5. DISCRETE DYNAMICAL SYSTEMS

In Chapter 5, we considered the dynamics of systems consisting of a single quantity in either discrete or continuous time. Here we consider the dynamics of certain systems consisting of several relating quantities in discrete time. These arise in a variety of settings and can have quite complicated behavior. However, we focus on the simplest of these systems, the linear systems. While biological behavior is almost always nonlinear, there are some important settings in which the behavior is linear. One of these is structured population dynamics in which environmental restrictions are unimportant or in which the model is intended for use only for a short period of time. The latter case is the situation of wildlife conservation, where the main goal is to determine whether a model population will grow or shrink in the long run. If it will grow, we may not care about the level to which it will grow, in which case the omission of environmental restrictions is unimportant. Another interesting example is genetic drift, in which we have mutations occurring at an approximately constant rate without effect on fitness. With natural selection absent, the genetic drift model is linear. In this case, we are dealing with processes that occur very slowly and our interest is in relating the time since the process began to the net amount of drift that has occurred.

We begin in Section 5.1 with an introduction to the dynamics of structured populations, by which we mean populations divided into classes by age, size, or stage. The models we obtain are analogous to exponential growth models for single quantities, except that the possibility of various distributions of population among the classes makes the exponential growth rate difficult to determine. We can, however, demonstrate for any example that the model does eventually tend toward exponential growth, and for problems with a limited number of classes, we can prescribe an intuitive method for determining both the eventual growth rate and the stable distribution of the population.

It is often the case that problems that can be solved by intuition can be solved more easily by a formal mathematical procedure based on prior conceptual development. The analysis of discrete linear dynamical systems is an outstanding example of this phenomenon. In Section 5.2, we develop some of the basic mathematical theory of matrix algebra, and then we apply this theory in Section 5.3 to the problem of determining the eventual growth rate and stable population distribution for structured models.

The mathematical tools developed in Section 5.2 are put to additional use in Section 5.4 with the development of the Cantor-Jukes model, the simplest model for measuring the extent to which two different species are genetically related. The matrix algebra theory of Sections 5.2 and 5.3 will also be used in Chapter 7 in the study of continuous systems, both linear and nonlinear. Indeed, this material is essential for much of the mathematical analysis done in biology, and the reader is advised to aim for its mastery.

5.1 Discrete Linear Systems

The simplest discrete model is the linear model,[1] which we may write as

$$N_{t+1} = \lambda N_t, \tag{5.1.1}$$

with λ a positive real number and N_0 given. We can rearrange this equation into a form that provides a simple interpretation of λ:

$$\lambda = \frac{N_{t+1}}{N_t} \tag{5.1.2}$$

is the constant factor by which the population is augmented at each time step. In particular, if $\lambda > 1$, the population grows without bound; however, if $\lambda < 1$, the population shrinks toward 0. The unique equilibrium $N = 0$ is asymptotically stable if $\lambda < 1$, neutrally stable if $\lambda = 0$, and unstable if $\lambda > 1$. We can write down an explicit solution for this simple model:

$$N_t = \lambda^t N_0. \tag{5.1.3}$$

In this chapter, we extend the basic ideas of the single linear model (5.1.1) to analogous models of populations with structure. After studying this section, you should be able to:

- Construct a structured discrete linear population model from a narrative description;

- Identify a narrative description from a structured discrete linear population model;

- Describe the general behavior of structured discrete linear population models in terms of the growth rate and stable population ratios; and

- Determine the long-term growth rate and stable population ratios for a structured discrete linear population model.

Simple structured models

structured population model: *a population model in which individuals are categorized according to some discrete or continuous property.*

We restrict consideration to models in which individuals are divided into discrete classes.

Example 5.1.1
A population consists of juveniles and adults and changes from year to year through survival and reproduction:

- Ten percent of the juveniles survive to become adults.

- All adults die after one year.

- Adults produce an average of 20 juveniles in their single season of life, through reproduction.

- Juveniles also reproduce, with an average of one juvenile offspring each.

---

[1]Section 5.1.

Using these assumptions, we create a model that consists of two equations that compute J_{t+1} and A_{t+1} in terms of the populations of both at time t:

$$J_{t+1} = J_t + 20A_t, \qquad A_{t+1} = 0.1J_t. \tag{5.1.4}$$

◇

As in the single equation model, we can run simulations with model 5.1.4 if we prescribe initial population values at time 0.

Example 5.1.2
Suppose we start with 200 juveniles and 10 adults; that is,

$$J_0 = 200, \qquad A_0 = 10.$$

Then the populations at time 1 are found using $t = 0$ in (5.1.4):

$$J_1 = J_0 + 20A_0 = 400, \qquad A_1 = 0.1J_0 = 20.$$

Similarly, we find

$$J_2 = J_1 + 20A_1 = 800, \qquad A_2 = 0.1J_1 = 40.$$

We see that the populations of both classes double in each time step. In other words,

$$\frac{J_{t+1}}{J_t} = 2, \qquad \frac{A_{t+1}}{A_t} = 2.$$

Alternatively, we may write an exact solution as

$$J_t = 2^t J_0, \qquad A_t = 2^t A_0.$$

◇

In Example 5.1.2, the structured model 5.1.4 displays behavior similar to the unstructured model 5.1.1. However, we should be careful not to draw too strong a conclusion. The simulation results are more complicated if we start with different initial populations.

Example 5.1.3
Consider the model 5.1.4 with an initial population of 100 juveniles and 10 adults. Table 5.3.1 shows the population growth over several time steps. The population does not double at each time step, which would require $J_{t+1}/J_t = 2$ and $A_{t+1}/A_t = 2$ for each t. However, the ratios of successive populations do approach 2 gradually as time increases. ◇

Time	0	1	2	3	4	5
J	100	300	500	1100	2100	4300
A	10	10	30	50	110	210
J_{t+1}/J_t		3	1.67	2.2	1.91	2.05
A_{t+1}/A_t		1	3	1.67	2.2	1.91
J/A	10	30	16.7	22	19.1	20.5

Table 5.1.1: Populations for Equation 5.1.4 with $J_0 = 100$, $A_0 = 10$

The comparison between Examples 5.1.3 and 5.1.4 is instructive. In both cases, the model predicts a growth rate of 2; however, the simulation produces a growth rate of exactly 2 in the first instance and only approaches 2 gradually in the second. Notice that the ratio of J to A was initially 20:1 in Example 5.1.3. Because both population classes doubled at each time step, the ratio of 20:1 was maintained. In Example 5.1.4, the ratio of J to A was initially 10:1. It did not remain at that value, but changed at each time step. The J:A ratio for this simulation appears to approach 20:1 as the growth rate approaches 2.

Examples 5.1.3 and 5.1.4 illustrate the basic behavior of discrete linear structured models. There is a special ratio of initial populations for which the model shows growth at a constant rate $\lambda > 0$.[2] If a simulation starts with a different ratio of initial populations, then the predicted populations will only gradually settle into a pattern with the same characteristic growth rate and proportions.

Finding the growth rate and stable distribution

We have discovered that discrete linear structured models have a characteristic growth rate and a characteristic population ratio for which that growth rate applies immediately. It would be nice to have a method for computing these quantities that does not require use of simulations. Indeed, we will develop a powerful mathematical technique in the next two sections, but first we solve the problem by using basic principles rather than sophisticated mathematical formalism.

Example 5.1.4
To find the characteristic growth rate and population ratios for model 5.1.4, we begin by assuming that the initial populations are $J_0 = j$ and $A_0 = a$, and that these have the correct ratio to produce a growth rate of λ that is correct at each time step. We may freely take $a = 1$, which will leave us with two unknowns, λ and j. We can find two equations for these unknowns by calculating J_1 and A_1 in two different ways.

- The model 5.1.4 allows us to compute J_1 and A_1 for any initial values. Here,

$$J_1 = J_0 + 20A_0 = j + 20a, \qquad A_1 = 0.1J_0 = 0.1j.$$

- If we choose j and a correctly, then these populations will change by a factor of λ in each time step; hence,

$$J_1 = \lambda J_0 = \lambda j, \qquad A_1 = \lambda A_0 = \lambda a.$$

Combining the results from our calculations yields two equations:

$$j + 20a = \lambda j, \qquad 0.1j = \lambda a.$$

We have two equations with three unknowns, but we do not need to find unique values for j and a. It is only the ratio j:a that matters. Hence, we do not lose any generality by arbitrarily choosing $a = 1$. With this simplification, the equations become

$$j + 20 = \lambda j, \qquad 0.1j = \lambda.$$

For reasons that will become more clear later, the best way to proceed is to solve one of the equations for j and substitute into the other to obtain an equation for λ. The second equation gives us $j = 10\lambda$, and then the first becomes

$$10\lambda + 20 = 10\lambda^2,$$

---

[2]Of course there could be "growth" at a rate less than 1 for some models.

or
$$0 = \lambda^2 - \lambda - 2 = (\lambda - 2)(\lambda + 1).$$

Mathematically, we have two results for λ, 2 and -1. From our simulations, we know that $\lambda = 2$ is the correct value. Given this result, we have $j = 10\lambda = 20$, yielding a population ratio of 20:1. ◇

General properties of linear discrete structured models

Example 5.1.4 suggests some mathematical questions. What sort of equation do we get for λ in general? How do we know which of the possible values for λ is correct? Here we answer these questions without proof. These issues will be explored in more depth later in the chapter.

- Linear discrete structured population models have a long-term growth rate λ, which is the largest positive root of a polynomial of degree n, where n is the number of classes comprising the population.

Example 5.1.5
A population consists of larvae, young adults, and older adults. One percent of larvae grow into young adults each year, and 30% of young adults survive to become older adults. Young adults have an average of 104 offspring each year, while older adults have an average of 160 offspring. Find the long-term growth rate and population ratio.

From the description, we obtain the model

$$
\begin{align}
L_{t+1} &= 104Y_t + 160A_t, & \text{(5.1.5)} \\
Y_{t+1} &= 0.01L_t, & \text{(5.1.6)} \\
A_{t+1} &= 0.3Y_t. & \text{(5.1.7)}
\end{align}
$$

Now suppose the initial populations are l, y, and a and the long-term growth rate is λ. Using the model, we obtain

$$L_1 = 104y + 160, \qquad Y_1 = 0.01l, \qquad A_1 = 0.3y.$$

The assumption that $l{:}y{:}a$ is the ratio of populations corresponding to growth at rate λ yields

$$L_1 = \lambda l, \qquad Y_1 = \lambda y, \qquad A_1 = \lambda a.$$

Combining these two sets of equations yields three equations for λ, l, y, and a:

$$
\begin{align}
104y + 160 &= \lambda l, \\
0.01l &= \lambda y, \\
0.3y &= \lambda a.
\end{align}
$$

Now we set $a = 1$ without loss of generality. Then we have

$$y = \frac{1}{0.3}\lambda$$

from the third equation and

$$l = 100\lambda y = \frac{100}{0.3}\lambda^2$$

from the second. The first equation then becomes

$$\frac{104}{0.3}\lambda + 160 = \frac{100}{0.3}\lambda^3,$$

or

$$104\lambda + 48 = 100\lambda^3,$$

or

$$f(\lambda) = 100\lambda^3 - 104\lambda - 48 = 0. \tag{5.1.8}$$

Equation 5.1.8 is a polynomial equation of third degree, with f called the **characteristic polynomial** for the model. There is no convenient solution formula for it, nor are we likely to be able to guess the correct factors. However, there is no difficulty getting an approximate solution by graphing. Figure 5.1.1 shows the function f, indicating that the solution is approximately $\lambda = 1.2$. In fact, this is seen to solve the equation exactly.

Once λ is known, we can quickly recover the other variables:

$$l = \frac{100}{0.3}\lambda^2 = 480, \qquad y = \frac{1}{0.3}\lambda = 4.$$

The stable age distribution has 480 larvae and 4 young adults for every adult. We can also compute the fractions of the adults in the population as $1/485 \approx 0.002$ older adults and $4/485 \approx 0.008$ younger adults. Figure 5.1.2 shows the populations and the relative growth rates for a simulation beginning with 1000 larvae, 50 young adults, and 5 old adults. Figure 5.1.2a shows the actual populations and Figure 5.1.2b shows the ratios of successive populations. ◇
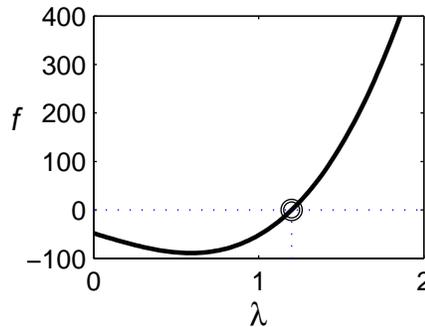

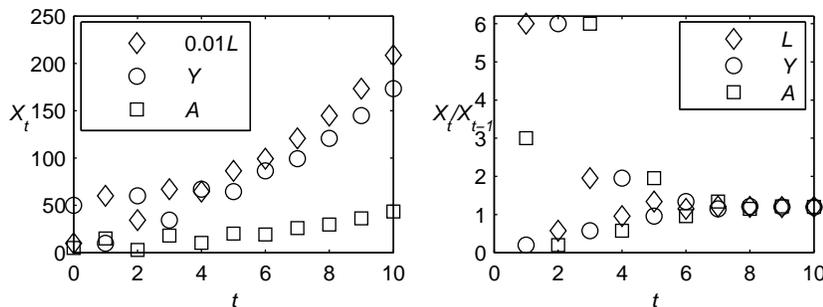Figure 5.1.1: The characteristic polynomial $f(\lambda)$ for Example 5.1.5.

Figure 5.1.2: Populations and population ratios for Example 5.1.5.

Note that plots of populations show the trend of accelerating growth, while plots of population ratios give more detailed information about the manner of growth.

5.2 A Matrix Algebra Primer

A working knowledge of matrix models in biology requires some understanding of topics in matrix algebra. We begin with the mathematical definitions needed so that linear discrete structured models can be written in matrix-vector form. We then study what for us is the central problem of matrix algebra: given a non-zero matrix $\mathbf{A}$, find (if possible) a non-zero vector $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$.

After studying this section, you should be able to:

- Multiply matrices times vectors;

- Write discrete linear systems in matrix-vector form;

- Compute determinants of 2×2 and 3×3 matrices;

- Determine whether an equation $\mathbf{A}\mathbf{x} = \mathbf{0}$ has nonzero solutions for a given nonzero matrix $\mathbf{A}$;

- Compute nonzero solutions of an equation $\mathbf{A}\mathbf{x} = \mathbf{0}$ when these exist.

Matrices and vectors

an n-vector $\mathbf{x}$: *a column of n numbers x_1, x_2, ..., x_n.*

Example 5.2.1
In Example 5.1.3, we considered a model of a population with classes J and A, and we had initial populations of $J_0 = 100$ and $A_0 = 10$. We can define a 2-vector $\mathbf{x}$ whose components are J and A. Thus,

$$\mathbf{x} = \begin{pmatrix} J \\ A \end{pmatrix}, \qquad \mathbf{x}_0 = \begin{pmatrix} J_0 \\ A_0 \end{pmatrix} = \begin{pmatrix} 100 \\ 10 \end{pmatrix}.$$

$\diamond$

an $n \times n$ square matrix $\mathbf{A}$: *a rectangular array of numbers with n rows and n columns, with a_{ij} as the entry in row i and column j.*

the main diagonal of an $n \times n$ square matrix $\mathbf{A}$: *the entries a_{11}, a_{22}, ..., a_{nn}.*

Example 5.2.2
The model of Example 5.1.3 can be written in the form

$$\begin{aligned} J_{t+1} &= 1J_t + 20A_t, \\ A_{t+1} &= 0.1J_t + 0A_t. \end{aligned}$$

The four coefficients in these equations can be arranged as a 2×2 matrix:

$$\mathbf{M} = \begin{pmatrix} 1 & 20 \\ 0.1 & 0 \end{pmatrix}.$$

$\diamond$

the $n \times n$ identity matrix I: *the matrix*

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

consisting of ones on the main diagonal and zeros everywhere else.

The identity matrix is useful notation to use when we want to subtract a common value λ from each entry on the main diagonal of a matrix $\mathbf{M}$.

Example 5.2.3
Let $\mathbf{M}$ be the matrix in Example 5.2.2. Let λ be an unknown number. Define a matrix $\mathbf{A}$ by $\mathbf{A} = \mathbf{M} - \lambda\mathbf{I}$. Then

$$\mathbf{A} = \begin{pmatrix} 1 & 20 \\ 0.1 & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 20 \\ 0.1 & 0 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} 1-\lambda & 20 \\ 0.1 & -\lambda \end{pmatrix}.$$

Notice that $\mathbf{A}$ is actually a family of matrices, with λ as the parameter. $\diamond$

the matrix product of an $n \times n$ square matrix A and an n-vector x: *the vector*

$$\mathbf{Ax} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{pmatrix}. \quad (5.2.1)$$

Example 5.2.4
Let $\mathbf{M}$ be the matrix of Examples 5.2.2 and 5.2.3 and let $\mathbf{x}_0$ be as in Example 5.2.1. We can multiply $\mathbf{x}_0$ on the left by $\mathbf{M}$:

$$\mathbf{Mx}_0 = \begin{pmatrix} 1 & 20 \\ 0.1 & 0 \end{pmatrix} \begin{pmatrix} 100 \\ 10 \end{pmatrix} = \begin{pmatrix} (1)(100) + (20)(10) \\ (0.1)(100) + (0)(10) \end{pmatrix} = \begin{pmatrix} 300 \\ 10 \end{pmatrix}.$$

$\diamond$

Note that the product $\mathbf{Ax}$ of a square matrix $\mathbf{A}$ and a vector $\mathbf{x}$ is defined only when the matrix is on the left and both matrix and vector have the same size n. The product also a vector of size n.

The population model in matrix notation

The alert reader may have noticed the correspondence between the preceding examples and Example 5.1.3. Using the matrix $\mathbf{M}$ and vector $\mathbf{x}$, we observe the relationship

$$\mathbf{Mx}_t = \begin{pmatrix} 1 & 20 \\ 0.1 & 0 \end{pmatrix} \begin{pmatrix} J_t \\ A_t \end{pmatrix} = \begin{pmatrix} J_t + 20A_t \\ 0.1J_t \end{pmatrix} = \begin{pmatrix} J_{t+1} \\ A_{t+1} \end{pmatrix} = \mathbf{x}_{t+1}.$$

We can do the same thing with any linear discrete structured model. Thus,

- **All discrete linear population models can be written in the form $\mathbf{x}_{t+1} = \mathbf{Mx}_t$.**

Example 5.2.5

In Example 5.1.5, we encountered the model

$$
\begin{aligned}
L_{t+1} &= 104Y_t + 160A_t, \\
Y_{t+1} &= 0.01L_t, \\
A_{t+1} &= 0.3Y_t.
\end{aligned}
$$

This model takes the form $\mathbf{x}_{t+1} = \mathbf{M}\mathbf{x}_t$, with

$$
\mathbf{x} = \begin{pmatrix} L \\ Y \\ A \end{pmatrix}, \qquad
\mathbf{M} = \begin{pmatrix} 0 & 104 & 160 \\ 0.01 & 0 & 0 \\ 0 & 0.3 & 0 \end{pmatrix}.
$$

If we start with 1000 larvae, 50 young adults, and 5 old adults, we can compute the populations at time 1 by matrix multiplication:

$$
\begin{pmatrix} L \\ Y \\ A \end{pmatrix} =
\begin{pmatrix} 0 & 104 & 160 \\ 0.01 & 0 & 0 \\ 0 & 0.3 & 0 \end{pmatrix}
\begin{pmatrix} 1000 \\ 50 \\ 5 \end{pmatrix} =
\begin{pmatrix} 6000 \\ 10 \\ 15 \end{pmatrix}.
$$

$\diamond$

The central problem of matrix algebra

Let $\mathbf{A}$ be an $n \times n$ matrix with at least one non-zero entry and let $\mathbf{0}$ be the n vector whose entries are all 0. For any given matrix $\mathbf{A}$, we are interested in finding non-zero solutions $\mathbf{x}$ for the equation

$$
\mathbf{A}\mathbf{x} = \mathbf{0}.
$$

It is instructive to examine the corresponding scalar problem. The equation is $ax = 0$, where $a \neq 0$. Of course the only solution is $x = 0$. This, however, is not necessarily the case for the matrix equation.

Example 5.2.6

Let $\mathbf{A}$, $\mathbf{x}$, and $\mathbf{y}$ be given by

$$
\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad
\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.
$$

Then

$$
\mathbf{A}\mathbf{x} = \mathbf{0},
$$

which demonstrates that there are nonzero solutions for this particular matrix. $\diamond$

Example 5.2.7

Are there any nonzero solutions to $\mathbf{I}\mathbf{x} = \mathbf{0}$, where $\mathbf{I}$ is the identity matrix?

Suppose

$$
\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},
$$

with x_1 and x_2 to be determined. Then

$$
\mathbf{I}\mathbf{x} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.
$$

The equation $\mathbf{Ix} = \mathbf{0}$ corresponds to the equation pair $x_1 = 0$, $x_2 = 0$. Clearly these equation have no non-zero solutions. ◇

Examples 5.2.6 and 5.2.7 demonstrate that nonzero solutions exist for some matrices, but not for others. Is there some way to predict whether a given matrix has nonzero solutions without first trying to find solutions? This is the crucial question for us to answer.

The determinant

We first need to develop the computational tool called the *determinant*. The determinant of a matrix is a number that is calculated using the entries in the matrix. A thorough treatment of the determinant requires a lot of linear algebra background. Fortunately, the formulas for the determinants of 2×2 and a 3×3 matrices are simple. We present them here as definitions.

the determinant of a 2×2 matrix A: *the quantity*

$$\det(\mathbf{A}) = ad - bc,$$

given

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

the determinant of a 3×3 matrix A: *the quantity*

$$\det(\mathbf{A}) = (aei + bfg + cdh) - (ceg + bdi + afh),$$

given

$$\mathbf{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}.$$

These formulas may seem almost random at first, but there is a pattern. Each of the positive terms is a product of elements aligned diagonally from top left to bottom right (lightly copying the first two columns of the 3×3 determinant to the right of the matrix, giving the appearance of a 3×5 matrix, helps to see this), and each of the negative terms is a product of elements aligned diagonally from top right to bottom left.[1]

Example 5.2.8
Let
$$\mathbf{A} = \begin{pmatrix} -1.2 & 104 & 160 \\ 0.01 & -1.2 & 0 \\ 0 & 0.3 & -1.2 \end{pmatrix}.$$

Then
$$\det(\mathbf{A}) = (-1.2)(-1.2)(-1.2) + (104)(0)(0) + (160)(0.01)(0.3) - (160)(-1.2)(0)$$
$$-(104)(0.01)(-1.2) - (-1.2)(0)(0.3) = -1.728 + 0.48 + 1.248 = 0.$$

◇

---

[1]These patterns do NOT hold in higher-dimensional determinants. The reader who wants to work with higher-dimensional matrices should consult a linear algebra book for a complete definition of the determinant.

The equation $\mathbf{Ax} = \mathbf{0}$

The determinant is an efficient way to identify matrices for which $\mathbf{Ax} = \mathbf{0}$ has non-zero solutions. We state the principal result without proof.

Theorem 5.2.1 *The equation $\mathbf{Ax} = \mathbf{0}$ has non-zero solutions for $\mathbf{x}$ if and only if $\det(\mathbf{A}) = 0$.*

A procedure for solving $\mathbf{Ax} = \mathbf{0}$ when $\det(\mathbf{A}) = 0$ is illustrated by an example.

Example 5.2.9
Let $\mathbf{A}$ be the matrix from Examples 5.2.2 and 5.2.3, and let x_1, x_2, and x_3 be the components of $\mathbf{x}$. Then $\mathbf{Ax} = \mathbf{0}$ is equivalent to the scalar equations

$$-1.2x_1 + 104x_2 + 160x_3 = 0, \quad 0.01x_1 - 1.2x_2 = 0, \quad 0.3x_2 - 1.2x_3 = 0.$$

Because $\det(\mathbf{A}) = 0$, we know that this system of equations has nonzero solutions. We can try to find a solution for which $x_3 = 1$; there is no harm in doing this, as we can try something else if it fails. Once we have $x_3 = 1$, then the third equation yields $x_2 = 4$. The second equation then yields $x_1 = 480$. We have apparently found a solution

$$\mathbf{x} = \begin{pmatrix} 480 \\ 4 \\ 1 \end{pmatrix}.$$

Note that we never used the first of the three equations. This equation is available as a check. Substituting the results into the first equation yields

$$(-1.2)(480) + (104)(4) + (160)(1) = 0,$$

which confirms that our solution is indeed correct. $\diamond$

Was the choice of $x_3 = 1$ special? No, it was convenient, but not necessary. We could have chosen any nonzero value for any of the three variables. We would have got a different answer, but all the answers are simple multiples of our answer. For example, had we started with $x_2 = 1$, we would have found $x_3 = 0.25$ and $x_1 = 120$. The ratio $x_1:x_2:x_3$ is 480:4:1 in all cases.

In general, the procedure of Example 5.2.9 could have failed only if the correct solutions required $x_3 = 0$. Had that been the case, we would simply have tried again with $x_3 = 0$ rather than $x_3 = 1$.

The matrix of Example 5.2.2 has a one-parameter family of solutions. Occasionally a matrix can have a larger family of solutions, but this does not occur in the models we will be considering.

5.3 Long-term Behavior of Linear Models

We studied the mathematics of discrete linear systems in Section 5.1, using unsophisticated scalar methods. Then we learned the basics of matrix algebra in Section 5.2. We are now ready to apply matrix methods to discrete linear systems.

After studying this section, you should be able to:

- Explain the biological significance of eigenvalues and eigenvectors;

- Compute (real-valued) eigenvalues and their associated eigenvectors;

- Determine the dominant eigenvalue of a matrix;

- Describe the long-term behavior of discrete linear models.

Eigenvalues[1] and eigenvectors

As we saw in Section 5.1, linear discrete population models exhibit growth at a uniform constant rate when the initial conditions are just right. Suppose $\mathbf{x}_{t+1} = \mathbf{M}\mathbf{x}_t$ and $\mathbf{x}_0 = \mathbf{v}$ is an initial condition for which growth occurs at a constant rate λ. Then we can calculate $\mathbf{x}_1$ in two ways:

$$\mathbf{x}_1 = \mathbf{M}\mathbf{x}_0 = \mathbf{M}\mathbf{v}, \qquad \mathbf{x}_1 = \lambda\mathbf{x}_0 = \lambda\mathbf{v}.$$

Combining these equations together, we have the problem

$$\mathbf{M}\mathbf{v} = \lambda\mathbf{v}. \tag{5.3.1}$$

This is the *eigenvalue problem* of matrix algebra:

eigenvalue problem for a nonzero matrix M: *the problem of finding nonzero solutions to Equation 5.3.1.*

eigenvalue: *given a nonzero matrix* $\mathbf{M}$, *a value of* λ *for which Equation 5.3.1 has nonzero solutions.*

eigenvector: *a vector* $\mathbf{v}$ *that solves the eigenvalue problem for a given eigenvalue*

To solve Equation 5.3.1, we must recast it in a more convenient form. Using the identity matrix $\mathbf{I}$,[2] we have $\mathbf{v} = \mathbf{I}\mathbf{v}$, and this allows us to write Equation 5.3.2 as

$$\mathbf{M}\mathbf{v} = \lambda\mathbf{v} = \lambda\mathbf{I}\mathbf{v},$$

or

$$\mathbf{M}\mathbf{v} - \lambda\mathbf{I}\mathbf{v} = \mathbf{0}.$$

This change allows us to use the distributive property of matrix multiplication (factoring out the $\mathbf{v}$) to get the equivalent equation

$$(\mathbf{M} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}. \tag{5.3.2}$$

The key to solving Equation 5.3.2 is to see that it is of the form $\mathbf{A}\mathbf{x} = \mathbf{0}$, which we studied in Section 5.2. By Theorem 5.2.1, the equation has a solution if and only if

$$\det(\mathbf{M} - \lambda\mathbf{I}) = 0. \tag{5.3.3}$$

---

[1] pronounced "eye-gen-value," with a hard g as in "get"
[2] See Section 5.2.

Example 5.3.1
Let $\mathbf{M}$ be the matrix from Example 5.2.2. We have

$$\mathbf{M} - \lambda\mathbf{I} = \begin{pmatrix} 1 & 20 \\ 0.1 & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1-\lambda & 20 \\ 0.1 & -\lambda \end{pmatrix}.$$

Thus,

$$0 = \det(\mathbf{M} - \lambda\mathbf{I}) = \begin{vmatrix} 1-\lambda & 20 \\ 0.1 & -\lambda \end{vmatrix} = (1-\lambda)(-\lambda) - (20)(0.1) = \lambda^2 - \lambda - 2 = (\lambda - 2)(\lambda + 1).$$

The matrix $\mathbf{M}$ has eigenvalues 2 and -1. $\diamond$

Once the eigenvalues are known, we can determine eigenvectors from Equation 5.3.2.

Example 5.3.2
With $\mathbf{M}$ from Example 5.3.1 and $\lambda = 2$, we have

$$\mathbf{M} - \lambda\mathbf{I} = \begin{pmatrix} -1 & 20 \\ 0.1 & -2 \end{pmatrix}.$$

The system $(\mathbf{M} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ is equivalent to the scalar equations

$$-J + 20A = 0, \qquad 0.1J - 2A = 0.$$

These equations are redundant, so any solution of one is a solution of the other. Arbitrarily taking $A = 1$, the first equation yields $J = 20$, and this solution also satisfies the second equation. Any vector of the form

$$\mathbf{x} = c_1 \begin{pmatrix} 20 \\ 1 \end{pmatrix}$$

is an eigenvector corresponding to $\lambda = 2$. Similarly, the reader should determine that any vector of the form

$$\mathbf{x} = c_2 \begin{pmatrix} 10 \\ -1 \end{pmatrix}$$

is an eigenvector corresponding to $\lambda = -1$. $\diamond$

Example 5.3.3
Let

$$\mathbf{M} = \begin{pmatrix} 0 & 104 & 160 \\ 0.01 & 0 & 0 \\ 0 & 0.3 & 0 \end{pmatrix},$$

as in Examples 5.2.5. We have

$$0 = \det(\mathbf{M} - \lambda\mathbf{I}) = \begin{vmatrix} -\lambda & 104 & 160 \\ 0.01 & -\lambda & 0 \\ 0 & 0.3 & -\lambda \end{vmatrix} = [-\lambda^3 + (160)(0.01)(0.3)] - [-(104)(0.1)(\lambda)] = -\lambda^3 + 10.4\lambda + 0.48.$$

We derived this equation in Example 5.1.5, where we found the root $\lambda = 1.2$.
The equation $(\mathbf{M} - 1.2\mathbf{I})\mathbf{x} = \mathbf{0}$ is equivalent to the scalar equations

$$-1.2J + 104Y + 160A = 0, \qquad 0.01J - 1.2Y = 0, \qquad 0.3Y - 1.2A = 0.$$

Taking $A = 1$, the third and second equations yield $Y = 4$ and $J = 480$ respectively. Substituting these values into the first equation confirms that the answers are correct. Note that it is easiest to find the eigenvector by breaking the matrix equation into a set of scalar equations. $\diamond$

Algorithm 5.3.1 summarizes the procedure for finding eigenvalues and eigenvectors.

Algorithm 5.3.1 *Eigenvalues and eigenvectors of a matrix* $\mathbf{M}$ *are the scalars* λ *and vectors* $\mathbf{x} \neq \mathbf{0}$ *that solve the equation*

$$\mathbf{M}\mathbf{x} = \lambda\mathbf{x}.$$

To find eigenvalues and eigenvectors,

1. *Find the polynomial equation that corresponds to the equation*

$$\det(\mathbf{M} - \lambda\mathbf{I}) = 0.$$

2. *Substitute an eigenvalue* λ *into the equation*

$$(\mathbf{M} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0},$$

then rewrite this equation as a set of scalar equations.

3. *Set one of the scalar unknowns to 1 and solve for the remaining unknowns.[3] This leaves one unused scalar equation to check the solutions.*

4. *If the last scalar equation does not check, there is an error; most likely the value of* λ *that you used is not actually an eigenvalue.*

Solutions of $\mathbf{x_{t+1}} = \mathbf{M}\mathbf{x_t}$

The eigenvalues and eigenvectors of a matrix $\mathbf{M}$ can be used to determine the set of all solutions to the matrix model $\mathbf{x_{t+1}} = \mathbf{M}\mathbf{x_t}$, according to the following theorem.

Theorem 5.3.1 *If* $\lambda_1, \ldots, \lambda_n$ *are distinct eigenvalues of an* $n \times n$ *matrix* $\mathbf{M}$ *with associated eigenvectors* $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$, *then all solutions of the equation*

$$\mathbf{x_{t+1}} = \mathbf{M}\mathbf{x_t}$$

have the form

$$\mathbf{x_t} = c_1\lambda_1^t\mathbf{x}^{(1)} + \cdots + c_n\lambda_n^t\mathbf{x}^{(n)},$$

where $c_1, \ldots, c_n$ *are constants.*

Example 5.3.4
Continuing with Examples 5.3.2, we have the solution

$$\mathbf{x_t} = c_1 2^t \begin{pmatrix} 20 \\ 1 \end{pmatrix} + c_2(-1)^t \begin{pmatrix} 10 \\ -1 \end{pmatrix}.$$

The values of c_1 and c_2 are determined by the initial data, which is $J(0) = 100$ and $A(0) = 10$. Thus,

$$\begin{pmatrix} 100 \\ 10 \end{pmatrix} = \mathbf{x_0} = c_1 \begin{pmatrix} 20 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 10 \\ -1 \end{pmatrix}.$$

---

[3]This won't work if you choose a scalar unknown whose value needs to be 0, but you can try again with that scalar unknown set to 0 instead.

This vector equation corresponds to the scalar equations

$$20c_1 + 10c_2 = 100, \qquad c_1 - c_2 = 10.$$

We can solve these equations by adding 10 times the second to the first, with the result $30c_1 = 200$; hence, $c_1 = 20/3$ and then $c_2 = -10/3$. Thus,

$$\mathbf{x_t} = \frac{20}{3}2^t \begin{pmatrix} 20 \\ 1 \end{pmatrix} - \frac{10}{3}(-1)^t \begin{pmatrix} 10 \\ -1 \end{pmatrix}.$$

$\diamond$

Long-term behavior

In ecological modeling, we are usually concerned with the general characteristics of models rather than the detailed results of specific simulations. The important thing to note about the solution in Example 5.3.4 is *not* the solution formula itself; rather, it is that 2^t dominates $(-1)^t$ as t increases. Regardless of the values of c_1 and c_2 (as long as $c_1 \neq 0$), the solution of the model ultimately approaches

$$\mathbf{x_t} = c_1 2^t \begin{pmatrix} 20 \\ 1 \end{pmatrix}.$$

For this reason, the eigenvalue that matters to the long-term behavior is the one with largest magnitude. The *value* of c_1, as long as it isn't 0, is not very important because the models are too crude to be used for quantitative prediction. The goal is to predict the qualitative behavior, and for this it is the dominant eigenvalue, and its associated eigenvector, that matter.

Theorem 5.3.2 *Suppose λ_1 has a larger magnitude than the other eigenvalues of a matrix $\mathbf{M}$. Then almost all solutions of $\mathbf{x_{t+1}} = \mathbf{M}\mathbf{x_t}$ eventually approach*

$$\mathbf{x_t} = c_1 \lambda_1^t \mathbf{x^{(1)}}.$$

Thus, the eigenvalue of largest magnitude determines the long-term growth rate and the corresponding eigenvector determines the long-term solution ratios.

Example 5.3.5
Let

$$\mathbf{M} = \begin{pmatrix} 0 & 104 & 160 \\ 0.01 & 0 & 0 \\ 0 & 0.3 & 0 \end{pmatrix}.$$

The dominant eigenvalue is $\lambda = 1.2$, which we found in Example 5.3.3. From Theorem 5.3.2, we have the qualitative result: the long-term behavior of the model is growth at a rate of 20% with the stable age distribution of 480:4:1. $\diamond$

5.4 Markov Chains

So far, we've used equations of the form

$$\mathbf{x}_{t+1} = \mathbf{M}\mathbf{x}_t$$

to model changes in a set of dynamic variables (the components of $\mathbf{x}$) that represent population sizes. The same mathematical structure applies to models that track dynamic changes in probabilities. These mathematical models have become useful tools in molecular biology, leading to discoveries about the development of species in evolutionary history. The full subject is very complicated, but we can get a general sense of the possibilities by examining the simplest mathematical model for genetic change. In this section, we consider the problem of estimating the *phylogenetic distance* between species. This concept refers to the overall amount of genetic difference between genomes; we'll give a precise mathematical definition later. Phylogenetic distance is an important concept because it has caused significant changes in biologists' understanding of the evolutionary relationships of species. Based on observable characteristics, scientists had long thought that chimpanzees were more closely related to gorillas than to humans. We now know that the phylogenetic distance between the chimpanzee genome and the human genome is less than that between the chimpanzee and gorilla genomes, and therefore chimpanzees are more closely related to humans than gorillas.[1]

Some scientific background

We've seen earlier that DNA carries information in the pattern of nucleotides that are attached to sites on the basic DNA molecule. The nucleotides come in four types, labeled A, G, C, and T.

A variety of processes govern changes in the pattern as generations pass. The most important processes are the copying process and the process whereby chromosomes from the male and female gametes are combined to make a genetically distinct offspring. In the special case of the Y chromosome, there is no contribution from the female gamete; hence, the only changes in the Y chromosome from one generation to the next are those caused by the rare copying error or other mutation.

Evolution happens because of two processes: mutation and natural selection. Mutation is caused by the occasional copying error, while natural selection is caused by differences in fitness of organisms having genetic differences. We can think of DNA strands in chromosomes as falling broadly into three categories.

1. Some DNA is in the form of genes that are crucial to species survival, such as the genes that determine the network of blood vessels in mammals. These genes are largely resistant to change because mutations tend to be harmful. The corresponding DNA may be different between species, but will likely be almost the same for individuals in a species.

2. Some DNA is in the form of genes that play at best a small role in species survival, such as the genes that determine hair color. The corresponding DNA shows significant variation within a population that persists over many generations. This DNA is useful for identification of individuals in a species.

---

[1] We are thinking specifically of the total amount of genetic difference in the genomes. One could still argue that the smaller number of differences between chimpanzee and human are more important than the larger number between chimpanzee and gorilla.

3. There is also DNA that is called *non-coding* and sometimes referred to as "junk DNA." This DNA does not affect the characteristics of the organism, but it is part of the genome as a residue of the evolutionary past. Non-coding DNA is particularly useful for gaining information about the relationships of species on the evolutionary family tree; it changes from mutation, but not from natural selection. This means that the rate at which non-coding DNA changes is independent of changing environmental conditions. Changes in non-coding DNA serve as a molecular clock for the measurement of evolutionary time.

A model for DNA change

Suppose we have sequenced a strand of non-coding DNA from the Y chromosome of a human and have for comparison a strand of DNA that corresponds to an ancestor of our species.[2] Assume that there are J nucleotides in the strand and let N be the unknown number of generations that have passed between the ancestral strand and the contemporary strand. For any position in the sequence, the nucleotide must be either A, G, C, or T. By comparing the ancestral and contemporary strands, we can measure the fraction of DNA sites that are different between the two strands. Let's call this value β. Our task is to construct a mathematical model that somehow connects the measured value of β with the unknown number of generations N.

Now let α be the probability of a mutation in one generation in any site. We expect the total number of mutations in one site over N generations to be αN and the total number of mutations in the strand to be $\alpha N J$. At first thought, this sounds easy. The total number of differences between the strands is βJ and the total number of mutations is $\alpha N J$. These should be equal, so $\alpha N = \beta$. However, this reasoning is flawed. If a site starts as A, mutates to G, and then mutates back to A, and there are no further changes, then $\alpha N J = 2$. However, the two strands are identical because the second mutation reversed the first one, so $\beta J = 0$. Thus, $\alpha N > \beta$, because some mutations do not increase the number of differences between the strands. We need a mathematical model that connects mutations in one generation with long-term sequence differences.

Let $p_A(n)$, $p_G(n)$, $p_C(n)$, and $p_T(n)$ be the probabilities of each given nucleotide at a particular site in the nth generation. Mutations from one generation to the next change these probabilities, and we must quantify these changes. The simplest assumption is that all possible changes are equally likely. Since α is the probability of change, and each nucleotide has three possible changes, the probability of any particular change is $\alpha/3$. Of course the probability of no change is $1 - \alpha$. With these assumptions, the probability that a site will contain the nucleotide A at time $n + 1$ can be calculated from the probabilities of the different nucleotides at time n as

$$p_A(n + 1) = (1 - \alpha)p_A(n) + \frac{\alpha}{3}p_G(n) + \frac{\alpha}{3}p_C(n) + \frac{\alpha}{3}p_T(n). \tag{5.4.1}$$

Following this reasoning, we can write a single matrix equation for the change in probabilities:

$$\mathbf{x}_{n+1} = \mathbf{M}\mathbf{x}_n, \tag{5.4.2}$$

where the probability vector $\mathbf{x}$ and the transition matrix $\mathbf{M}$ are

$$\mathbf{x} = \begin{pmatrix} p_A \\ p_G \\ p_C \\ p_T \end{pmatrix}, \qquad \mathbf{M} = \begin{pmatrix} 1 - \alpha & \alpha/3 & \alpha/3 & \alpha/3 \\ \alpha/3 & 1 - \alpha & \alpha/3 & \alpha/3 \\ \alpha/3 & \alpha/3 & 1 - \alpha & \alpha/3 \\ \alpha/3 & \alpha/3 & \alpha/3 & 1 - \alpha \end{pmatrix}. \tag{5.4.3}$$

[2]We set aside all questions regarding the plausibility of these assumptions, focusing only on the mathematics.

Any model, such as 5.4.2, that represents dynamic changes in vectors of probabilities is called a **Markov chain** model. The specific model we are examining is called the **Cantor-Jukes** model. Other models do not assume that all transitions occur with the same probability. We use Cantor-Jukes for our example because in an educational context it is better to omit relatively minor complications and emphasize the main ideas.

One additional difficulty is that α is not easy to measure. This turns out not to be very important, because our main interest is not in the number of generations, but rather in the total amount of genetic change. As we saw earlier, the total number of mutations per site over N generations is αN. This quantity is a measure of the total amount of genetic change from the ancestral strand to the contemporary strand. We define the **phylogenetic distance** d as the expected fraction of mutations per site from the ancestral sequence to the contemporary sequence; thus,

$$d = \alpha N \tag{5.4.4}$$

for the Cantor-Jukes model. Our problem is to determine this Cantor-Jukes distance from measured values of β.

Analysis of the DNA change model

There is an extensive literature on Markov chains, and they have many applications beyond biology. A thorough treatment of Markov chains requires more linear algebra than we have presented in this chapter. However, we have just enough linear algebra background for our specific problem.

Our analysis requires the following facts, which the reader can verify.

Theorem 5.4.1

$$\mathbf{v_1} = \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}, \qquad \mathbf{v_2} = \begin{pmatrix} 3/4 \\ -1/4 \\ -1/4 \\ -1/4 \end{pmatrix}$$

are eigenvectors of the matrix $\mathbf{M}$, *with corresponding eigenvalues* $\lambda_1 = 1$ *and* $\lambda_2 = 1 - \frac{4}{3}\alpha$.

Now define the vector $\mathbf{u}$ by

$$\mathbf{u} = 3\mathbf{M}^N(\mathbf{v_1} + \mathbf{v_2}).$$

We proceed to calculate $\mathbf{u}$ by two different methods.

Example 5.4.1
To calculate $\mathbf{u}$ for the Cantor-Jukes model, we can observe that $\mathbf{M}^N$ is the matrix that contains the overall transition probabilities for N generations. We know what this matrix is—each nucleotide has probability β of being different from generation 0 to generation N, with each possible change given equal probability. Thus,

$$\mathbf{M}^N = \begin{pmatrix} 1-\beta & \beta/3 & \beta/3 & \beta/3 \\ \beta/3 & 1-\beta & \beta/3 & \beta/3 \\ \beta/3 & \beta/3 & 1-\beta & \beta/3 \\ \beta/3 & \beta/3 & \beta/3 & 1-\beta \end{pmatrix}.$$

Therefore

$$\mathbf{u} = 3 \begin{pmatrix} 1-\beta & \beta/3 & \beta/3 & \beta/3 \\ \beta/3 & 1-\beta & \beta/3 & \beta/3 \\ \beta/3 & \beta/3 & 1-\beta & \beta/3 \\ \beta/3 & \beta/3 & \beta/3 & 1-\beta \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 3(1-\beta) \\ \beta \\ \beta \\ \beta \end{pmatrix}.$$

$\diamond$

Example 5.4.2

Another way to calculate $\mathbf{u}$ for the Cantor-Jukes model is to use the special relationship of eigenvalues and eigenvectors, namely the equations $\mathbf{M}\mathbf{v_1} = \lambda_1\mathbf{v_1} = \mathbf{v_1}$ and $\mathbf{M}\mathbf{v_2} = \lambda_2\mathbf{v_2}$. We have, in turn,

$$\mathbf{M}(\mathbf{v_1} + \mathbf{v_2}) = \mathbf{M}\mathbf{v_1} + \mathbf{M}\mathbf{v_2} = \mathbf{v_1} + \lambda_2\mathbf{v_2},$$

$$\mathbf{M^2}(\mathbf{v_1} + \mathbf{v_2}) = \mathbf{M}(\mathbf{M}(\mathbf{v_1} + \mathbf{v_2})) = \mathbf{M}(\mathbf{v_1} + \lambda_2\mathbf{v_2}) = \mathbf{M}\mathbf{v_1} + \lambda_2\mathbf{M}\mathbf{v_2} = \mathbf{v_1} + \lambda_2^2\mathbf{v_2},$$

$$\mathbf{M^3}(\mathbf{v_1} + \mathbf{v_2}) = \mathbf{M}(\mathbf{M^2}(\mathbf{v_1} + \mathbf{v_2})) = \cdots = \mathbf{v_1} + \lambda_2^3\mathbf{v_2},$$

and ultimately

$$\mathbf{u} = 3\mathbf{M}^N(\mathbf{v_1} + \mathbf{v_2}) = 3\mathbf{v_1} + 3\lambda_2^N\mathbf{v_2} = \begin{pmatrix} 3/4 + 9/4\lambda_2^N \\ 3/4 - 3/4\lambda_2^N \\ 3/4 - 3/4\lambda_2^N \\ 3/4 - 3/4\lambda_2^N \end{pmatrix}.$$

$\diamond$

Examples 5.4.1 and 5.4.2 use two different methods to calculate the same quantity. One is based on the expression of $\mathbf{M}^N$ in terms of the change probability β and the other uses the special relationship between eigenvalues and eigenvectors to avoid having to compute $\mathbf{M}^N$. Both must yield the same result, and thus

$$\beta = \frac{3}{4} - \frac{3}{4}\lambda_2^N = \frac{3}{4} - \frac{3}{4}\left(1 - \frac{4}{3}\alpha\right)^N. \tag{5.4.5}$$

We can solve this equation for N, with the elegant result

$$N = \frac{\ln\left(1 - \frac{4}{3}\beta\right)}{\ln\left(1 - \frac{4}{3}\alpha\right)} \tag{5.4.6}$$

Thus, the Cantor-Jukes distance is

$$d = \alpha\frac{\ln\left(1 - \frac{4}{3}\beta\right)}{\ln\left(1 - \frac{4}{3}\alpha\right)} = \frac{\alpha}{\ln\left(1 - \frac{4}{3}\alpha\right)}\ln\left(1 - \frac{4}{3}\beta\right). \tag{5.4.7}$$

The requirement of having to know α is somewhat restrictive, but we can use a linear approximation to eliminate α from the result. Consider that α is quite small because very few mutations occur in one generation. Hence, there is only a slight error introduced by replacing the nonlinear function $\ln\left(1 - \frac{4}{3}\alpha\right)$ by its linear approximation. Using calculus, we have

$$\ln\left(1 - \frac{4}{3}\alpha\right) \approx -\frac{4}{3}\alpha;$$

substituting this approximation into Equation 5.4.7 yields the result

$$d = -\frac{3}{4}\ln\left(1 - \frac{4}{3}\beta\right). \tag{5.4.8}$$

This formula allows scientists to convert a measured value of β into a number that represents phylogenetic distance according to the Cantor-Jukes model. Note that $d > 0$, in spite of the negative sign in the formula, because $0 < 1 - 4\beta/3 < 1$ for $0 < \beta < 3/4$.