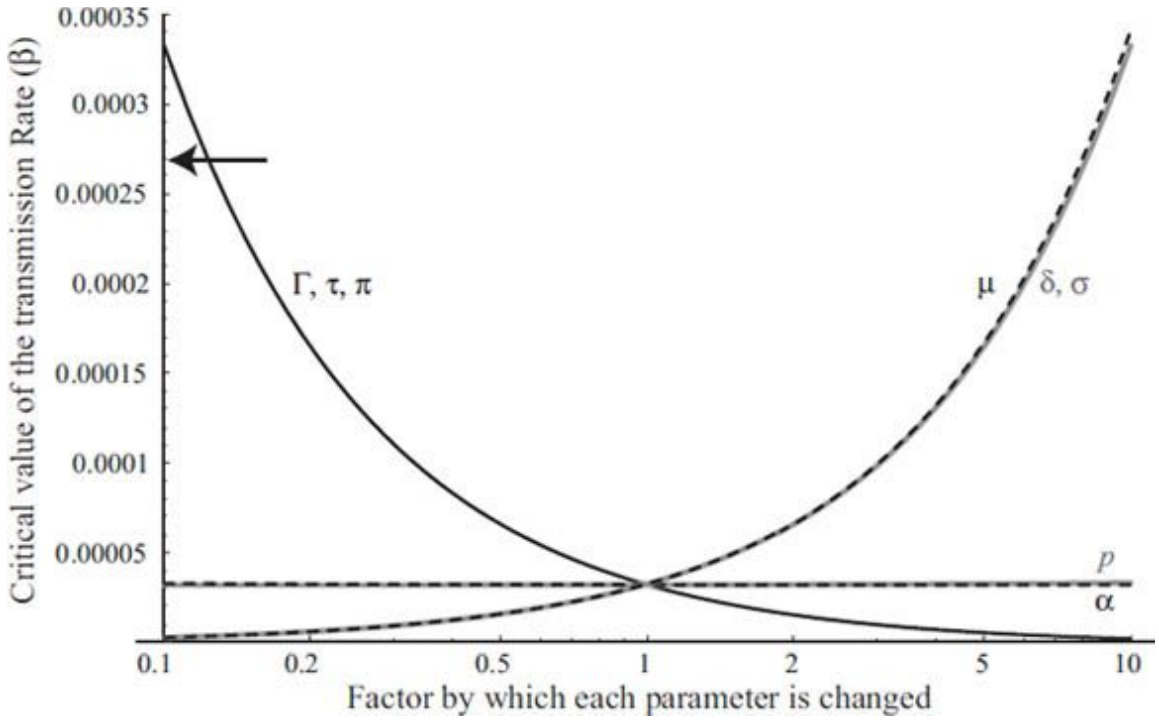

Figure 8.4: The critical value of the transmission rate for HIV to become established. The critical value of β (8.31) above which viral counts are expected to increase is shown as each parameter is varied by a factor from 0.1 to 10 (horizontal axis). Except for the parameter being varied, the parameters are those used by Phillips (1996): $\Gamma = 1.36$, $\mu = 1.36 \times 10^3$, $\tau = 0.2$, $p = 0.1$, $\alpha = 3.6 \times 10^{-2}$, $\sigma = 2$, $\delta = 0.33$, $\pi = 100$. His choice of $\beta = 0.00027$ is marked by the arrow. The critical value is fairly insensitive to p, the probability that HIV becomes latent within a newly infected cell, and to α, the rate at which latently infected cells become activated (these parameters affect the speed of invasion, but not whether invasion occurs). In contrast, the critical value rises rapidly, making it more difficult for HIV to become established within the body, if the death rate of infected (δ) or uninfected (α) CD4+ cells increases or if the death rate of viruses in the bloodstream (σ) increases. Conversely, the critical value declines and becomes easier to satisfy if the birth rate (Γ) or maturation rate (τ) of CD4+ cells increases or if the rate of production of virus particles per actively infected cell increases (π).

8.3 Equilibria and Stability for Nonlinear Discrete-Time Models

We now turn our attention to nonlinear models in discrete time. We focus on models with two variables and generalize these results to models with more variables in Recipes 8.3 and 8.4. The method is very similar to that used in continuous time, with the main difference being the way in which the eigenvalues are interpreted to determine stability.

Consider a general, two-variable, nonlinear model having the form

$$\begin{aligned} x_1(t+1) &= f(x_1(t), x_2(t)), \\ x_2(t+1) &= g(x_1(t), x_2(t)), \end{aligned} \tag{8.32}$$

where f and g are arbitrary functions of the variables x_1 and x_2. As with continuous-time models, the first thing we must do is find the equilibria of the model, and again there can be several. At each equilibrium the variables must be unchanging, and therefore $x_1(t+1) = x_1(t) = \hat{x}_1$ and $x_2(t+1) = x_2(t) = \hat{x}_2$. This gives two equilibrium conditions, $f(\hat{x}_1, \hat{x}_2) = \hat{x}_1$ and $g(\hat{x}_1, \hat{x}_2) = \hat{x}_2$, which must both be satisfied simultaneously at any equilibrium.

Now we would like to know what happens to the system if we start it very close to an equilibrium. Will it move in toward the equilibrium (and hence that equilibrium is locally stable), or will it move away (and hence that equilibrium is unstable)? Again, we will derive the method used to assess stability, but this method results in a recipe (Recipe 8.3) that can be used without rederiving it every time.

Let us start the system a small amount ε_1 from $\hat{x}_1$, and a small amount ε_2 from $\hat{x}_2$, so that $\varepsilon_1 = x_1(t) - \hat{x}_1$ and $\varepsilon_2 = x_2(t) - \hat{x}_2$ are the deviations of x_1 and x_2 from their equilibrium values. If the equilibrium point $(\hat{x}_1, \hat{x}_2)$ is locally stable, then these deviations ε_1 and ε_2 must decay to zero as time passes (meaning that $x_1(t) \to \hat{x}_1$ and $x_2(t) \to \hat{x}_2$ as time passes). To determine whether or not an equilibrium is stable, we need a pair of equations that tell us the dynamics of ε_1 and ε_2. These can be obtained as

$$\begin{aligned} \varepsilon_1(t+1) &= x_1(t+1) - \hat{x}_1 \\ &= f(x_1(t), x_2(t)) - \hat{x}_1 \\ &= f(\hat{x}_1 + \varepsilon_1(t), \hat{x}_2 + \varepsilon_2(t)) - \hat{x}_1 \end{aligned} \tag{8.33a}$$

and

$$\begin{aligned} \varepsilon_2(t+1) &= x_2(t+1) - \hat{x}_2 \\ &= g(x_1(t), x_2(t)) - \hat{x}_2 \\ &= g(\hat{x}_1 + \varepsilon_1(t), \hat{x}_2 + \varepsilon_2(t)) - \hat{x}_2. \end{aligned} \tag{8.33b}$$

We can again write the multiple-variable Taylor series (Box 8.1) for each of the functions f and g near the equilibrium point $(\hat{x}_1, \hat{x}_2)$, just as we did for continuous-time models. The Taylor series of (8.33a) with respect to ε_1 and ε_2 near (0,0) is

$$\varepsilon_1(t+1) = \left(f(\hat{x}_1, \hat{x}_2) + \left.\frac{\partial f}{\partial x_1}\right|_{x_1=\hat{x}_1,\, x_2=\hat{x}_2} \varepsilon_1(t) + \left.\frac{\partial f}{x_2}\right|_{x_1=\hat{x}_1,\, x_2=\hat{x}_2} \varepsilon_2(t) + \text{higher-power terms} \right) - \hat{x}_1 \tag{8.34a}$$

$$= \left.\frac{\partial f}{\partial x_1}\right|_{x_1=\hat{x}_1,\, x_2=\hat{x}_2} \varepsilon_1(t) + \left.\frac{\partial f}{\partial x_2}\right|_{x_1=\hat{x}_1,\, x_2=\hat{x}_2} \varepsilon_2(t) + \text{higher-power terms.} \tag{8.34b}$$

Expression (8.34b) follows from (8.34a) and from the fact that $f(\hat{x}_1, \hat{x}_2) = \hat{x}_1$ at equilibrium. "Higher-power terms" in the above expression are terms that involve higher powers of ε_1 and ε_2, and the derivatives in this expression are evaluated at the equilibrium (hence they are constants). Consequently, if we start the system very close to the equilibrium, so that ε_1 and ε_2 are very small, then all the terms having higher powers in these deviations will be extremely small and can be ignored. This gives

$$\varepsilon_1(t+1) = \left.\frac{\partial f}{\partial x_1}\right|_{x_1=\hat{x}_1,\, x_2=\hat{x}_2} \varepsilon_1(t) + \left.\frac{\partial f}{\partial x_2}\right|_{x_1=\hat{x}_1,\, x_2=\hat{x}_2} \varepsilon_2(t). \tag{8.34c}$$

These calculations can be repeated to obtain a linear approximation for the function g in (8.33b) near the equilibrium point $(\hat{x}_1, \hat{x}_2)$:

$$\varepsilon_2(t+1) = \left.\frac{\partial g}{\partial x_1}\right|_{x_1=\hat{x}_1,\, x_2=\hat{x}_2} \varepsilon_1(t) + \left.\frac{\partial g}{\partial x_2}\right|_{x_1=\hat{x}_1,\, x_2=\hat{x}_2} \varepsilon_2(t). \tag{8.34d}$$

Finally, we can write this pair of recursions equations in matrix form as

$$\begin{pmatrix} \varepsilon_1(t+1) \\ \varepsilon_2(t+1) \end{pmatrix} = \left. \begin{pmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} \\ \dfrac{\partial g}{\partial x_1} & \dfrac{\partial g}{\partial x_2} \end{pmatrix} \right|_{x_1 = \hat{x}_1, x_2 = \hat{x}_2} \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{pmatrix}, \tag{8.35}$$

The matrix in (8.35) is again a Jacobian matrix and is sometimes referred to as a *stability matrix*. The system of equations is linear in the two deviations ε_1 and ε_2. Thus, (8.35) allows us to describe the nonlinear dynamics in terms of linear equations near an equilibrium of interest. Having done so, we can apply the techniques described in Chapter 7 for linear models. If the deviations ε_1 and ε_2 decay to zero over time, then the equilibrium is locally stable. Conversely, if the deviations ε_1 and ε_2 grow over time, then the equilibrium is unstable. Specifically, the equilibrium is stable if the absolute value of every eigenvalue is less than one ($|\lambda_1| < 1$ and $|\lambda_2| < 1$; see section 7.4). And, again, we must remember to repeat this stability analysis for every one of the equilibria of interest.

We summarize the above methods and generalize them to models with multiple variables in the following definitions and rules:

Definition 8.3: A General Nonlinear Model in Discrete Time

A general, nonlinear, discrete-time model with n dynamic variables $x_1, \ldots, x_n$ can be written as

$$\begin{aligned} x_1(t+1) &= f_1(x_1(t), x_2(t), \cdots, x_n(t)), \\ x_2(t+1) &= f_2(x_1(t), x_2(t), \cdots, x_n(t)), \\ &\vdots \qquad\qquad \vdots \\ x_n(t+1) &= f_n(x_1(t), x_2(t), \cdots, x_n(t)), \end{aligned}$$

where $f_1, f_2, \ldots, f_n$ denote different functions that map the variables from one time step to the next.

Recipe 8.3 then describes how to find equilibria of such models:

Recipe 8.3

Equilibria of a Nonlinear Multivariable Model in Discrete Time

Equilibria are found by determining the values of the variables that cause all of the variables to be the same in the next time step: $f_1(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = \hat{x}_1, f_2(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = \hat{x}_2, \ldots, f_n(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = \hat{x}_n$. This results in n equations in n unknowns. Any point that satisfies *all* of these conditions simultaneously is an equilibrium. To identify the equilibria:

Step 1: Factor each equation $f_i(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) - \hat{x}_i = 0$.

Step 2: Identify all possible solutions to one of these equations (start with the simplest one).

Step 3: Plug each possible solution into the remaining equations and repeat the above steps until equilibrium values for all variables are identified.

For nonlinear models, there may be more than one equilibrium. Depending on the complexity of the model, it may or may not be possible to identify all of the equilibria explicitly.

Recipe 8.4 describes how to assess the stability of equilibria:

Recipe 8.4

Stability of a Nonlinear Multivariable Model in Discrete Time

To determine whether an equilibrium of interest is stable:

Step 1: Evaluate the Jacobian matrix (Definition 8.2) at the equilibrium of interest $\mathbf{J}|_{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n}$. This matrix is often called the *local stability matrix*.

Step 2: Solve the characteristic polynomial $\text{Det}\,(\mathbf{J}|_{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n} - \lambda \mathbf{I}) = 0$, which is an nth-degree polynomial.

Step 3: The n solutions ("roots") to this characteristic polynomial are the n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Step 4: The equilibrium will be locally stable if the absolute values of all n eigenvalues are less than one.

- For complex eigenvalues $\lambda = A \pm Bi$, stability requires that the absolute value $\sqrt{A^2 + B^2}$ be less than one.
- For real eigenvalues, stability requires that both $\lambda < 1$ and $-1 < \lambda$.

Equivalently, the equilibrium is locally stable if the eigenvalue with the largest absolute value (*the leading eigenvalue*) has an absolute value less than one. (If the leading eigenvalue equals one exactly, the local stability analysis is inconclusive, and higher-order terms must be considered.)

Once again, whether the eigenvalues are real or complex provides information about the behavior near the equilibrium. If they are complex, then the system will spiral around the equilibrium along some axes (for details, see Box 9.2).

Example: Density-Dependent Natural Selection

We now apply the above methods to investigate how ecological interactions within a species might evolve. Specifically, we will ask when a new allele that experiences competition in a different manner from a resident allele can invade a population. To answer this question, we use the recursions developed in Problem 3.17 of Chapter 3 that track the dynamics of population size and the frequency of an allele that affects reproductive success. This two-variable model has the form

$$\begin{aligned} N(t+1) &= \overline{W}\, N(t), \\ p(t+1) &= \frac{p(t)\, W_{AA}(N(t)) + (1 - p(t))\, W_{Aa}(N(t))}{\overline{W}}\, p(t). \end{aligned} \tag{8.36}$$

Here $N(t)$ and $p(t)$ are the population size and the frequency of allele A at time t. $\overline{W}$ is the mean fitness of the population and is a function of the current population size and allele frequency: $\overline{W} = p^2\, W_{AA}(N) + 2p(1-p)\, W_{Aa}(N) + (1-p)^2\, W_{aa}(N)$.

The $W_{ij}(N)$ are the fitnesses of individuals with genotype ij, which are assumed to be decreasing functions of the population size due to competition for resources. We assume that some genotypes are more sensitive to competition than others; for example, some individuals might be better able to switch to alternative resources in the face of competition, or they might be less likely to engage in lethal battles over resources. There are many possible ways in which we could model how fitness declines with the population size (some are listed in Problem 5.1). Here, we use the functions $W_{AA}(N) = (1 + r)e^{-\alpha_{AA}N}$, $W_{Aa}(N) = (1 + r)e^{-\alpha_{Aa}N}$, and $W_{aa}(N) = (1 + r)e^{-\alpha_{aa}N}$, which imply that fitness declines exponentially with population size. This is similar to the Ricker model but now the rate of decline depends on genotype (a large α_{ij} means that the fitness of genotype ij decreases very quickly with increases in population size; Figure 8.5). The term r represents the intrinsic growth rate, which we assume is positive and the same for all genotypes.

We begin by finding equilibria by setting $N(t + 1) = N(t) = \hat{N}$ and $p(t + 1) = p(t) = \hat{p}$, giving the equations

$$\begin{aligned}
\hat{N} &= \left(\hat{p}^2 W_{AA}(\hat{N}) + 2\hat{p}(1 - \hat{p}) W_{Aa}(\hat{N}) + (1 - \hat{p})^2 W_{aa}(\hat{N})\right)\hat{N}, \\
\hat{p} &= \frac{\hat{p} W_{AA}(\hat{N}) + (1 - \hat{p}) W_{Aa}(\hat{N})}{\hat{p}^2 W_{AA}(\hat{N}) + 2\hat{p}(1 - \hat{p}) W_{Aa}(\hat{N}) + (1 - \hat{p})^2 W_{aa}(\hat{N})}\hat{p}.
\end{aligned} \tag{8.37}$$

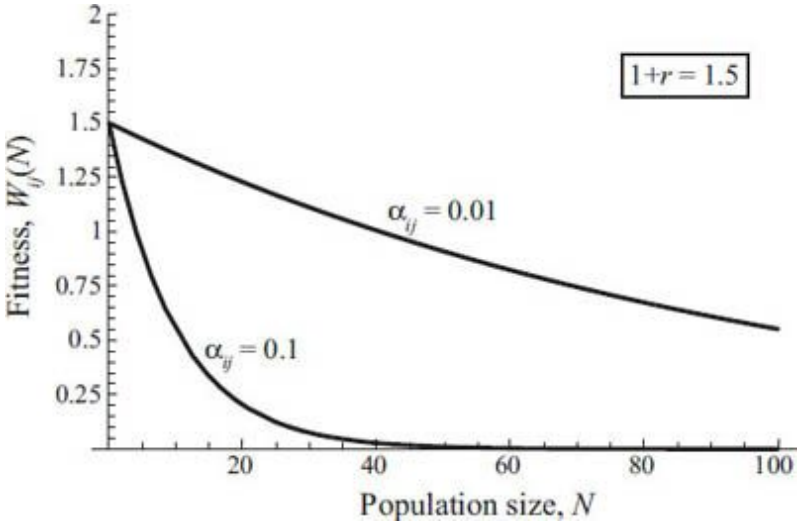

Figure 8.5: Fitness as a function of population size. The fitness function, $W_{ij}(N) = (1 + r)\, e^{-\alpha_{ij}N}$, is plotted against population size, N. Fitness declines more quickly with population size when α_{ij} is larger.

We have not yet substituted the specific forms of W_{ij} into the equilibrium conditions, but doing so produces a pair of complicated equations. We could analyze these expressions to obtain all of the possible equilibria (Recipe 8.3), but instead we will search for a specific equilibrium that is relevant to our biological question (Problem 8.2 gives more practice obtaining equilibria using the more formal route).

We are interested in knowing whether a rare allele can invade a resident population with a different competition coefficient, so let us focus on finding an equilibrium where the resident allele is fixed; i.e., $\hat{p} = 1$. Because there is no mutation in the model, the population must remain at $p = 1$ (allele A fixed) if it ever reaches this point. Substituting $\hat{p} = 1$ into the equilibrium condition for the population size gives $\hat{N} = W_{AA}(\hat{N})\,\hat{N}$, where $W_{AA}(\hat{N}) = (1 + r)e^{-\alpha_{AA}\hat{N}}$. One equilibrium of this model is $\hat{N} = 0$, representing extinction, and a second equilibrium occurs at $\hat{p} = 1$ and $\hat{N} = \ln(1 + r)/\alpha_{AA}$. We shall focus on the equilibrium with a positive population size. It is the stability of this equilibrium that interests us, and we need not search for any remaining equilibria.

We now perform a stability analysis of the equilibrium with $\hat{p} = 1$ and $\hat{N} = \ln(1 + r)/\alpha_{AA}$. Biologically, if this equilibrium is unstable, then either the population can be invaded by the a allele, or the population size moves away from $\hat{N} = \ln(1 + r)/\alpha_{AA}$ (or both). The general Jacobian matrix is rather large, but when evaluated at $\hat{p} = 1$, $\hat{N} = \ln(1 + r)/\alpha_{AA}$ it simplifies to

$$\mathbf{J} = \begin{pmatrix} 1 - \ln(1 + r) & \dfrac{2(1 - (1 + r)^{1-(\alpha_{Aa}/\alpha_{AA})})\ln(1 + r)}{\alpha_{AA}} \\ 0 & (1 + r)^{1-(\alpha_{Aa}/\alpha_{AA})} \end{pmatrix}. \qquad (8.38)$$

Because (8.38) is upper triangular, its eigenvalues are $\lambda_1 = 1 - \ln(1 + r)$ and $\lambda_2 = (1 + r)^{1-(\alpha_{Aa}/\alpha_{AA})}$ (Rule P2.12).

According to Recipe 8.4, the equilibrium will be stable if both eigenvalues are less than one in absolute value. Because we assumed that the

intrinsic growth rate is positive, $\ln(1 + r)$ will always be positive, so that $\lambda_1 = 1 - \ln(1 + r) < 1$. Even so, we must also check to make sure that $-1 < \lambda_1$ for the equilibrium to be stable. This condition requires that $\ln(1 + r) < 2$, implying that r must be less than $e^2 - 1$. This indicates that the rate of population growth must be small enough if the equilibrium is to be stable; the same condition was found in Problem 5.7, where we analyzed the local stability of the standard Ricker model with only one genotype. If the intrinsic growth rate is too large, the population overshoots the carrying capacity by such a large degree that the equilibrium becomes unstable.

Stability of the equilibrium with $\hat{p} = 1$ and $\hat{N} = \ln(1 + r)/\alpha_{AA}$ also requires that the second eigenvalue λ_2 be less than one in absolute value. Because $(1 + r)$ is positive, $\lambda_2 = (1 + r)^{1-(\alpha_{Aa}/\alpha_{AA})}$ must be positive. Therefore, we need only check that $(1 + r)^{1-(\alpha_{Aa}/\alpha_{AA})} < 1$ for stability. Taking the natural logarithm of both sides, stability requires that $(1 - \alpha_{Aa}/\alpha_{AA}) \ln(1 + r) < 0$. Given the assumption of a positive intrinsic growth rate, $\ln(1 + r) > 0$, and this condition is met only if $\alpha_{AA} < \alpha_{Aa}$. When $\alpha_{AA} < \alpha_{Aa}$, the fitness of AA individuals declines less rapidly with population size than the fitness of Aa individuals. Furthermore, according to our fitness functions, this stability condition implies that $W_{AA}(N)$ is greater than $W_{Aa}(N)$ at any population size. This condition is thus analogous to the requirement that $W_{AA} > W_{Aa}$ for stability in the one-gene diploid model of selection in a population of constant size (see discussion after equation (5.22)).

Overall, the resident equilibrium is locally stable only if both $r < e^2 - 1$ and $\alpha_{AA} < \alpha_{Aa}$ (Figure 8.6). The first condition depends only on the intrinsic growth rate and not on attributes of each genotype, while the second condition depends only on how sensitive each genotype is to population size. Intuitively, the first condition determines whether the system is ecologically stable at the genetic equilibrium ($\hat{p} = 1$), while the second condition determines whether the system is genetically stable at the ecological equilibrium ($\hat{N} = \ln(1 + r)/\alpha_{AA}$). Accurate predictions about stability can be made only by considering both the displacement from the ecological equilibrium *and* the displacement from the genetic equilibrium whenever growth rates depend on genotype.

Assuming that the ecological equilibrium is stable in the resident population before the introduction of the new allele, we can use these

stability conditions to answer our original question. A new allele will invade a population whenever it is less sensitive to competition than the resident.

Example: The Evolutionary Dynamics of Two Genes

The genomes of most organisms contain thousands of genes, yet the evolutionary models we have considered so far have tracked changes at one gene in isolation of the rest of the genome. Is it reasonable to ignore neighboring genes? Do the evolutionary dynamics of one gene depend on the evolutionary dynamics of other genes in important ways? Ideally, to answer these questions, we would model a whole genome (see, for example, Barton 1995; Barton and Turelli 1991; Kirkpatrick et al. 2002; Turelli and Barton 1990). Yet we can get a lot of insight by just moving up from one gene to two (see also Bürger 2000; Christiansen 2000; Crow and Kimura 1970; Karlin 1975). If adding a second gene causes major changes to the results, then we know that one-gene models are potentially misleading.

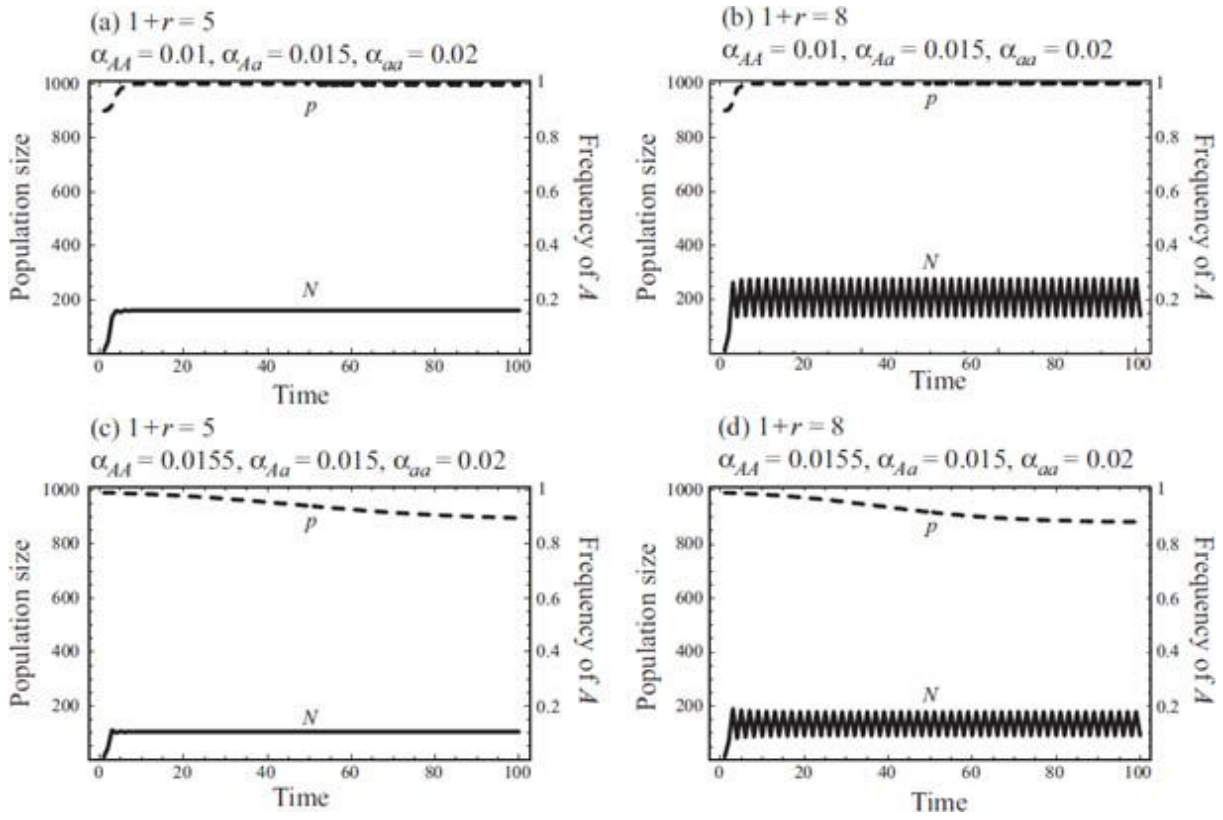

Figure 8.6: The evolution of competition coefficients. Plots of the population size (solid) and allele frequency (dashed) dynamics for model (8.27) with initial conditions near the equilibrium point, $\hat{p} = 1, \hat{N} = \ln(1 + r)/\alpha_{AA}$. (a) The equilibrium is stable because both $1 + r < e^2$ and $\alpha_{AA} < \alpha_{Aa}$.

(b) The equilibrium is unstable because $1 + r > e^2$. (c) The equilibrium is unstable because $\alpha_{AA} > \alpha_{Aa}$, (d) The equilibrium is unstable because both $1 + r > e^2$ and $\alpha_{AA} > \alpha_{Aa}$.

We denote the two genes by **A** and **B** and assume that there are two alleles at each gene, which we denote A_1, A_2 and B_1, B_2. There are thus four possible combinations of these alleles that can be found on any chromosome: A_1B_1 (with frequency x_1), A_1B_2 (with frequency x_2), A_2B_1 (with frequency x_3), and A_2B_2 (with frequency x_4). The frequency of these four types must sum to one (i.e., $x_1 + x_2 + x_3 + x_4 = 1$).

Let us now consider the life cycle given in Figure 8.7. If we assume that gametes (sperm and eggs) come together at random, then we can census the population at the gamete stage where there are only four types of chromosomes (as opposed to all the possible combinations of diploid individuals: A_1B_1/A_1B_1, A_1B_1/A_1B_2, etc.). Developing the recursions for this model requires a lot of bookkeeping as we go around the life cycle, which is best organized in the form of a table (Table 8.1). In the gamete pool, there are male gametes (sperm) and female gametes (eggs), each of which contains one of the four possible chromosome types. As a result, there are $4 \times 4 = 16$ different kinds of unions that produce diploid individuals (given in the first column of Table 8.1). If eggs and sperm unite at random to form diploid individuals, then the frequency of these different unions is given in the second column of Table 8.1.

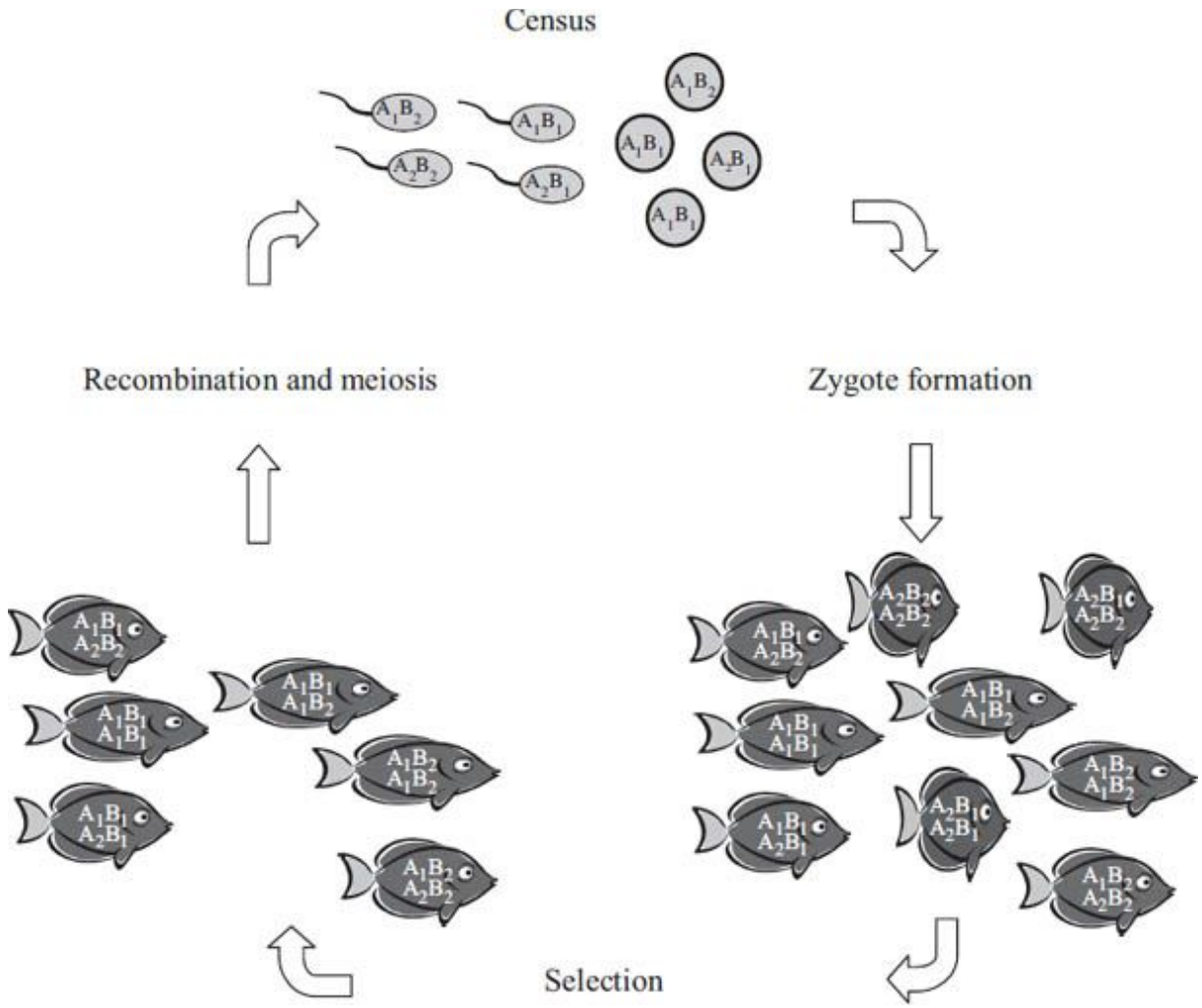

Figure 8.7: Life cycle diagram for the two-gene model

The next phase in the life cycle involves natural selection, and we suppose that certain diploid genotypes survive better than others. There are many different ways in which we might assign fitnesses to the different genotypes. Table 8.2 describes a very general scheme in which the fitness of an individual depends on whether it inherited a particular chromosome from its mother or father. For most genes, this parent-of-origin effect is absent (in which case $w_{ij} = w_{ji}$) but we allow for this possibility here to keep the model more general. The third column of Table 8.1 specifies the frequency of the different possible diploid individuals after selection has acted.

At this point, gametes are produced. In our previous models involving a single gene, we did not worry about where, in the genome, the gene of interest was found. Now we must be more explicit because the types of gametes that can be produced by a parent depend on the distance between

the two genes within the genome. If the two genes are on different chromosomes (or very far apart on the same chromosome), then the alleles will assort independently of one another according to Mendel's second law (Figure 8.8). If they are on the same chromosome, a chromosome could be passed intact to a gamete, carrying the same combination of alleles as found on a parental chromosome. Alternatively, recombination can occur in the parent, producing a chromo-some with a different combination of alleles (Figure 8.9). We can develop a general model that allows for any genomic arrangement by letting r equal the probability that recombination occurs between the two genes, where $r = 1/2$ corresponds to the case of independent assortment (see Figure 8.9).

TABLE 8.1:
Life cycle table with mating and selection

Female × male	Frequency of union	Frequency after selection[a]	Gamete frequencies			
			A_1B_1	A_1B_2	A_2B_1	A_2B_2
$A_1B_1 \times A_1B_1$	x_1x_1	$x_1x_1\frac{w_{11}}{\overline{w}}$	1	0	0	0
$A_1B_1 \times A_1B_2$	x_1x_2	$x_1x_2\frac{w_{12}}{\overline{w}}$	1/2	1/2	0	0
$A_1B_1 \times A_2B_1$	x_1x_3	$x_1x_3\frac{w_{13}}{\overline{w}}$	1/2	0	1/2	0
$A_1B_1 \times A_2B_2$	x_1x_4	$x_1x_4\frac{w_{14}}{\overline{w}}$	$(1-r)/2$	$r/2$	$r/2$	$(1-r)/2$
$A_1B_2 \times A_1B_1$	x_2x_1	$x_2x_1\frac{w_{21}}{\overline{w}}$	1/2	1/2	0	0
$A_1B_2 \times A_1B_2$	x_2x_2	$x_2x_2\frac{w_{22}}{\overline{w}}$	0	1	0	0
$A_1B_2 \times A_2B_1$	x_2x_3	$x_2x_3\frac{w_{23}}{\overline{w}}$	$r/2$	$(1-r)/2$	$(1-r)/2$	$r/2$
$A_1B_2 \times A_2B_2$	x_2x_4	$x_2x_4\frac{w_{24}}{\overline{w}}$	0	1/2	0	1/2
$A_2B_1 \times A_1B_1$	x_3x_1	$x_3x_1\frac{w_{31}}{\overline{w}}$	1/2	0	1/2	0
$A_2B_1 \times A_1B_2$	x_3x_2	$x_3x_2\frac{w_{32}}{\overline{w}}$	$r/2$	$(1-r)/2$	$(1-r)/2$	$r/2$
$A_2B_1 \times A_2B_1$	x_3x_3	$x_3x_3\frac{w_{33}}{\overline{w}}$	0	0	1	0
$A_2B_1 \times A_2B_2$	x_3x_4	$x_3x_4\frac{w_{34}}{\overline{w}}$	0	0	1/2	1/2
$A_2B_2 \times A_1B_1$	x_4x_1	$x_4x_1\frac{w_{41}}{\overline{w}}$	$(1-r)/2$	$r/2$	$r/2$	$(1-r)/2$
$A_2B_2 \times A_1B_2$	x_4x_2	$x_4x_2\frac{w_{42}}{\overline{w}}$	0	1/2	0	1/2
$A_2B_2 \times A_2B_1$	x_4x_3	$x_4x_3\frac{w_{43}}{\overline{w}}$	0	0	1/2	1/2
$A_2B_2 \times A_2B_2$	x_4x_4	$x_4x_4\frac{w_{44}}{\overline{w}}$	0	0	0	1

[a] $\overline{w}$ is chosen such that all elements in the column "Frequency after selection" sum to 1.

The final four columns of Table 8.1 specify the fraction of each type of gamete produced by a diploid individual. For example, consider an A_1B_1/A_2B_2 parent (row 4 of Table 8.1). With probability $1 - r$, recombination does not occur between the two genes (see Figure 8.9), in which case half of the gametes carry the A_1B_1 chromosome and half carry the A_2B_2 chromosome. Thus, we place $(1 - r)/2$ in both the fourth and seventh columns of row 4. With probability r, recombination does occur between the two genes, in which case half of the gametes carry the A_1B_2 chromosome

and half carry the A_2B_1 chromosome. Thus, we place $r/2$ in both the fifth and sixth columns of row 4. The calculations for the other rows are analogous, but the recombination rate only matters for some of the genotypes. For example, an A_1B_1/A_1B_2 parent can produce only A_1B_1 and A_1B_2 gametes regardless of the rate of recombination. Only when both genes are heterozygous does recombination affect the array of gametes produced. This property serves as a useful check when deriving models with recombination. Another important check is to make sure that the sum of the gamete frequencies produced by any one genotype is one. Indeed, across any row, the last four columns do sum to one.

TABLE 8.2
Genotypic fitnesses in a diploid model of selection

Chromosome from mother	Chromosome from father:			
	A_1B_1 (freq. x_1)	A_1B_2 (freq. x_2)	A_2B_1 (freq. x_3)	A_2B_2 (freq. x_4)
A_1B_1 (freq. x_1)	w_{11}	w_{12}	w_{13}	w_{14}
A_1B_2 (freq. x_2)	w_{21}	w_{22}	w_{23}	w_{24}
A_2B_1 (freq. x_3)	w_{31}	w_{32}	w_{33}	w_{34}
A_2B_2 (freq. x_4)	w_{41}	w_{42}	w_{43}	w_{44}

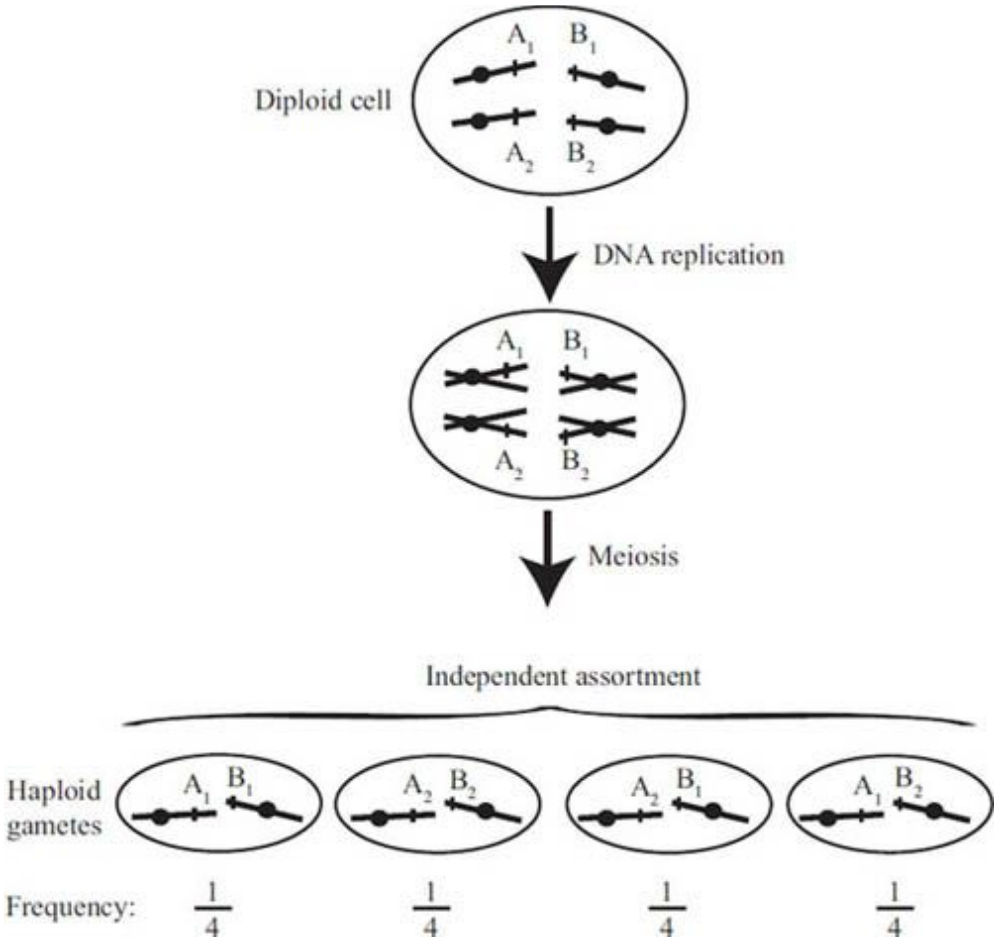

Figure 8.8: Independent assortment of chromosomes. Two chromosomes are shown in the nucleus of a diploid cell (each line represents a double helix of DNA, with a centromere located at the circle). During the production of gametes, the DNA replicates (producing X-shaped pairs of chromosomes) and then undergoes two meiotic divisions to produce four haploid cells. Because assortment of chromosomes is independent across different chromosomes, each of the four gamete types is equally likely.

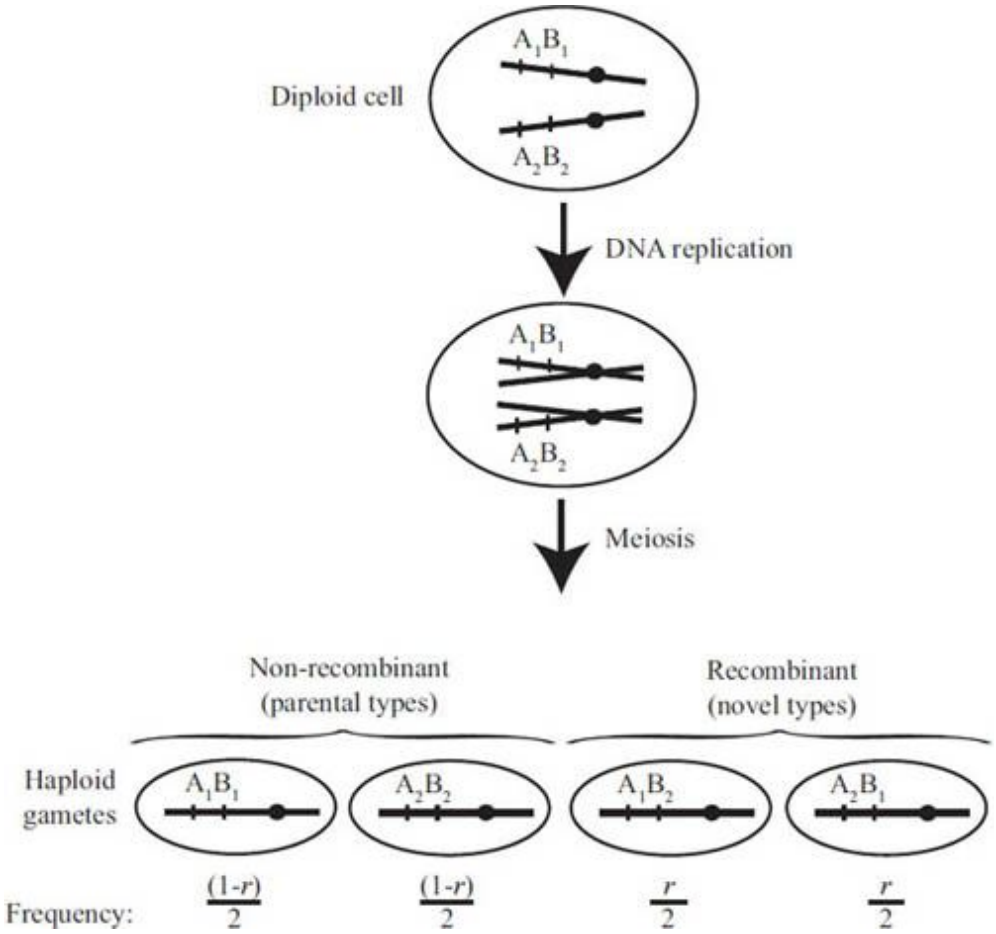

Figure 8.9: Recombination and meiosis. Two genes (**A** and **B**) on the same chromosome are illustrated. After DNA replication but before cell division, recombination can occur between the chromosomes, creating chromosomes with different combinations of the alleles than the parental chromosomes. The rate of recombination between two genes, r, measures the probability that a chromosome carries an **A** allele from one parental chromosome and a **B** allele from the other parental chromosome. When $r = 1/2$, all gamete types are equally like to be produced, whether recombinant or non-recombinant, as was the case for genes on different chromosomes (Figure 8.8).

Once gamete production is complete, the life cycle then begins again. To tabulate the frequency of the A_1B_1 gamete in the next generation (i.e., $x_1(t + 1)$) as a function of the frequency of the four types in the current generation, we multiply column 3 (which gives the postselection frequency of each parental genotype) by column 4 (which gives the probability that this parent produces an A_1B_1 gamete) and then sum these products over all rows. We get

$$x_1(t+1) = x_1x_1\frac{w_{11}}{\overline{w}} + \frac{1}{2}x_1x_2\frac{w_{12}}{\overline{w}} + \frac{1}{2}x_1x_3\frac{w_{13}}{\overline{w}} + \frac{1-r}{2}x_1x_4\frac{w_{14}}{\overline{w}} + \frac{1}{2}x_2x_1\frac{w_{21}}{\overline{w}} + \frac{r}{2}x_2x_3\frac{w_{23}}{\overline{w}} + \frac{1}{2}x_3x_1\frac{w_{31}}{\overline{w}} + \frac{r}{2}x_3x_2\frac{w_{32}}{\overline{w}} + \frac{1-r}{2}x_4x_1\frac{w_{41}}{\overline{w}}, \quad (8.39)$$

where the gamete frequencies on the right are measured at time t and where $\overline{W}$ is the population mean fitness and is equal to $\overline{w} = \sum_{i=1}^{4}\sum_{j=1}^{4} x_i(t)\, x_j(t)\, w_{ij}$.

To simplify this recursion, we bring together the terms that do not involve the recombination rate:

$$x_1\left(x_1\frac{w_{11}}{\overline{w}} + \frac{1}{2}x_2\frac{w_{12}}{\overline{w}} + \frac{1}{2}x_3\frac{w_{13}}{\overline{w}} + \frac{1}{2}x_4\frac{w_{14}}{\overline{w}} + \frac{1}{2}x_2\frac{w_{21}}{\overline{w}} + \frac{1}{2}x_3\frac{w_{31}}{\overline{w}} + \frac{1}{2}x_4\frac{w_{41}}{\overline{w}}\right), \quad (8.40)$$

which can be written more compactly as

$$x_1(t)\sum_{j=1}^{4} x_j(t)\left(\frac{w_{1j}+w_{j1}}{2\overline{w}}\right).$$

The remaining terms involve the recombination rate and are

$$\frac{-r}{2}x_1x_4\frac{w_{14}}{\overline{w}} + \frac{r}{2}x_2x_3\frac{w_{23}}{\overline{w}} + \frac{r}{2}x_3x_2\frac{w_{32}}{\overline{w}} + \frac{-r}{2}x_4x_1\frac{w_{41}}{\overline{w}}. \quad (8.41)$$

We can rewrite (8.41) as $-r\,D^*$, where

$$D^* = x_1x_4\left(\frac{w_{14}+w_{41}}{2\overline{w}}\right) - x_2x_3\left(\frac{w_{23}+w_{32}}{2\overline{w}}\right). \quad (8.42)$$

A nonrandom association between alleles at different genes is referred to as *linkage disequilibrium*. A random association is referred to as linkage equilibrium.

D^* is known as the *linkage disequilibrium*, measured after selection in (8.42). We use an asterisk to distinguish this measure from the linkage

disequilibrium before selection, $D = x_1x_4 - x_2x_3$. If you have not encountered the concept of linkage disequilibrium before, the main point to remember is that it describes the association between alleles carried at gene **A** and at gene **B**. When linkage disequilibrium is positive, a chromosome is more likely to carry a B_1 allele if it carries an A_1 allele (and more likely to carry a B_2 allele if it carries A_2) than you would predict based on random associations among the alleles within the population. In other words, there is a correlation between the alleles carried at the two genes when disequilibrium is present. Recall that we encountered this idea of a genetic correlation in the example of sexual selection in Chapter 7 (see also Figure 7.8).

Following the same procedure for the remaining three gamete types, we can write the four recursion equations in a nice compact form as

$$\begin{aligned}
x_1(t+1) &= x_1(t)\left(\sum_{j=1}^{4} x_j(t)\left(\frac{w_{1j}+w_{j1}}{2\overline{w}}\right)\right) - rD^*,\\
x_2(t+1) &= x_2(t)\left(\sum_{j=1}^{4} x_j(t)\left(\frac{w_{2j}+w_{j2}}{2\overline{w}}\right)\right) + rD^*,\\
x_3(t+1) &= x_3(t)\left(\sum_{j=1}^{4} x_j(t)\left(\frac{w_{3j}+w_{j3}}{2\overline{w}}\right)\right) + rD^*,\\
x_4(t+1) &= x_4(t)\left(\sum_{j=1}^{4} x_j(t)\left(\frac{w_{4j}+w_{j4}}{2\overline{w}}\right)\right) - rD^*,
\end{aligned} \tag{8.43}$$

Model (8.43) is a very complicated system of nonlinear equations. Even finding all of the equilibria of this model is an impossible task. Here we ask only one question with this model that illustrates a stability analysis without being overly complicated. Suppose that the two alleles A_1 and B_1 are fixed in the population (i.e., $x_1 = 1$). Under what conditions can the alternative alleles, A_2 and/or B_2 invade (Crow and Kimura 1965)? To answer this question we first need to calculate the stability matrix for (8.43) and evaluate it at the equilibrium $\hat{x}_1 = 1, \hat{x}_2 = 0, \hat{x}_3 = 0, \hat{x}_4 = 0$. This task is greatly aided by mathematical software such as Maple or *Mathematica* and results in the matrix

$$\mathbf{J} = \begin{pmatrix} 0 & -\frac{w_{12} + w_{21}}{2w_{11}} & -\frac{w_{13} + w_{31}}{2w_{11}} & -(1 - r)\frac{w_{14} + w_{41}}{2w_{11}} \\ 0 & \frac{w_{12} + w_{21}}{2w_{11}} & 0 & r\frac{w_{14} + w_{41}}{2w_{11}} \\ 0 & 0 & \frac{w_{13} + w_{31}}{2w_{11}} & r\frac{w_{14} + w_{41}}{2w_{11}} \\ 0 & 0 & 0 & (1 - r)\frac{w_{14} + w_{41}}{2w_{11}} \end{pmatrix} \quad (8.44)$$

The stability matrix (8.44) is upper triangular, and therefore its four eigenvalues are given by the diagonal elements (Rule P2.26) $\lambda_1 = 0$, $\lambda_2 = (w_{12} + w_{21})/(2w_{11})$, $\lambda_3 = (w_{13} + w_{31})/(2w_{11})$, and $\lambda_4 = (1 - r)(w_{14} + w_{41})/(2w_{11})$. For an equilibrium to be stable in a discrete-time model, we must show that the absolute values of all eigenvalues are less than one. None of the eigenvalues can be negative (because fitness cannot be negative), and therefore we need only check whether the eigenvalues are less than one.

The first eigenvalue $\lambda_1 = 0$ reveals that there is some direction in which the system always goes to zero. In this case, the genotype frequencies are constrained to satisfy a particular relationship: they must sum to one. Because of this constraint, there are effectively only three dimensions in which evolution occurs, even though there are four equations in (8.43).

For λ_2 to be less than one requires that $w_{11} > (w_{12} + w_{21})/2$. That is, the fitness of an A_1B_1/A_1B_1 individual must be larger than the average fitness of individuals that are heterozygous at the **B** gene and homozygous for the A_1 allele at the **A** gene. When this condition holds, the A_1B_2 chromosome is, on average, selected against. Similarly, for λ_3 to be less than one requires that $w_{11} > (w_{13} + w_{31})/2$. That is, the fitness of an A_1B_1/A_1B_1 individual must be larger than the average fitness of individuals that are heterozygous at the **A** gene and homozygous for the B_1 allele at the **B** gene. When this condition holds, the A_2B_1 chromosome is, on average, selected against. Finally, for λ_4 to be less than one requires that $w_{11} > (1 - r)(w_{14} + w_{41})/2$. This inequality is more interesting and involves the recombination rate. For stability of the equilibrium with A_1 and B_1 fixed, the fitness of an A_1B_1/A_1B_1 individual must be larger than the average fitness of individuals that are heterozygous at both genes times the probability that no recombination occurs.

How does the presence of two genes affect the stability of the equilibrium? The equilibrium is always unstable if the A_1B_2 or A_2B_1 chromosome has higher fitness than the resident ($\lambda_2 > 1$ or $\lambda_3 > 1$), just as we would predict from the one-gene model (section 5.3.2). Unlike with one-gene models, even if the A_1B_2 and A_2B_1 chromosomes are less fit than the residents ($\lambda_2 < 1$ and $\lambda_3 < 1$), the equilibrium can still be unstable if the fitness of A_1B_1/A_2B_2 individuals is higher than the fitness of A_1B_1/A_1B_1 individuals (and the recombination rate is low enough, $(1 - r)(w_{14} + w_{41})/2 > w_{11}$).

When the recombination rate is very low, the new A_2B_2 chromosome spreads if it has higher fitness than the resident. Recombination with the resident A_1B_1 breaks apart the good A_2B_2 combination of alleles, hindering the spread of the A_2B_2 chromosome, even if A_2B_2 causes its carriers to have higher fitness. Interestingly, this demonstrates that recombination can prevent the spread of gene combinations that are selectively favorable by breaking them apart.

If the A_2 and B_2 alleles decrease fitness on their own but increase fitness when combined, the two-gene model predicts that the new alleles can invade a population of A_1B_1/A_1B_1 individuals if $\lambda_4 > 1$ (Figure 8.10). This requires that the two genes interact to affect fitness, a phenomenon known as *epistasis*. As we shall discuss in greater depth in the next chapter, a one-gene model is generally not sufficient to predict the outcome of evolution whenever there are such fitness interactions among alleles at different genes.

8.4 Perturbation Techniques for Approximating Eigenvalues

Although we can often calculate and interpret the eigenvalues of a model, this is certainly not always true. When we can't, different techniques can be tried to determine the stability of an equilibrium. We have already mentioned one technique—using the Routh-Hurwitz conditions (Box 8.2). Another technique involves a graphical analysis of the characteristic polynomial (Sup. Mat. 8.1), which we use to analyze the discrete-time Lotka-Volterra

model of competition given by equation (3.14). Alternatively, it is often possible to approximate the leading eigenvalue using the perturbation techniques introduced in Chapter 5.

In Box 5.1, we introduced *perturbation analysis* for approximating the solution to an equation. In that context, we were interested in obtaining an expression that provided an approximation to an equilibrium. But perturbation analysis can be useful for finding approximate solutions to any equation, including the characteristic polynomial of a matrix. In this latter context, it thereby provides an approximation for the eigenvalues of a matrix.

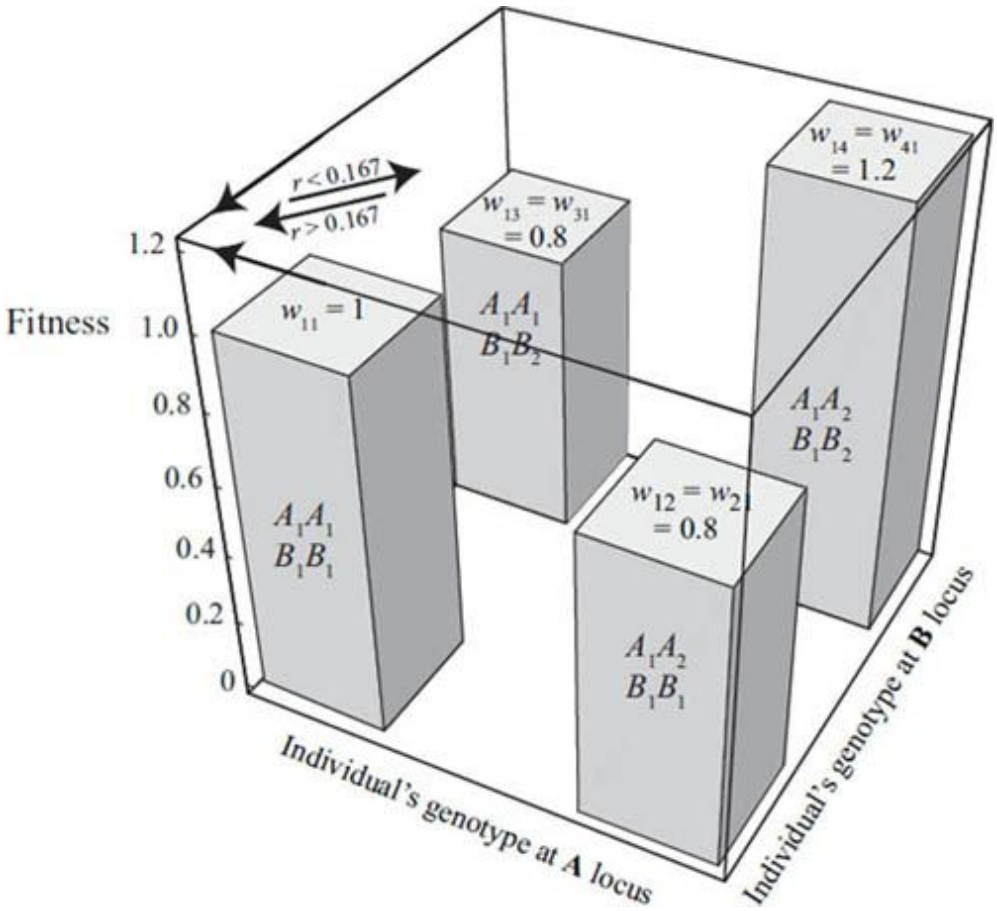

Figure 8.10: The bars give the fitnesses of the four relevant genotypes when alleles A_2 and B_2 are first introduced into a population of A_1B_1 chromosomes. The resident $A_1A_1B_1B_1$ individuals are more fit than $A_1A_1B_1B_2$ and $A_1A_2B_1B_1$ individuals, so that neither A_2 nor B_2 could spread on its own (arrows along the axes). Carrying both genetic changes, however, confers a 20% fitness advantage to $A_1A_2B_1B_2$ individuals. Even so, the A_2 and B_2 alleles spread only if they are sufficiently linked ($r <$ 0.167) such that $\lambda_4 = (1 - r)(w_{14} + w_{41})/(2w_{11}) > 1$ (diagonal arrow).

As in any perturbation analysis, we must first identify a parameter (or a set of parameters) that is small. For example, we might be willing to assume that the mutation rate or the effect of a new allele is small in an evolutionary model, or we might assume that the difference between two species is small in an ecological model. For the perturbation method to be useful in a stability analysis, it must be possible to determine the eigenvalues when the small parameter is set to zero. One situation in which perturbation analysis is particularly useful is when the leading eigenvalue falls on the boundary between stability and instability (i.e., equals one in a discrete-time model or zero in a continuous-time model). In this case, introducing the small parameter with some non-zero value might cause a critical shift in the leading eigenvalue. If the leading eigenvalue increases, then including the small parameter causes the equilibrium to become unstable. If the leading eigenvalue decreases, then including the small parameter causes the equilibrium to become stable.

An *approximate stability analysis* can be performed when a parameter is thought to be near a special value, using perturbation methods.

To begin, we assume that we have a characteristic polynomial that we want to solve for the leading eigenvalue λ. We then rewrite the characteristic polynomial as a function of the small parameter ζ; i.e., $f(\zeta) = 0$ (see Box 5.1). Next, we write the leading eigenvalue λ as a sum of terms involving powers of the small parameter ζ:

$$\lambda = \lambda_0 + \lambda_1 \zeta + \lambda_2 \zeta^2 + \lambda_3 \zeta^3 + \cdots . \tag{8.45}$$

You can think of each successive ζ^i term as providing a more refined estimate for the eigenvalue. Although there will be more than one eigenvalue in a multiple-variable model, we are particularly interested in the behavior of the leading eigenvalue. Thus, we set λ_0 equal to the value of the leading eigenvalue when the small parameter is set to zero. Next, we plug (8.45) into the characteristic polynomial $f(\zeta)$ and take the Taylor series of the characteristic polynomial with respect to ζ, around the point $\zeta = 0$ (Primer 1). Finally, we determine the values of the λ_i in (8.45) that cause each term in the Taylor series to equal zero. Such values ensure that the characteristic polynomial does indeed equal zero (see Box 5.1 for more details).

Knowing λ_1 is typically sufficient to determine the direction in which the eigenvalue changes with the addition of the small parameter. When λ_0 in (8.45) lies on the border between stability and instability (i.e., $\lambda_0 = 1$ in discrete time and $\lambda_0 = 0$ in continuous time), the sign of λ_1 in (8.45) determines whether the balance tips toward stability or instability for nonzero values of the small parameter ζ.

Example: The Evolution of Haploid and Diploid Organisms

We saw in Chapter 3 that an important genetic distinction between some organisms is whether they are haploid (i.e., they carry only one copy of each gene) or diploid (i.e., they carry two copies of each gene). For example, humans are primarily diploid because the majority of their life cycle is carried out in the diploid state, even though they have a haploid stage (sperm and eggs). In contrast, many fungi, algae, and unicellular organisms are primarily haploid. Why have some organisms evolved to become diploid while others have evolved to become haploid? Let us construct a model to gain some insight into this question.

Consider an organism that reproduces sexually and is capable of growth and development in either the haploid or diploid stage. Because the organism is sexual, it will necessarily pass through a haploid stage after meiosis and a diploid stage after the union of gametes (Figure 8.11). To allow the proportion of time spent in each state to evolve, we suppose that there is a gene that alters the life cycle. At this gene, allele C_1 causes meiosis to occur early in life (before natural selection has acted), resulting in a predominantly haploid life cycle (left pathway, Figure 8.11). In contrast, allele C_2 causes a delay in meiosis until after selection, resulting in a predominantly diploid life cycle (right pathway, Figure 8.11). To simplify matters, we assume that C_2 is dominant, so that C_1C_2 individuals are also diploid.

We also suppose that there are no intrinsic costs or benefits to being haploid or diploid, and therefore the frequency of allele C_2 would remain constant over time if it were the only gene considered in the model. There are certainly other genes in the genome under selection, however, and we thus include a second gene under selection with two alleles, where allele A is the most fit but mutates regularly to a less fit allele, a, at rate μ. Specifically, we assume that the fitnesses of individuals are

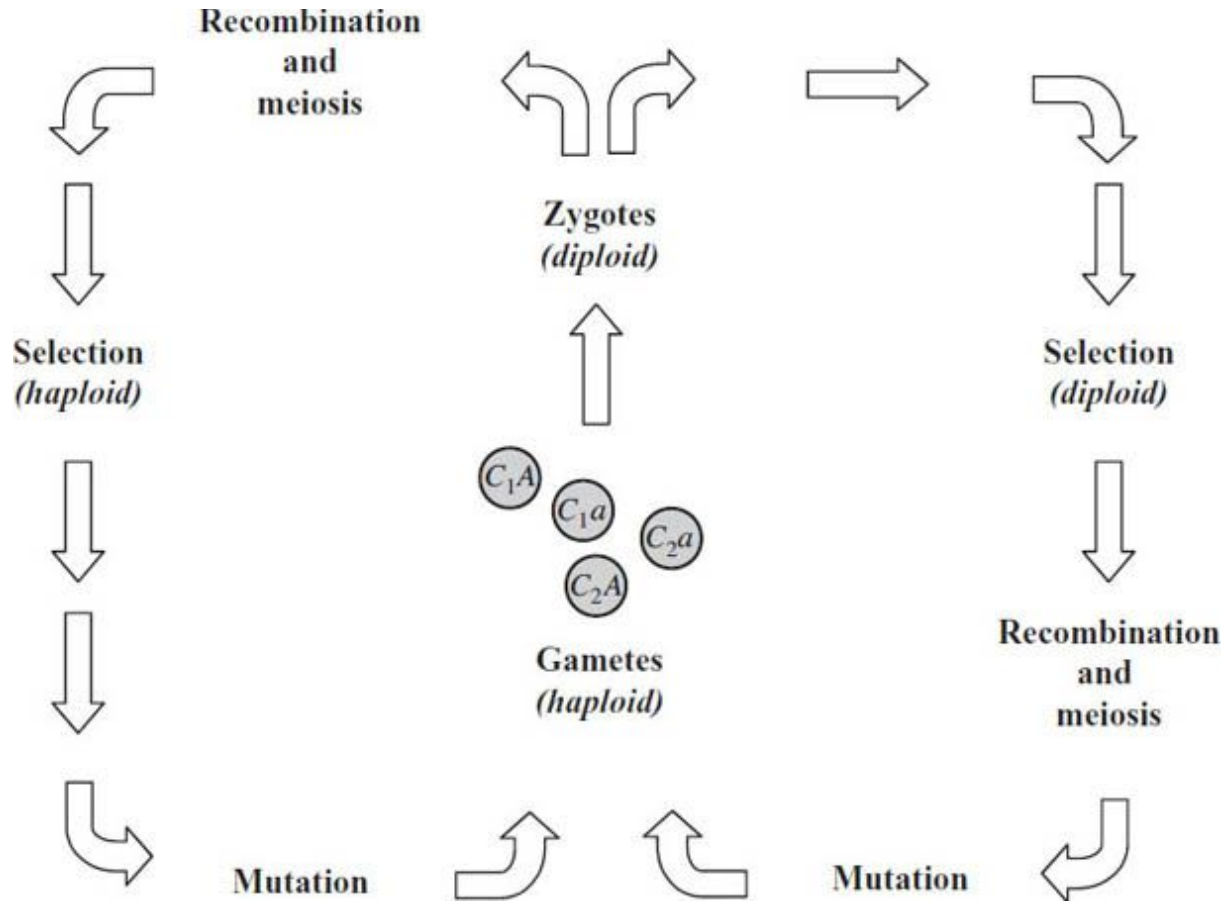

Figure 8.11: Evolution of haploidy and diploidy. A general life cycle is shown in which an organism either undergoes meiosis early and experiences selection as a haploid (left pathway) or undergoes meiosis late and experiences selection as a diploid (right pathway).

Haploid fitness		Diploid fitness		
A	a	AA	Aa	aa
1	$1-s$	1	$1-hs$	$1-s$

where h is the coefficient of dominance and s is the selective disadvantage. Both h and s are assumed positive so that the a allele is deleterious. We also assume that, at the end of the life cycle, all individuals produce haploid gametes that unite at random to begin the next generation (Figure 8.11).

We census the population at the gamete stage immediately after mutation, letting the frequency of C_1A, C_1a, C_2A, and C_2a gametes equal x_1, x_2, x_3, and x_4, respectively (Figure 8.11). Using a mating table like Table 8.1, it is possible to track the frequency of these genotypes as they unite, undergo

early meiosis (in C_1C_1 individuals), undergo selection, and then undergo late meiosis (in C_1C_2 and C_2C_2 individuals). The only difference from Table 8.1 is that C_1C_1 individuals experience selection only after meiosis. Working through such a table, the frequencies of the four chromosome types after meiosis but before mutation are

$$\begin{aligned}
x_1' &= (x_1^2 + x_1x_2 + x_1x_3 + (1-r)\,x_1x_4\,(1-hs) + r\,x_2x_3\,(1-hs))/\overline{W},\\
x_2' &= (x_1x_2\,(1-s) + r\,x_1x_4\,(1-hs) + x_2^2\,(1-s)\\
&\quad + (1-r)\,x_2x_3\,(1-hs) + x_2x_4\,(1-s))/\overline{W},\\
x_3' &= (x_1x_3 + x_1x_4\,r\,(1-hs) + x_2x_3\,(1-r)\,(1-hs) + x_3^2\\
&\quad + x_3x_4\,(1-hs))/\overline{W}\\
x_4' &= (x_1x_4(1-r)\,(1-hs) + x_2x_3\,(1-hs)\,r + x_2x_4\,(1-s)\\
&\quad + x_3x_4\,(1-hs) + x_4^2\,(1-s))/\overline{W},
\end{aligned} \tag{8.46a}$$

where r is the rate of recombination between the two genes, and $\overline{W}$ is the mean fitness of the population (the sum of the numerators on the right-hand side). Finally, we allow mutations to occur from A to a (mutations occurring in the reverse direction are assumed to be very rare and are ignored), which gives us the frequency of the four types of gametes in the next generation:

$$\begin{aligned}
x_1'' &= (1-\mu)x_1',\\
x_2'' &= \mu\,x_1' + x_2',\\
x_3'' &= (1-\mu)x_3',\\
x_4'' &= \mu\,x_3' + x_4'.
\end{aligned} \tag{8.46b}$$

If we want to understand how life cycles evolve, one approach is to assume that the population has a certain life cycle and then determine when a mutation altering the life cycle can invade. To this end, let's consider a haploid population, with allele C_1 fixed, into which we will introduce the C_2 allele, thereby generating diploid individuals.

First, we must determine the equilibrium reached by a haploid population when only allele C_1 is present ($x_3 = x_4 = 0$). In Problem 8.3, you are asked to find the equilibrium where selection is balanced by mutation.

This equilibrium occurs at $\hat{\chi}_1 = 1 - \mu/s$ and $\hat{\chi}_2 = \mu/s$, as we found in the one-gene model (Problem 5.4).

Next, we explore what happens after C_2 arises at some small frequency, causing its carriers to remain diploid throughout selection. The allele will ultimately decrease in frequency if the equilibrium $\hat{\chi}_1 = 1 - \mu/s$, $\hat{\chi}_2 = \mu/s$, $\hat{\chi}_3 = 0$, $\hat{\chi}_4 = 0$ is locally stable, and the population will remain haploid. Therefore, we can determine when we expect the allele to die out by performing a local stability analysis of this equilibrium. The stability matrix for this model is

$$\mathbf{J} = \left(\begin{array}{cccc} \frac{\partial x_1''}{\partial x_1} & \frac{\partial x_1''}{\partial x_2} & \frac{\partial x_1''}{\partial x_3} & \frac{\partial x_1''}{\partial x_4} \\ \frac{\partial x_2''}{\partial x_1} & \frac{\partial x_2''}{\partial x_2} & \frac{\partial x_2''}{\partial x_3} & \frac{\partial x_2''}{\partial x_4} \\ \frac{\partial x_3''}{\partial x_1} & \frac{\partial x_3''}{\partial x_2} & \frac{\partial x_3''}{\partial x_3} & \frac{\partial x_3''}{\partial x_4} \\ \frac{\partial x_4''}{\partial x_1} & \frac{\partial x_4''}{\partial x_2} & \frac{\partial x_4''}{\partial x_3} & \frac{\partial x_4''}{\partial x_4} \end{array} \right) \Bigg|_{\substack{x_1 = 1 - \mu/s,\ x_2 = \mu/s, \\ x_3 = 0, x_4 = 0}} \tag{8.47}$$

where the x_i'' are given by equations (8.46). Analyzing this matrix is aided by the fact that the 2×2 submatrix on the bottom left is full of zeros:

$$\left(\begin{array}{cc} \frac{\partial x_3''}{\partial x_1} & \frac{\partial x_3''}{\partial x_2} \\ \frac{\partial x_4''}{\partial x_1} & \frac{\partial x_4''}{\partial x_2} \end{array} \right) \Bigg|_{\substack{x_1 = 1 - \mu/s,\ x_2 = \mu/s, \\ x_3 = 0, x_4 = 0}} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus, matrix (8.47) has a block-triangular form (see Primer 2), and its four eigenvalues are given by the eigenvalues of the two submatrices along the diagonal (Rule P2.27), which we label **A** and **B**:

$$\mathbf{A} = \left(\begin{array}{cc} \dfrac{\partial x_1''}{\partial x_1} & \dfrac{\partial x_1''}{\partial x_2} \\ \dfrac{\partial x_2''}{\partial x_1} & \dfrac{\partial x_2''}{\partial x_2} \end{array} \right) \Bigg|_{\substack{x_1 = 1-\mu/s,\, x_2 = \mu/s, \\ x_3 = 0, x_4 = 0}} = \left(\begin{array}{cc} \dfrac{\mu(1-s)}{(1-\mu)s} & -\dfrac{(s-\mu)(1-s)}{(1-\mu)s} \\ -\dfrac{\mu(1-s)}{(1-\mu)s} & \dfrac{(s-\mu)(1-s)}{(1-\mu)s} \end{array} \right) \tag{8.48a}$$

$$\mathbf{B} = \left(\begin{array}{cc} \dfrac{\partial x_3''}{\partial x_3} & \dfrac{\partial x_3''}{\partial x_4} \\ \dfrac{\partial x_4''}{\partial x_3} & \dfrac{\partial x_4''}{\partial x_4} \end{array} \right) \Bigg|_{\substack{x_1 = 1-\mu/s,\, x_2 = \mu/s, \\ x_3 = 0, x_4 = 0}}$$

$$= \left(\begin{array}{cc} 1 - \dfrac{\mu(r + hs - rhs)}{s} & \dfrac{(s-\mu)(1-hs)\,r}{s} \\ \dfrac{\mu(r + s - rhs) - \mu^2(r + hs - rhs)}{(1-\mu)s} & 1 - \dfrac{(s-\mu)(hs + (1-\mu)(1-hs)\,r)}{(1-\mu)s} \end{array} \right). \tag{8.48b}$$

From the partial derivatives contained in the submatrix **A**, we can tell that this submatrix describes the sensitivity of the recursions for x_1 and x_2 to displacements in x_1 and x_2. The eigenvalues of this submatrix thus describe the stability of the equilibrium in the absence of the new mutant C_2 allele. Using Rule P2.20 and factoring, these eigenvalues are given by $\lambda = 0$ and $\lambda = (1 - s)/(1 - \mu)$. The zero eigenvalue indicates that the model has only three effective dimensions, and again this occurs because all four gamete frequencies must sum to one. The eigenvalue of $(1 - s)/(1 - \mu)$ is positive and less than one as long as the mutation rate is small relative to selection ($\mu < s$), which is both a reasonable assumption and necessary for $\hat{x}_1 = 1 - \mu/s$ to be a valid equilibrium frequency. Thus, when the C_2 allele is absent, the mutation-selection balance equilibrium is stable when it exists.

The real question of interest is whether this equilibrium is stable if we perturb it by introducing the C_2 allele (i.e., if we have x_3 and/or x_4 not equal to zero). As seen by the partial derivatives contained in submatrix **B**, the eigenvalues of **B** address this question. These eigenvalues can be calculated, but they are ugly and difficult to interpret. To obtain interpretable results, we

use the perturbation method under the assumption that the mutation rate is very small, $\mu = \zeta$.

First, we must find λ_0, the leading term in the eigenvalue (8.45), by setting the mutation rate to zero in submatrix **B**. The submatrix is then triangular and has eigenvalues equal to the diagonal elements; $\lambda_0 = 1$ and $\lambda_0 = (1 - hs)(1 - r)$. Because we have assumed that all of the parameters are positive, $\lambda_0 = 1$ is the leading eigenvalue (i.e., the one with the largest absolute value).

Because the leading term $\lambda_0 = 1$ falls on the border between stability and instability, we must seek out the next-order term λ_1 in (8.45). Plugging (8.45) in for λ in the characteristic polynomial of **B**, taking the Taylor series with respect to ζ, and keeping only terms to first order in ζ, we get an equation that λ_1 must satisfy:

$$-\mu\,(r\,(1 - 2h)(1 - hs) - h^2 s - \lambda_1\,(r + s - rhs)) = 0, \tag{8.49}$$

which we have written in terms of the original parameter μ (see Problem 8.9). Solving (8.49) for λ_1 and plugging λ_0 and λ_1 into (8.45), we get an approximation for the leading eigenvalue:

$$\lambda = 1 + \frac{r\,(1 - 2h)(1 - hs) - h^2 s}{r + s - rhs}\,\mu + O(\mu^2), \tag{8.50}$$

which is extremely close to the exact numerical value (Figure 8.12). The denominator of (8.50) is always positive, but the numerator is more difficult to interpret. If $h > 1/2$ then all terms in the numerator are negative and λ is less than one, indicating that diploids cannot invade a haploid population. If $h < 1/2$, then the two terms in the numerator are of opposite sign, and diploidy can invade ($\lambda > 1$) as long as the first term is larger. This requires that $r > h^2 s/((1 - 2h)(1 - hs))$.

What do these results mean biologically? Intuitively, one might expect diploidy to be favored because mutant alleles can be "masked" by the good copy of the allele in heterozygotes. This does provide diploidy with an advantage, but this advantage is counterbalanced by the fact that a diploid individual has two chances of carrying a deleterious mutation. Thus, only when masking is strong enough ($h < 1/2$) would the average fitness of

diploids be better than haploids, assuming that they have the same frequency of deleterious alleles.

The final twist to our result, however, is that diploidy spreads only if recombination rates are high enough, even when mutations are better masked in diploid individuals ($h < 1/2$). Because of masking among diploid ancestors, mutant alleles are more likely to survive and persist among the descendants carrying the diploid allele, C_2. Consequently, chromosomes with the diploid allele C_2 are more loaded with deleterious mutations, while chromosomes with the haploid allele C_1 are more effectively purged of deleterious mutations. The tighter the recombination rate between the ploidy and selected genes, the greater the difference in mutant allele frequency expected between C_1-bearing and C_2-bearing chromosomes, and the less likely diploids are to invade a population. While diploidy protects the individual from its burden of deleterious mutations, it does so at the expense of future generations, which are more likely to inherit deleterious mutations. Thus, diploidy is favored only when there is enough genetic mixing among the chromosomes in a population (see also Otto 1994; Otto and Goldstein 1992; Otto and Marks 1996; Perrot et al. 1991).

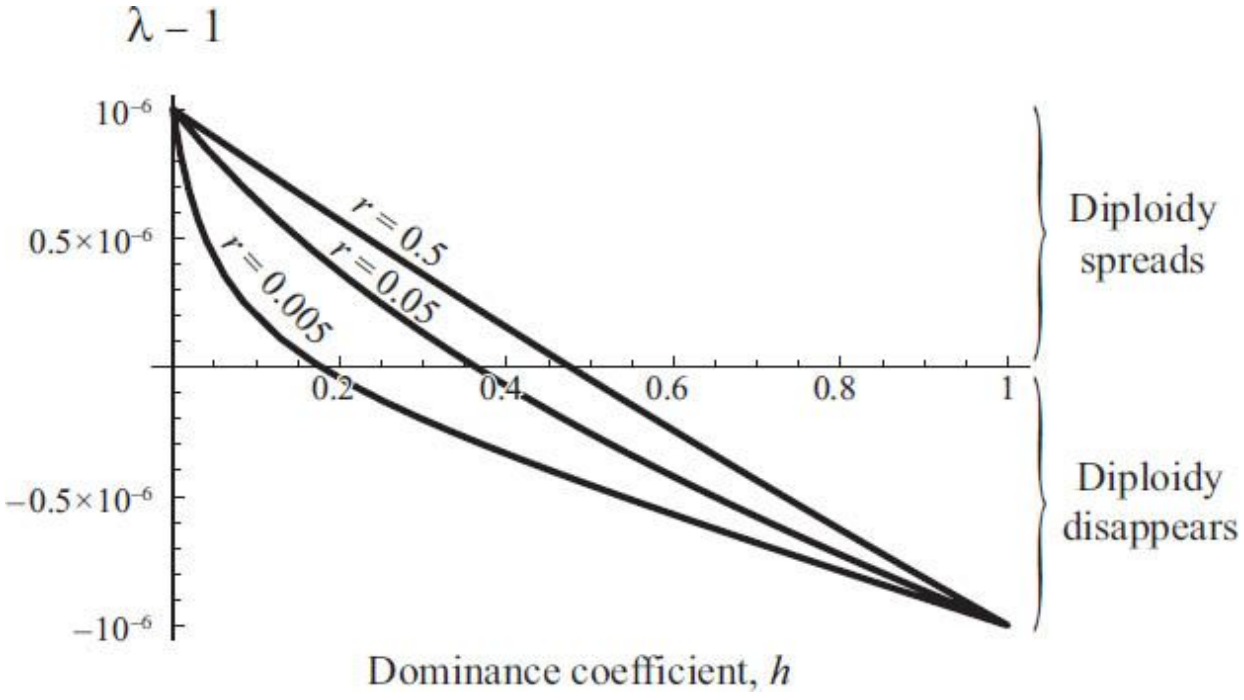

Figure 8.12: Evolution of diploidy. The leading eigenvalue minus one ($\lambda - 1$) is shown as a function of the dominance coefficient. A mutation causing diploid life cycles spreads within a haploid population only when the curve lies above the horizontal axis. This requires that h is small enough (λ

– 1 > 0); the larger the recombination rate between the genes, the broader the range of dominance conditions under which diploidy evolves. For the parameters chosen ($s = 0.1$, $\mu = 10^{-6}$), the curves for the leading eigenvalue based on the approximation (8.50) or a numerical evaluation of the eigenvalues of the matrix (8.48b) cannot be distinguished.

8.5 Concluding Message

In this chapter we have presented the general techniques for finding equilibria and determining their stability properties in non-linear models with multiple variables. This concludes our tour of the main techniques for analyzing equilibria of dynamical models. To summarize, we write a nonlinear model involving n dynamic variables $x_1, \ldots, x_n$, using n differential equations of the form $dx_i/dt = f_i(x_1, x_2, \ldots, x_n)$ (continuous time) or n recursion equations of the form $x_i(t + 1) = f_i(x_1(t), x_2(t), \ldots, x_n(t))$ (discrete time).

At any equilibrium of such a model, none of the variables changes over time. In continuous time, the equilibria are found by determining the values of the variables that simultaneously cause $dx_1/dt = 0$, $dx_2/dt = 0, \ldots, dx_n/dt = 0$. In discrete time, the equilibria are found by determining the values of the variables that simultaneously cause $f_1(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = \hat{x}_1$, $f_2(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = \hat{x}_2, \ldots,$ $f_n(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = \hat{x}_n$. Nonlinear models can have multiple equilibria, each of which has its own stability properties.

For each equilibrium, we can determine whether it is stable or unstable by performing a local stability analysis. A local stability analysis or linearization approximates the nonlinear equations with linear equations that are accurate near the equilibrium and that can be analyzed using the techniques of Chapter 7. First, we find the Jacobian matrix (Definition 8.2), which consists of the partial derivatives of each function with respect to each variable. Next, we determine the eigenvalues of this matrix evaluated at an equilibrium. For a continuous-time model, the equilibrium is stable as long as the real parts of all eigenvalues are negative (Recipe 8.2). For a discrete-time model, the equilibrium is stable as long as the absolute values of all eigenvalues are less than one (Recipe 8.4).

For some models with multiple variables, it is possible to go beyond analyzing the stability of equilibrium points. In the next chapter, we will see how to obtain general solutions for some such models. As with the one-variable models considered in Chapter 6, the general solution to multiple-variable models provides us with complete information about the model's behavior.

Problems

Problem 8.1: For each pair of hypothetical eigenvalues from a stability analysis, specify whether the equilibrium is stable or unstable; also specify when cycles are expected around the equilibrium because the eigenvalues are complex (you can assume that p is a proportion, and therefore lies between 0 and 1). (a) $r_1 = 0.5$ and $r_2 = -2/3$ in a continuous-time model. (b) $\lambda_1 = 0.5$ and $\lambda_2 = -2/3$ in a discrete-time model. (c) $r_1 = (1 + \sqrt{1 + p})/2$ and $r_2 = (1 - \sqrt{1 + p})/2$ in a continuous-time model. (d) $\lambda_1 = (1 + \sqrt{1 + p})/2$ and $\lambda_2 = (1 - \sqrt{1 + p})/2$ in a discrete-time model. (e) $r_1 = (1 + \sqrt{-p})/2$ and $r_2 = (1 - \sqrt{-p})/2$ in a continuous-time model. (f) $\lambda_1 = (1 + \sqrt{-p})/2$ and $\lambda_2 = (1 - \sqrt{-p})/2$ in a discrete-time model.

Problem 8.2: In the text, we explored a diploid version of density-dependent natural selection (8.36). A haploid version of this model is

$$N(t = 1) = \overline{W}(N(t), p(t))\, N(t),$$

$$p(t + 1) = \frac{W_A(N(t))}{\overline{W}(N(t), p(t))}\, p(t),$$

where $\overline{W}(N(t), p(t)) = p(t)\, W_A(N(t)) + (1 - p(t))\, W_a(N(t))$, $W_A(N) = (1 + r)e^{-\alpha_A N}$, and $W_a(N) = (1 + r)e^{-\alpha_a N}$. Identify all of the equilibria of this model.

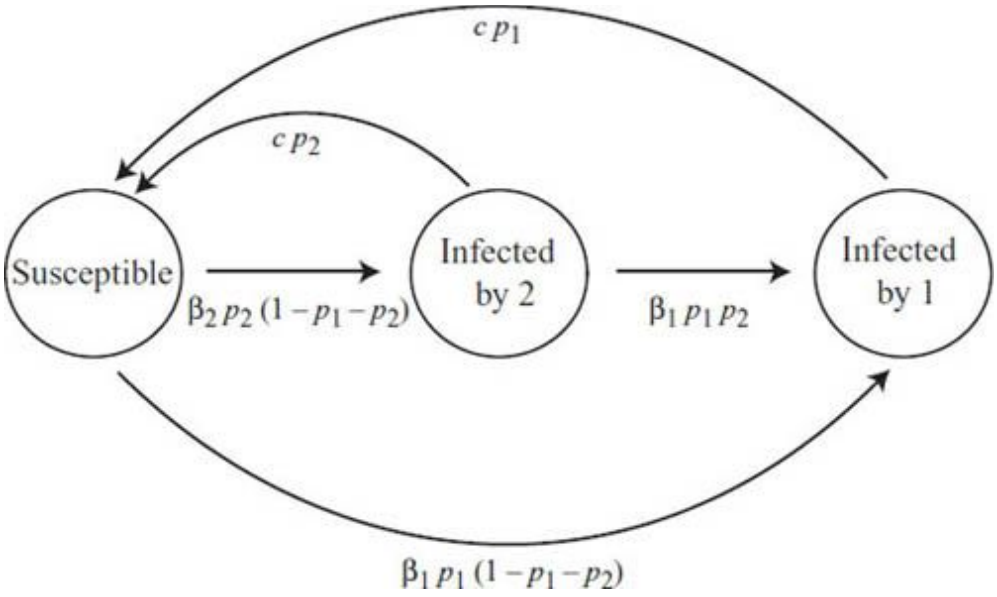

Figure 8.13: A flow diagram of infection and clearance.

Problem 8.3: Find the two equilibria of (8.46) under the assumption that the allele C_2 is absent. Say why your answer does not depend on the recombination rate r or the dominance coefficient h.

Problem 8.4: Here we construct an epidemiological model in which there are two parasite types (labeled 1 and 2). To simplify matters, we assume that the total host population is constant in size, and we track the frequency of the host population that is infected by each type of parasite in continuous time. We further assume that a host can be infected by at most one type of parasite and that there is an asymmetry such that parasite 1 can infect and immediately take over a host infected with parsite 2 but not vice versa (this is sometimes referred to as superinfection: Levin and Pimentel 1981; Nowak and May 1994). At any given time, some fraction of the hosts will be susceptible ($1 - p_1 - p_2$), some fraction will be infected by parasite 1 (p_1), and some fraction will be infected by parasite 2 (p_2) (Figure 8.13).

Assuming a mass-action infection process, susceptible hosts are infected with parasite 2 at rate $\beta_2 p_2(1 - p_1 - p_2)$ (i.e., proportional to the number of hosts currently infected with parasite 2 and to the number of susceptible hosts). Hosts infected with parasite 2 clear the infection at rate $c\,p_2$ or can be superinfected by parasite 1 at rate $\beta_1 p_1 p_2$. Parasite 1 infects both susceptible hosts and hosts infected by parasite 2 at a total rate of $\beta_1 p_1(1 - p_1 - p_2) + \beta_1 p_1 p_2 = \beta_1 p_1(1 - p_1)$. Hosts infected with parasite 1 also clear the infection at rate $c\,p_1$. These assumptions generate the model

$$\frac{dp_1}{dt} = \beta_1 p_1 (1 - p_1) - c\,p_1,$$

$$\frac{dp_2}{dt} = \beta_2 p_2 (1 - p_1 - p_2) - \beta_1 p_1 p_2 - c\,p_2.$$

In the following, the quantities β_1/c and β_2/c will arise frequently. Whenever possible, simplify the notation by defining $R_1 = \beta_1/c$ and $R_2 = \beta_2/c$, which represent the infectivities of the parasites relative to their clearance rates. Intuitively we might expect to find four equilibria in this model: (i) one with both parasites absent, (ii) one with parasite 1 present and parasite 2 absent, (iii) one with parasite 2 present and parasite 1 absent, and (iv) one with both parasites present.

(a) Find the four equilibria and express them in terms of the quantities, R_1 and R_2. (b) For equilibria (i)–(iv), determine the conditions under which they are biologically feasible (i.e., represent fractions between 0 and 1), expressed in terms of the quantities R_1 and R_2. (c) Calculate the general Jacobian matrix for the differential equations, without plugging in a particular equilibrium. (d) Suppose $R_1 < 1$ and $R_2 < 1$. Which equilibria are biologically feasible? For each feasible equilibrium, determine if it is locally stable or unstable. Provide a biological interpretation of the assumption that $R_1 < 1$ and $R_2 < 1$ to explain your findings. (e) Suppose $R_1 > 1$ and $R_2 < 1$. Which equilibria are biologically feasible? For each feasible equilibrium, determine if it is locally stable or unstable. Provide a biological interpretation of the assumption that $R_1 > 1$ and $R_2 < 1$ to explain your findings. (f) Suppose $R_1 > 1$ and $R_2 > 1$ and $R_2 < R_1^2$. Which equilbria are biologically feasible? For each feasible equilibrium, determine if it is locally stable or unstable. Provide a biological interpretation of the assumption that $R_1 > 1$ and $R_2 > 1$ and $R_2 < R_1^2$ to explain your findings. (g) Suppose $R_1 > 1$ and $R_2 > 1$ and $R_2 > R_1^2$. Which equilibria are biologically feasible? For each feasible equilibrium, determine if it is locally stable or unstable. Provide a biological interpretation of the assumption that $R_1 > 1$ and $R_2 > 1$ and $R_2 > R_1^2$ to explain your findings.

Problem 8.5: Consider the following predator-prey model:

$$\frac{dR}{dt} = rR\left(1 - \frac{R}{K}\right) - P\frac{cR}{a + R}$$

$$\frac{dP}{dt} = \frac{\varepsilon cR}{a + R}P - dP.$$

R and P represent the number of prey and predators respectively, r is the per capita growth rate of the prey when their numbers are small and the predator is absent, K is the carrying capacity of the prey in the absence of the predator, a and c are parameters governing the shape of the functional response, ε is the conversion efficiency of consumed prey into new predators, and d is the per capita mortality rate of the predator. All parameters are positive.

(a) Provide a biological interpretation for the above model. In particular, discuss the nature of the prey dynamics in the absence of the predator and the predator dynamics in the absence of the prey. Also, discuss whether the consumption of prey by any given predator individual is affected by prey and predator population sizes. If it is affected by these population sizes, discuss how and provide a reason why this might be the case in reality. (b) Find all equilibria of the model. You should find three of them, only one of which has both species present. Label the equilibrium with neither species present as (i), with one species as (ii), and with both species as (iii). Give the conditions on the parameters under which each equilibrium is biologically feasible. (c) Calculate the Jacobian matrix for the above model but do not evaluate it at any particular

equilibrium. (d) Determine the conditions under which equilibrium (i) is *unstable* and provide a biological interpretation of the result. (e) Determine the conditions under which equilibrium (ii) is *unstable* and provide a biological interpretation of the result. (f) Show that equilibria (i) and (ii) are both unstable when equilibrium (iii) is feasible. (g) Given that equilibrium (iii) is biologically feasible, determine the conditions under which it is *stable*. Use the fact that the equilibrium is stable if the determinant of the 2×2 Jacobian matrix is positive and if its trace is negative. (h) As the carrying capacity of the prey, K, increases, there comes a point at which the equilibrium (iii) loses its stability. Determine this critical value of K in terms of the other parameters. (For K larger than this critical value, predator-prey cycles occur.)

Problem 8.6: Consider the following discrete-time model for the population dynamics of a host species and a parasitoid (a parasitoid is an insect whose larvae develop inside other insects):

$$N(t+1) = R\,N(t)\exp(-a\,P(t)),$$

$$P(t+1) = a\,N(t)\,(1-\exp(-a\,P(t))).$$

Here, N is the host population size and P is the parasitoid population size, and a and R are positive parameters. (a) Find all equilibria of model (2). (b) Calculate the Jacobian matrix and evaluate it at each equilibrium obtained in (a). (c) Determine the conditions under which the equilibrium where both species are absent is unstable. (d) Provide a biological interpretation for your answer to (c).

Problem 8.7: Here, we analyze the dynamics of a source-sink model of population dynamics. Consider two populations of a single species, one in a good environment (the "source" population 1) and one in a poor environment (the "sink" population 2). The number of individuals in the two populations is n_1 and n_2, respectively. In the good environment, the population is able to grow logistically with a carrying capacity K_1 and an intrinsic rate of growth r_1. Each generation, a proportion m of the adults migrate from the good to the poor environment. In the poor environment, each individual has R offspring. R is less than one and the population is unable to replace itself. Thus, the sink population is maintained by the constant input of migrants every generation from the source population. Equations describing this model are

$$n_1(t+1) = (1-m)\,n_1(t) + n_1(t)\,r_1\left(1 - \frac{n_1(t)}{K}\right),$$

$$n_2(t+1) = R\,n_2(t) + m\,n_1(t).$$

(a) What are the equilibrium population sizes for the two populations in this model? (b) Under what conditions is the source-sink metapopulation stable?

Problem 8.8: Here we analyze the Volterra model of predator-prey dynamics in discrete time. We let $P(t)$ equal the number of predators at time t and $H(t)$ equal the number of prey. The parameters of the model are: the per capita growth of the prey in the absence of the predator (r), the per capita probability that a predator contacts and kills a prey (β), the per capita growth of the predator following the consumption of prey (c), and the death rate of predators (δ). With these definitions, the discrete time predator-prey model is

$$H(t+1) = H(t) + r\,H(t) - \beta\,H(t)\,P(t),$$

$$P(t+1) = P(t) + c\,H(t)\,P(t) - \delta\,P(t).$$

(a) Determine the two equilibria of these equations. Double check that the numbers of *both* predators and prey do not change over time when started at an equilibrium. (b) Determine the local stability matrix that approximates these equations near the equilibrium with both species absent. Repeat, finding the local stability matrix near the equilibrium with both species present. (c) Find the eigenvalues for the two matrices in (b). (d) From these eigenvalues, determine whether each equilibrium is stable or unstable, assuming that every parameter is positive. [Recall that an equilibrium in a discrete-time model is stable if all of its eigenvalues are less than one in magnitude, where the magnitude of a complex eigenvalue $A + B\,i$ is $\sqrt{A^2 + B^2}$.]

Problem 8.9: Here we derive equation (8.50) of the text, which gives the eigenvalue near one in the model of haploid-diploid evolution using the stability matrix (8.48b),

$$\mathbf{B} = \begin{pmatrix} 1 - \dfrac{\mu\,(r + hs - rhs)}{s} & \dfrac{(s-\mu)(1-hs)\,r}{s} \\ \dfrac{\mu\,(r + s - rhs) - \mu^2\,(r + hs - rhs)}{(1-\mu)\,s} & 1 - \dfrac{(s-\mu)(hs + (1-\mu)(1-hs)\,r)}{(1-\mu)\,s} \end{pmatrix}.$$

(a) Find the characteristic polynomial $\mathrm{Det}(\mathbf{B} - \lambda\mathbf{I}) = 0$. (b) Replace μ with ζ, replace λ with $\lambda_0 + \lambda_1\,\zeta + \lambda_2\,\zeta^2 + \lambda_3\,\zeta^3$, and replace λ_0 with 1 (based on the case where $\mu = 0$). (c) Perform a Taylor series approximation of this characteristic polynomial with respect to ζ around the point $\zeta = 0$ to linear order in ζ (Primer 1). (d) Undo the transformation by replacing ζ with μ, and show that your result is equivalent to (8.50). Note that your result does not depend on λ_2 or λ_3, nor would it depend on any higher-order term in (8.45). [Use a mathematical software package if available.]

Problem 8.10: The Lotka-Volterra model of competition between two species in discrete-time can be described by the recursion equations

$$n_1(t+1) = n_1(t) + r_1\,n_1(t)\left(1 - \frac{n_1(t) + \alpha_{12}\,n_2(t)}{K_1}\right),$$

$$n_2(t+1) = n_2(t) + r_2\,n_2(t)\left(1 - \frac{n_2(t) + \alpha_{21}\,n_1(t)}{K_2}\right).$$

(a) Determine the four equilibria and confirm that $\hat{n}_1 = (K_1 - \alpha_{12}\,K_2)/(1 - \alpha_{12}\,\alpha_{21})$ and $\hat{n}_2 = (K_2 - \alpha_{21}\,K_1)/(1 - \alpha_{12}\,\alpha_{21})$ is the only equilibrium with both species present. (b) Determine the local stability matrices for the three equilibria in which both species are *not* maintained (stability of the equilibrium with both species present is analyzed in Supplementary Material 8.1). (c) Use Rule P2.26 to find the eigenvalues for these three equilibria. (d) [Challenging] Interpret the stability conditions for these three equilibria using the fact that, near an equilibrium,

$$1 + r_i\left(1 - \frac{\hat{n}_i + \alpha_{ij}\,\hat{n}_j}{K_i}\right)$$

is the number of offspring per parent of species i, which can never be negative.

Problem 8.11: The Lotka-Volterra model of competition in continuous time can be described by the differential equations

$$\frac{dn_1}{dt} = r_1 n_1(t)\left(1 - \frac{n_1(t) + \alpha_{12} n_2(t)}{K_1}\right),$$

$$\frac{dn_2}{dt} = r_2 n_2(t)\left(1 - \frac{n_2(t) + \alpha_{21} n_1(t)}{K_2}\right).$$

(a) Show that $\hat{n}_1 = (K_1 - \alpha_{12} K_2)/(1 - \alpha_{12}\alpha_{21})$ and $\hat{n}_2 = (K_2 - \alpha_{21} K_1)/(1 - \alpha_{12}\alpha_{21})$ is an equilibrium of the continuous-time model as well as the discrete-time model (Problem 8.10) and that it is the only equilibrium with both species present. (b) Determine the local stability matrix for this equilibrium. How does this matrix differ from the discrete-time equivalent (S8.1.5) of Supplemental Material 8.1? (c) Use the Routh-Hurwitz conditions to determine when the equilibrium with both species is stable. Contrast these conditions to the discrete-time model analyzed in Supplemental Material 8.1.

Problem 8.12: In Box 2.5, we described the model of HIV infection developed by Blower et al. (2000) to predict changes in HIV incidence in the San Francisco community of gay males. Before treatment begins, the model describing changes to the number of uninfected (X) and infected (Y) individuals simplifies to

$$\frac{dX}{dt} = \pi - c\beta\frac{Y}{Y+X}X - \mu X,$$

$$\frac{dY}{dt} = c\beta\frac{Y}{Y+X}X - (\mu + \nu)Y,$$

where π describes the immigration rate of males, c is the number of partners per male, β is the rate of infection per sexual contact with an infected male, μ is the normal mortality or emigration rate, and ν is the additional mortality rate suffered by infected individuals. (a) Determine the equilibria of this model and describe the conditions under which each equilibrium is biologically feasible. (b) Use a local stability analysis to determine the conditions under which HIV will spread when rare within a population. (c) Use the trace and determinant conditions (Rule P2.25) to determine whether the equilibrium with both infected and uninfected individuals present is stable. Would you expect to observe cycling around this equilibrium?

Problem 8.13: Show that the general Routh-Hurwitz conditions (8.2.12) of Box 8.2 are equivalent to the conditions for a 3×3 matrix (8.2.6) of Box 8.2.

Problem 8.14: In this problem, we analyze a nonlinear crab-anemone model that allows the number of juvenile crabs to depend on the current population size and incorporates competition among these juveniles for available sea anemones. The recursions for this model are

$$B(t+1) = B(t) - \mu_B B(t) + (B(t) + R(t) + G(t))\, b\, p_B\, (1 - B(t)/N_B),$$
$$R(t+1) = R(t) - \mu_R R(t) + (B(t) + R(t) + G(t))\, b\, p_R\, (1 - R(t)/N_R),$$

$$G(t+1) = G(t) - \mu_G\, G(t) + (B(t) + R(t) + G(t))\, b\, P_G\, (1 - G(t)/N_G),$$

(see Problem 7.8). (a) Write down the equilibrium conditions for these recursions and solve for the mortality rates (do *not* solve them for the equilibrium). That is, determine what μ_B, μ_R, and μ_G are as functions of the other parameters and $\hat{B}$, $\hat{R}$, and $\hat{G}$. (b) Using the data of Baeza and Stotz (2001), solve for the mortality rates relative to the per capita birth rate b. Set $N_B = 69$, $N_R = 115$, $N_G = 179$, $p_B \approx 0.14$, $p_G \approx p_R \approx 0.43$, and assume that the observed numbers of occupied anemones in Figure 7.10a are near the equilibrium so that $\hat{B} \approx 1$, $\hat{R} \approx 33$, $\hat{G} \approx 38$. (c) Compare your predictions from (b) to the predictions $\mu_B \approx 0.14b$, $\mu_R \approx 0.0092b$, and $\mu_G \approx 0.0090b$ obtained from model (iii) in Section 7.4. Remember that b represents the per capita birth rate in this model but the total birth rate for the whole population in (7.33a).

Problem 8.15: [Challenging]. Consider model (8.36) but where the fitness expressions $W_{ij}(N)$ are left unspecified. Here you will again focus on the equilibrium with only the resident allele A. (a) Set $\hat{p} = 1$ in equation (8.37) and show that the equilibrium population size must then satisfy the equation $W_{AA}(\hat{N}) = 1$. Provide a biological interpretation of this result. (b) Even though we do not have an explicit expression for $\hat{N}$, we can still use the conditions $\hat{p} = 1$ and $W_{AA}(\hat{N}) = 1$ to simplify the Jacobian matrix at the equilibrium with allele A fixed. Show that this matrix can be written as

$$\mathbf{J} = \begin{pmatrix} W_{AA}(\hat{N}) + \hat{N}\dfrac{dW_{AA}}{dN}\Big|_{\hat{N}} & -2\hat{N}\,(W_{Aa}(\hat{N}) - W_{AA}(\hat{N})) \\ 0 & W_{Aa}(\hat{N})/W_{AA}(\hat{N}) \end{pmatrix}.$$

(c) What are the eigenvalues of the Jacobian matrix in (b) (use the fact that $W_{AA}(\hat{n}) = 1$ to simplify these expressions as much as possible). (d) Provide an argument that, as with model (8.36) of the text, these two eigenvalues correspond, respectively, to a condition for ecological stability of the equilibrium with allele A fixed and to a condition for the stability of the genetic equilibrium with the population size fixed.

Further Reading

For a more complete mathematical treatment of population genetics models consult

- Bürger, R. 2000. *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Chichester.
- Christiansen, F. B. 1999. *Population Genetics of Multiple Loci*. Wiley, Chichester.
- Crow, J. F., and M. Kimura. 1970. *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- Hartl. D. L., and A. G. Clark. 1989. *Principles of Population Genetics*. Sinauer Associates, Sunderland. Mass.

References

Baeza, J. A., and W. Stotz. 2003. Host-use and selection of differently colored sea anemones by the symbiotic crab *Allopetrolisthes spinifrons*. *J. Exp. Mar. Biol. Ed*. 284:25–39.

Barton, N. H. 1995. A general model for the evolution of recombination. *Genet. Res*. 65:123–144.

Barton, N. H., and M. Turelli. 1991. Natural and sexual selection on many loci. *Genetics* 127:229–255.

Blower, S. M., H. B. Gershengorn, and R. M. Grant. 2000. A tale of two futures: HIV and antiretroviral therapy in San Francisco. *Science* 287:650–654

Bürger, R. 2000. *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Chichester.

Christiansen, F. B. 2000. *Population Genetics of Multiple Loci*. Wiley, Chichester.

Crow, J. F., and M. Kimura. 1965. Evolution in sexual and asexual populations. *Amer. Nat*. 99:439–450.

Crow, J. F., and M. Kimura. 1970. *An Introduction to Population Genetics Theory*. Harper & Row, New York.

Edelstein-Keshet, L. 1988. *Mathematical Models in Biology*. McGraw-Hill, New York.

Hastings, A. 1997. *Population Biology: Concepts and Models*. Springer, New York.

Karlin, S. 1975. General two-locus selection models: some objectives, results and interpretations. *Theoretical Population Biology* 7:364–398.

Kirkpatrick, M., T. Johnson, and N. Barton. 2002. General models of multilocus evolution. *Genetics* 161:1727–1750.

Levin, S., and D. Pimentel. 1981. Selection of intermediate rates of increase in parasite-host systems. *Am*. *Nat*. 117:308–315.

Nowak, M. A., and R. M. May. 1994. Superinfection and the evolution of parasite virulence. *Proc. R. Soc. London, Ser. B* 255:81–89.

Otto, S. P. 1994. The role of deleterious and beneficial mutations in the evolution of ploidy levels. *Lectures on Mathematics in the Life Sciences* 25:69–96.

Otto, S. P., and D. B. Goldstein. 1992. Recombination and the evolution of diploidy. *Genetics* 131:745–751.

Otto, S. P., and J. Marks. 1996. Mating systems and the evolutionary transition between haploidy and diploidy. *Biol. J. Linnean Soc*. 57:197–218.

Perrot, V., S. Richerd, and M. Valero. 1991. Transition from haploidy to diploidy. *Nature* 351:315–317.

Phillips, A. N. 1996. Reduction of HIV concentration during acute infection: Independence from a specific immune response. *Science* 271:497–499.

Turelli, M., and N. H. Barton. 1990. Dynamics of polygenic characters under selection. *Theor. Popul. Biol*. 38:1–57.