Reference Readings & Homework Exercises for

Discrete Difference Equations

Written by

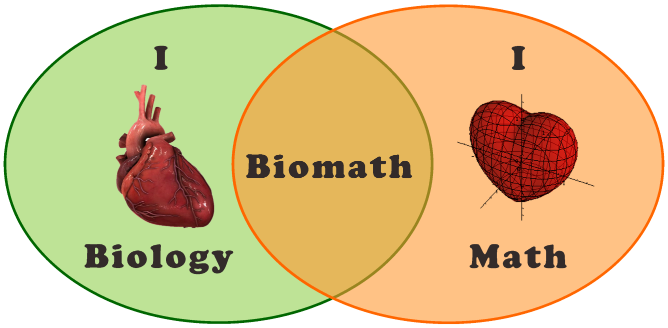
Prof. Erin N. Bodine

Prepared for

Math 214: Discrete Math Modeling with Biological Applications

Rhodes College

Last Updated: February 23, 2018



1 An Introduction to Discrete Mathematical Modeling

The field of mathematics provides many different means for modeling the world around us. Some mathematical tools are more appropriate for modeling certain phenomenon and less appropriate for other phenomenon. In this course we focus on how to use mathematics to simulate biological situations in which it is reasonable to assume the underlying variable of the model is discrete, that is, it can be written as a sequence of values. In contrast, the real number line is continuous and for any two values a and b (no matter how close a and b are on the number line), we can always find another value that is between a and b. We may have to increase our precision to achieve this, but it can always be done. A sequence of discrete values is a set of points along the real number line. Thus, with a discrete model, we are assuming that we can ignore the space between the points.

As one example, suppose we grow a cell culture in petri dish and measure the cell density (cells per unit area) every hour for 24 hours. This would create a sequence of 25 data points (if we include the initial cell density). We know that the cell culture is growing during the time between when we take our measurements, but measuring the population each hour over the course of a day will give us a reasonable representation of how the culture grows over the course of one day. Correspondingly, when we build a mathematical model to represent the growth of the cell culture, it is reasonable for the model to only predict how the culture grows from hour to hour. Thus, the underlying variable, time, is represented in the mathematical model as increasing in discrete, one hour increments.

For another example, consider a population which breeds seasonally, that is, only once a year. If we wanted to determine how the population size is changing over the course of many years, we could collect data to estimate the size of the population at the same time each year (say shortly after the breeding season ends). We know that between the times at which we estimate the population size some individuals will die and during the breeding season many new individuals will be born, however, we ignore the changes in population size from day to day, or week to week, and look only at how the population size changes from year to year. Thus, when we build a mathematical model of this population, it is reasonable for the model to only predict the population size for each year shortly after the breeding season. Here, the underlying variable, time, is represented in the mathematical model as increasing in discrete, one year increments.

2 A Motivating Example: Growth of an *E. coli* colony

Let us consider an actual data set for the cell growth example described above. The data in Table 1 gives the optical density of an *Escherichia coli* (commonly referred to as *E. coli*) cell culture measured every 30 minutes for 3 hours. The culture was grown in a nutrient solution at 37°C.

Table 1: Optical Density of *E. coli* cell culture.

Time (hours)	0.0	0.5	1.0	1.5	2.0	2.5	3.0
Cell Density (OD_{600})	0.055	0.120	0.231	0.360	0.516	0.821	1.300

***Escherichia coli*:** Abbreviated as *E. coli*, this bacteria lives in the lower intestines of endotherms (warm blooded organisms), including humans. There are variety of strains of *E. coli* many of which are part of the natural microbiome of an endotherm's intestinal tract. However, there are some strains of *E. coli* which can cause severe and adverse reactions in their hosts including vomiting, fever, aches, and diarrhea (in humans).

If we plot the data in Table 1 we obtain the graph in Figure 1. When we examine this data it would appear that the cell density is growing exponentially. How might we approximate this sequence of cell densities using an algebraic expression, that is, a mathematical model?

In general, when we want to describe a sequence mathematically, we want to express how the $(n+1)^{th}$ term in the sequence, x_{n+1} is determined using previous n terms in the sequence, $x_n,\ x_{n-1},\ x_{n-2},\ldots,\ x_1,\ x_0$. Mathematically, we express the term x_{n+1} as a function of the previous n terms, that is

$$x_{n+1} = f(x_n,\ x_{n-1},\ x_{n-2},\ldots,\ x_1,\ x_0). \tag{1}$$

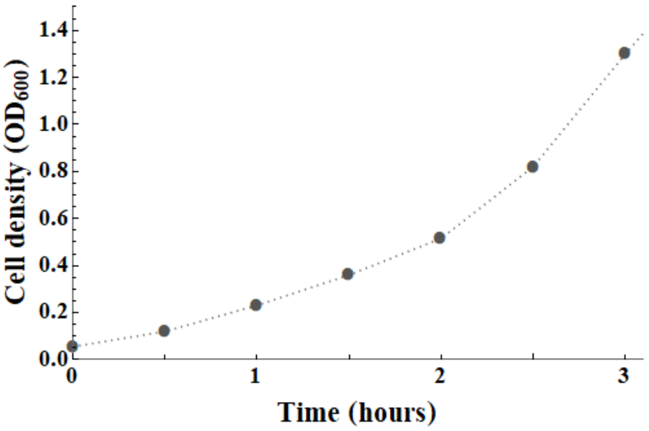


Figure 1: Optical density of the *E. coli* cell culture measured every 30 minutes over 3 hours.

Equations of the form of Equation (1) are known as ***discrete difference equations***.

Optical Density: How does one "count" the number of cells in a cell culture? The number of cells in a culture could be quite large (on the order of millions or billions). It would indeed be a tedious task to sit down and count the total number of cells in such a culture. In fact, if the culture is growing (the number of cells are increasing), by the time you counted the first 200 cells, the size of culture would likely have changed. Instead of counting the number of cells in a culture of volume V, we sample the culture, that is, we take a representative portion of the culture and count the number of cells in the sample. Suppose the sample has N_s cells, and a volume of V_s. If we assume that the cell density (cells per unit volume) are the same in the whole culture as in the sample, then $N/V = N_s/V_s$, where N is the total number of cells in the sample. Thus, $N = V\,(N_s/V_s)$. The ratio N_s/V_s is proportional to the optical density (i.e. $N_s/V_s = a \cdot OD$ where a is a constant and OD is the optical density) and is measured using a spectrophotometer. The sample is placed in a cuvette (essentially a rectangular test tube) and the spectrophotometer shines a beam of light (of a specified wavelength) through the cuvette. The denser the cell culture, the less light will shine through the cuvette. By measuring the amount of light that does not make it through the cuvette (light that scatters due to the cells in the cuvette), the spectrophotometer measures the optical density of the cell culture sample. The intensity of the light scattering, and thus the measure of optical density, depends on the wavelength of light being used. The wavelength of light used, measured in nanometers (nm), is denoted by a subscript of OD. Thus, an optical density measure taken using a light with wavelength 600 nm is indicated by OD$_{600}$.

For simplicity, let us assume that the next value in the cell density sequence can be determined using only the previous value in the sequence. Thus, x_{n+1} is a function of x_n, that is $x_{n+1} = f\left(x_n\right)$, which is called a ***first order difference equation***. Often we can determine the relationship between two consecutive terms in a sequence, x_n and x_{n+1}, by examining either the difference or ratio between those two terms, that is $x_{n+1} - x_n$ or x_{n+1}/x_n, respectively. For the data presented in Table 1, let x_n represent the cell density of the n^{th} data point, where $n = 0$ denotes the data point at 0 hours, $n = 1$ denotes the data point at 30 minutes, etc. Thus,

$$x_0 = 0.055,\ x_1 = 0.120,\ x_2 = 0.231,\ x_3 = 0.360,\ x_4 = 0.561,\ x_5 = 0.821,\ x_6 = 1.300, \tag{2}$$

where the underlying variable, time, is represented as increasing in discrete 30 minute increments. Table 2 shows both the difference and ratio of consecutive terms in the sequence representing the cell density data from Table 1. The difference between consecutive terms in the sequence are increasing as n increases. However, with the exception of the first two ratios, the ratios of consecutive terms appears to remain relatively constant as n increases. Suppose the ratios of consecutive terms were constant with a value c as n increased, then

$$\frac{x_{n+1}}{x_n} = c \quad \Rightarrow \quad x_{n+1} = c x_n,$$

and thus we have written x_{n+1} as a function of x_n (where $f(x_n) = c x_n$). Given that the ratios of consecutive terms show little fluctuation, we might assume that the fluctuations we see are due to measurement error (either by the spectrophotometer used to measure the optical density or by the human taking the measurements). If we assume that the ratios of consecutive terms would be constant if not for measurement error, then what would be the value of c? We could use the average (mean value, or more specifically the arithmetic mean) of all of the ratios: 1.712. We could also use the geometric mean of all of the ratios: 1.694. As a third option, we could use the median of all of the ratios: 1.585. The arithmetic mean, geometric mean, and median are all ***measures of central tendency*** which provide different ways of describing the "center" of a set of values, in this case the center of the ratio values. Thus, we have three options for our model:

Table 2: Differences & ratios of consecutive terms of Sequence (2).

n	$x_{n+1} - x_n$	x_{n+1}/x_n
0	0.065	2.18
1	0.111	1.93
2	0.129	1.56
3	0.156	1.43
4	0.305	1.59
5	0.479	1.58

$$x_{n+1} = 1.712 x_n \text{ (model 1)}, \quad x_{n+1} = 1.693 x_n \text{ (model 2)} \quad \text{and} \quad x_{n+1} = 1.585 x_n \text{ (model 3)}.$$

Arithmetic vs. Geometric Mean In general, mean or average values are a way to describe the "center" of a set of values. Statisticians refer to mean values as ***measures of central tendency***. Other measures of central tendency are the median and the mode.

The ***arithmetic mean*** is what we typically think of as the average value of a set of numbers, that is we take the sum of a set of values and divide by the number of values in the set. In mathematical notation we denote the arithmetic mean of a set of n values $x_1, x_2, \ldots, x_n$ as

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

The ***geometric mean*** is an alternative method for calculating a mean or average value. Instead of taking a sum of the values, a product of the values is used. Specifically, we take the product of a set of n values and then take the n^{th} root of that product. In mathematical notation we denote the geometric mean of a set of n values $x_1, x_2, \ldots, x_n$ as

$$\left(\prod_{i=1}^{n} x_i\right)^{1/n} = \sqrt[n]{x_1\, x_2\, \cdots\, x_n}.$$

The geometric mean is the more appropriate average value to use when you are calculating the "center" of a set of proportional data like exponential growth rates.

How do we decide which model is "best"? First, what does it mean for one model to be better than another model? In this case, we are interested in seeing which model is best able to accurately reproduce the data. When we use model 1 with $x_0 = 0.055$, we generate the sequence

$$x_0 = 0.055,\ x_1 = 0.094,\ x_2 = 0.161,\ x_3 = 0.276,\ x_4 = 0.472,\ x_5 = 0.809,\ x_6 = 1.385. \tag{3}$$

When we use model 2 with $x_0 = 0.055$, we generate the sequence

$$x_0 = 0.055,\ x_1 = 0.093,\ x_2 = 0.158,\ x_3 = 0.267,\ x_4 = 0.452,\ x_5 = 0.765,\ x_6 = 1.295. \tag{4}$$

When we use model 3 with $x_0 = 0.055$, we generate the sequence

$$x_0 = 0.055,\ x_1 = 0.087,\ x_2 = 0.138,\ x_3 = 0.219,\ x_4 = 0.347,\ x_5 = 0.550,\ x_6 = 0.872. \tag{5}$$

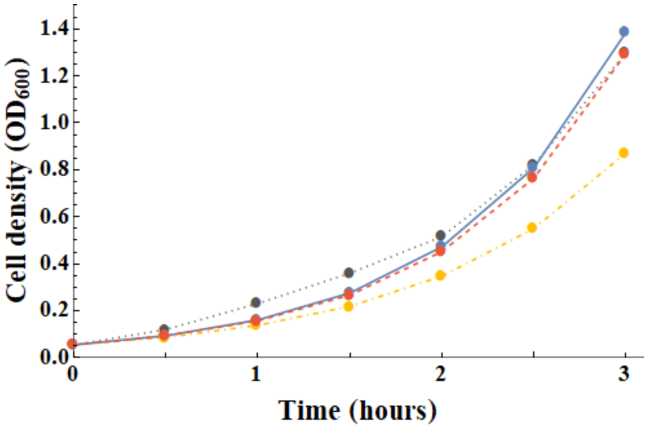


Figure 2: Optical density of the *E. coli* cell culture measured every 30 minutes over 3 hours (dotted grey), and estimated optical density using model 1 (solid blue), model 2 (dashed red), and model 3 (dot-dash yellow).

Figure 2 show a plot of the sequence of the original data (gray) against the sequences generated by model 1 (blue), model 2 (red), and model 3 (yellow). Which model appears to approximate the data best? From the Figure 2 it is easy to see that model 3 (using the median value of the ratios to approximate the value of c) is not approximating the data as well as the other two models. However, between model 1 and model 2, it is difficult to tell which model is better at approximating the data. Thus, we need a way to quantify, or to measure, which model is better at approximating the data.

If we want to quantify which model is "better", we need a criteria for what it means for one model to be "better" than another. One criteria we could use is the better model will have smaller differences between the model values and the data points at each time step. Let $\hat{x}_n$ represent the value of the cell density data at time step n, and x_n represent the value of the cell density at time step n predicted by the model. Then we define the error between the data and the model as

$$\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 . \tag{6}$$

Then, model 1 has an error of 0.02194, model 2 has an error of 0.02196, and model 3 has an error of 0.31481. Notice that the error for model 3 is significantly larger than the error for the other two models which corroborates the analysis from our visual inspection of Figure 2. The error for models 1 and 2 are very close, though model 2 has a slightly higher error than model 1, making model 1 the "best" model given our criteria.

3 Discrete Exponential Growth Model

The equation $x_{n+1} = ax_n$ is known as the ***discrete exponential growth model***, and sequences generated by functions of the form $x_{n+1} = ax_n$ are known as ***geometric sequences***. Given a value of x_0, then

$$\begin{aligned} x_1 &= ax_0 \\ x_2 &= ax_1 = a(ax_0) = a^2x_0 \\ x_3 &= ax_2 = a(a^2x_0) = a^3x_0 \\ &\vdots \\ x_n &= ax_{n-1} = a^nx_0. \end{aligned}$$

The equation

$$x_n = a^n x_0 \tag{7}$$

is known as the ***general solution*** or ***closed form solution*** of the equation $x_{n+1} = ax_n$. Note, for Equation 7, if you know the initial value of the sequence, x_0, you are able to determine the value of the n^{th} term of the sequence without iterating through all n terms (which could be tedious for large n).

Suppose $a > 1$, then $x_{n+1} = ax_n > x_n$. This means that when $a > 1$ the $(n+1)^{th}$ term in the sequence is always larger than the term that came before it, the n^{th} term. We describe such a sequence as an ***increasing sequence***. In the example where x_n represents the cell density of the *E. coli* population at time step n, $a > 1$ implies that the cell culture is growing; specifically, it is becoming more dense. By examining the closed form solution $x_{n+1} = a^n x_0$, we can see that as n becomes increasingly large (denoted by $n \to \infty$), the value of $a^n \to \infty$ if $a > 1$. Thus, as we allow n to increase towards infinity, the values of the sequence always increase towards infinity. This makes intuitive sense since this discrete difference equation models exponential growth.

If $a < 1$, then $x_{n+1} = ax_n < x_n$ which means the $(n+1)^{th}$ term in the sequence is always smaller than the term that came before it, the n^{th} term, and thus we have a ***decreasing sequence***. By examining the close form solution $x_{n+1} = a^n x_0$, we can see that as $n \to \infty$, $a^n \to 0$ when $a < 1$. This also makes intuitive sense since the amount of x is being reduced by some proportion with each increment of n. Note, in the special case when $a = 1$, $x_{n+1} = x_n$ for all n, and thus every value in the sequence is the same.

Let $a = 1 + r$, then the discrete exponential growth model becomes

$$x_{n+1} = (1 + r)x_n = x_n + rx_n. \tag{8}$$

The red term, x_n, represents the amount of x at time step n, and the blue term, rx_n, represents a proportion of x at time step n. Thus, the amount of x at time step $n + 1$ is the amount of x at time step n plus a proportion of x at time step n. Notice if $r < 0$ (which corresponds to $a < 1$), then we will be subtracting a proportion of x at time step n.

If, for example, we knew a population was increasing 3% each year, we could represent the population size as

$$x_{n+1} = x_n + 0.03x_n = 1.03x_n,$$

where x_n represents the population size at year n. Note that this difference equation produces an increasing sequence and thus, $x_n \to \infty$ as $n \to \infty$. If, on the other hand, the population was decreasing by 3% each year, then we would represent the population size as

$$x_{n+1} = x_n - 0.03x_n = 0.97x_n,$$

where x_n represents the population size at year n. This difference equation produces a decreasing sequence and thus, $x_n \to 0$ as $n \to \infty$.

Note that Equation (8) can be rewritten as

$$x_{n+1} - x_n = rx_n, \tag{9}$$

which denotes the difference in the amount of x from time n to time $n+1$. The equation in this form reveals why we refer to these types of equations as difference equations. Note, the change in x over one time step is a proportion (r) of the amount at the previous time step (x_n). Thus, we say "the change is x is *proportional* to the amount of x." Furthermore, if we rewrite Equation (9) as

$$\frac{x_{n+1} - x_n}{x_n} = r, \tag{10}$$

then we have a direct way to estimate the value of r for a model. Table 3 shows the values of the difference quotients in Equation (10) for $n = 0, 1, \ldots, 5$. We can approximate a value for r by taking the arithmetic mean, geometric mean, or median of the values in the second column of Table 3, which are 0.712, 0.671, and 0.578, respectively. Since $a = 1 + r$, these correspond to values of $a = 1.712$ (exactly the same as in model 1), $a = 1.671$ (similar as in model 2), and $a = 1.578$ (similar as in model 3), respectively.

Table 3: Difference quotients of consecutive terms of Sequence (2).

n	$(x_{n+1} - x_n)/x_n$
0	1.182
1	0.925
2	0.558
3	0.433
4	0.591
5	0.583

Example 1.

We determined that a reasonable model for the density of *E. coli* cells given the data presented in Table 1 is

$$x_{n+1} = 1.712x_n.$$

(a) Find the general solution to this difference equation.

(b) Does this difference equation model produce an increasing or decreasing sequence?

(c) Given this model, what cell density would we predict at 6 hours? at 10 hours? at 24 hours? Do these values make biological sense?

(d) What happens to the population of *E. coli* cells as $n \to \infty$? Is this biologically reasonable?

The solutions to this example will be worked in class.

Example 2.

The half-life of caffeine in the human body is about 6 hours, thus every 6 hours the amount of caffeine the body decreases by half (assuming no other intake of caffeine). Suppose you consume a cup of coffee with 100 mg of caffeine at 7 am.

(a) Use a discrete difference equation model to express how much caffeine is in the body over time. How much time does each time step n represent? Express the model in closed form solution.

(b) Does the model generate an increasing or decreasing sequence? Does this make biological sense?

(c) How much caffeine is left in the body at 1 pm? at 7 pm? at 1 am?

(d) Suppose you drank two cups of coffee (200 mg) at 7 am instead of one cup of coffee. How much caffeine is left in the body at 1 pm? at 7 pm? at 1 am?

(e) More accurately, the half-life of caffeine in the human body is 5.7 hours [8] and the amount of caffeine in the average 8 oz cup of coffee is 95 mg. How would you change the model in (a) to reflect this?

The solutions to this example will be worked in class.

Example 3.

The Forest and Wildlife Research Center at Mississippi State University measured the survival and mortality rates of bobcats (*Lynx rufus*) in and near the Tallahala Wildlife Management Area through their Carnivore Ecology Research Project [9]. They studied 68 bobcats (28 males and 40 females) and found that the mean annual survival was 75% for males and 84% for females.

(a) Use a discrete difference equation model to express how many male bobcats (of the original 28) are still alive at year n. Assume that there are 28 male bobcats at year $n = 0$. Express the model in closed form solution.

(b) Use a discrete difference equation model to express how many female bobcats (of the original 40) are still alive at year n. Assume that there are 40 female bobcats are year $n = 0$. Express the model in closed form solution.

(c) Both models generate decreasing sequences. Which is decreasing more quickly?

(d) What is the first year in which there will be no male bobcats left (i.e. less than 1 male bobcat) of the original 28? What is the first year in which there will be no female bobcats left of the original 40?

The solutions to this example will be worked in class.

Exercise 1 – Wild Hares

Suppose a population of wild hares increases by 12% each year, and there are currently 200 hares. Let x_n be the number of hares in the population at year n.

(a) Find the difference equation representing x_{n+1} as a function of x_n.

(b) Find the general solution to the difference equation found in (a).

(c) Find the number of hares in the population 6 years from now.

(d) What happens to the population of hares as $n \to \infty$? Is this biologically reasonable?

Exercise 2 – Elk Herd

Let x_n be the number of adult elk in a newly introduced elk herd in a wildlife management area in the Cumberland Mountains in Tennesseee. There are currently 50 adult elk in the population.

(a) Suppose the number of adults in the population increases by 10% each year. Use a discrete difference equation model to express how the number of adult elk in the population at year n (express your model in the closed form solution). Does your model generate an increasing or decreasing sequence? Determine how many adult elk are in the population after 20 years.

(b) Suppose that the probability that an adult elk survives in a given year is 90%. Use a discrete difference equation model to express how many of the originally introduced adult still alive at year n (express your model in the close form solution). Does your model generate an increasing or decreasing sequence? Determine how many of the originally introduced elk are still alive after 20 years. Elk live, on average, 15 years in the wild. Does your answer make biological sense?

Exercise 3 – Treating Heart Disease

Digoxin is used in the treatment of heart disease. Doctors must prescribe an amount of medicine that keeps the concentration of digoxin in the bloodstream above a minimum level of effectiveness without exceeding the maximum level for safety (these levels vary among patients). For an initial dosage of 0.5 mg in the bloodstream, the table below shows the amount of digoxin (x) remaining in the bloodstream of a particular patient after n days.

Time (days)	0	1	2	3	4	5	6	7	8
Digoxin (mg)	0.500	0.345	0.238	0.164	0.113	0.078	0.054	0.037	0.026

(a) Suppose the amount of digoxin in the bloodstream of the patient can be modeled using $x_{n+1} = ax_n$, where x_n is the amount of digoxin (in mg) in the bloodstream on day n. Use the ratio of consecutive terms in the sequence to determine the value of a.

(b) Express your model in its closed form solution.

(c) In this patient the digoxin is effective if it remains above 0.01 mg, however digoxin toxicity occurs when the amount of digoxin in the bloodstream exceeds 0.6 mg. When would you recommend this patient take his or her next dose of digoxin.
Note, we will revisit this example and the idea of timing between doses once we have explored some more sophisticated difference equation models.

4 Discrete Model with Exponential & Constant Growth

In the discrete exponential growth model, the change in x (represented as $x_{n+1} - x_n$) was proportional to x (that is x_n), thus we had the equation

$$x_{n+1} - x_n = rx_n,$$

where r is the proportionality constant which represents the rate of change. However, what if the change is not proportional, but constant, that is what if

$$x_{n+1} - x_n = b, \tag{11}$$

where b is a constant. Equation (11) can be rewritten as

$$x_{n+1} = x_n + b, \tag{12}$$

and we will refer to this model as the ***constant growth model***. The sequence generated by Equation (12) is known as an ***arithmetic sequence***. As with the discrete exponential growth model, we can find a closed form solution of the constant growth model. Given a value for x_0, then

$$\begin{aligned}
x_1 &= x_0 + b \\
x_2 &= x_1 + b = (x_0 + b) + b = x_0 + 2b \\
x_3 &= x_2 + b = (x_0 + 2b) + b = x_0 + 3b \\
&\vdots \\
x_n &= x_0 + nb.
\end{aligned}$$

Note, if $b > 0$ then the values of x_n increase as $n \to \infty$, but if $b < 0$ then the value of x_n decrease as $n \to \infty$. In particular, the values of x_n can become negative if $b < 0$ and n is large enough, even if $x_0 > 0$. This may or may not be biologically reasonable depending on what is being modeled.

In general, we will not use the constant growth model on its own, but in combination with the exponential growth model, that is

$$x_{n+1} - x_n = \underbrace{rx_n}_{\substack{\text{exponential}\\\text{growth}}} + \underbrace{b}_{\substack{\text{constant}\\\text{growth}}}. \tag{13}$$

We can rearrange Equation (13) to obtain

$$x_{n+1} = (r+1)x_n + b. \tag{14}$$

In this section we will examine a variety of models (known as ***first order linear difference equations***) of the form

$$x_{n+1} = \underbrace{ax_n}_{\substack{\text{exponential}\\\text{growth}}} + \underbrace{b}_{\substack{\text{constant}\\\text{growth}}}, \tag{15}$$

where $a = r+1$ from Equation (14). Notice if $a = 1$, then Equation (15) reduces to the constant growth model represented by Equation (12).

Before we look at specific biological phenomenon which can be modeled using Equation (15), we shall derive the closed form solution (see below) and discuss the notion of the model begin at equilibrium and how to determine the long term behavior of a model (see the next section). Given x_0, then the closed form solution is derived as

$$\begin{aligned} x_1 &= ax_0 + b \\ x_2 &= ax_1 + b = a\left[ax_0 + b\right] + b = a^2x_0 + (a+1)b \\ x_3 &= ax_2 + b = a\left[a^2x_0 + (a+1)b\right] = a^3x_0 + \left(a^2 + a + 1\right)b \\ &\vdots \\ x_n &= a^nx_0 + \left(a^{n-1} + \cdots + a^2 + a + 1\right)b. \end{aligned}$$

This is a way to express the closed form solution, but it seems rather unwieldy. It would be nice if that $(n-1)^{th}$ order polynomial in terms of a weren't mucking up the equation. With some good old fashioned polynomial long division (think back to your fondest memories of Algebra II), it can be shown that

$$1 - a^n = (1-a)\left(1 + a + a^2 + \cdots + a^{n-1}\right) \quad \Rightarrow \quad \frac{1-a^n}{1-a} = 1 + a + a^2 + \cdots + a^{n-1},$$

provided $a \neq 1$. Thus, we can rewrite the closed form solution as

$$x_n = a^nx_0 + \frac{1-a^n}{1-a}b, \tag{16}$$

which feels much more manageable as an equation, and will be easier to work with as we progress forward. However, we can rearrange Equation (16) to make it easier to remember and work with,

$$x_n = a^nx_0 + (1-a^n)\frac{b}{1-a} = a^nx_0 - a^n\frac{b}{1-a} + \frac{b}{1-a} = a^n\left(x_0 - \frac{b}{1-a}\right) + \frac{b}{1-a}.$$

If we let $x^* = \dfrac{b}{1-a}$, then the rearranged close form solution becomes

$$\boxed{x_n = a^n\left(x_0 - x^*\right) + x^*} \tag{17}$$

We will see in the next section that the quantity $x^* = b/(1-a)$ has particular meaning as we analyze the long term behavior of the $x_{n+1} = ax_n + b$ model (i.e. as n gets large).

Example 4.

Recall that the half-life of caffeine in the human body is approximately 6 hours. Suppose you drink 2 cups of coffee (200 mg of caffeine) each morning at 8 am.

(a) Propose a model of the form $x_{n+1} = ax_n + b$ to represent the amount of caffeine in your body each morning at 8 am. What is the value of a? What is the value of b? How much time does one time step n represent?

(b) Compute the values of the terms x_1, x_2, x_3, and x_4. Describe the pattern you see in the terms of the sequence x_n.

(c) Write the Matlab code that would generate the first 10 terms in the sequence determined by the model proposed in (a).

(d) What is the closed form solution of the model proposed in (a)?

(e) Using the closed form solution determine what happens to the terms x_n as $n \to \infty$. Describe what this means biologically.

The solutions to this example will be worked in class.

Exercise 4 – Dixogin Dosing

Recall that digoxin is used in the treatment of heart disease. Suppose a patient takes a dose of 0.5 mg of dixogin every day at the same time, and that each day at the time the next dose is taken 69% of the amount of dixogin in the body from the previous day remains.

(a) Propose a model of the form $x_{n+1} = ax_n + b$, where x_n represents the amount of dixogin in the patient's body on the n^{th} day of taking the prescribed medication. What is the value of a? What is the value of b? How much time does one time step n represent?

(b) Compute the first 5 terms in the sequence.

(c) Express your model in its closed form solution.

(d) Using the closed form solution determine what happens to the terms x_n as $n \to \infty$. Describe what this means biologically.

5 Equilibria & Long-Term Behavior

Equilibria

The term equilibrium refers to being in a state of balance and remaining as you are. In the case of discrete difference equation models, a number x^* is called an ***equilibrium point*** or ***fixed point*** of the difference equation $x_{n+1} = f(x_n)$ if $x^* = f(x^*)$. Let us examine this concept for the few discrete difference equation models we have encountered so far.

EXPONENTIAL GROWTH MODEL · Recall the exponential growth model has the form $x_{n+1} = ax_n$, where a is a constant. The model will be in equilibrium when $x^* = ax^*$ which can only occur if $a = 1$. Thus, when $a \neq 1$, there is no fixed point for the model which makes sense because the values of x_n are either always increasing ($a > 1$) or always decreasing ($a < 1$). When $a = 1$, $x_n = x_0$ for all n.

CONSTANT GROWTH MODEL · Recall the constant growth model has the form $x_{n+1} = x_n + b$, where b is a constant. The model will be in equilibrium when $x^* = x^* + b$ which can only occur if $b = 0$. Thus, when $b \neq 0$, there is no fixed point for the model which make sense because the values of x_n are either always increasing ($b > 0$) or always decreasing ($b < 0$). When $b = 0$, $x_n = x_0$ for all n.

EXPONENTIAL & CONSTANT GROWTH MODEL · Recall the difference equation model with exponential and constant growth has the form $x_{n+1} = ax_n + b$, where a and b are constants. The model will be in equilibrium when $x^* = ax^* + b$, then $(1-a)x^* = b$, and thus the equilibrium point is

$$\boxed{x^* = \frac{b}{1-a}} \tag{18}$$

This is different from the previous two examples because this does not require that a and b take on certain values for the model to have a fixed point, and the fixed point is not necessarily the initial condition x_0. When we examine this model as it pertains to different biological phenomenon, we will see the different ways in which we can interpret the meaning of the equilibrium value x^*.

Long-Term Behavior

When you are examining the features of a model $x_{n+1} = f(x_n)$, it if often useful to know what happens to the model as $n \to \infty$. This is known as the ***long-term behavior*** of the model. We already examined the long term behavior of the exponential growth model and constant growth model under various parameter scenarios ($a > 1$, $a = 1$, $a < 1$ for the exponential growth model, and $b > 0$, $b = 0$, $b < 0$ for the constant growth model).

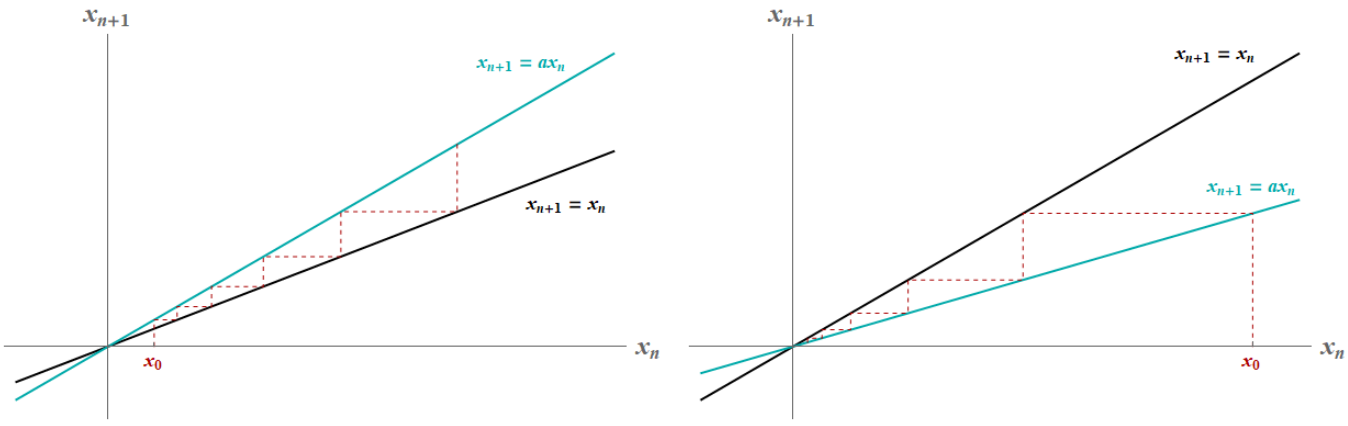
It turns out there is another, more visual way to examine the long-term behavior of these models called ***cobweb diagram analysis***. On the x_n, x_{n+1} plane (where x_n is the horizontal axis and x_{n+1} is the vertical axis), we plot $x_{n+1} = f(x_n)$ and $x_{n+1} = x_n$. On the x_n (horizontal) axis, locate x_0. Draw a vertical line from $(x_0, 0)$ on the horizontal axis to (x_0, x_1) on the $x_{n+1} = f(x_n)$ curve. Your height on the vertical axis now represents x_1. Then, starting with $n = 0$

1. Draw a horizontal line from (x_n, x_{n+1}) on the curve $x_{n+1} = f(x_n)$ to (x_{n+1}, x_{n+1}) on the line $x_{n+1} = x_n$.
2. Draw a vertical line from (x_{n+1}, x_{n+1}) on the line $x_{n+1} = x_n$ to (x_{n+1}, x_{n+2}) on the curve $x_{n+1} = f(x_n)$.
3. Repeat steps 1 & 2 with increasing values of n as many times as needed to see a pattern.

Figure 3 shows what this looks like for the discrete exponential growth model when (a) $a > 1$ and (b) $a < 1$. In each graph the blue curve represents $x_{n+1} = ax_n$ and the black line represents $x_{n+1} = x_n$. The dotted red line show the "cobwebbing". In Figure 3a (where $a > 1$) we can see that the cobwebbing (dotted red line) steps up and away from x_0 as n increases. This is because the values of x_n (and thus x_{n+1}) are increasing as n increases. In Figure 3b (where $0 < a < 1$) we can see that the cobwebbing (dotted red line) steps down towards $(0, 0)$ as n increases.

This method may seem like overkill for these simple difference equations but it will be very handy for analyzing the long-term behavior of $x_{n+1} = ax_n + b$.

Note that for the difference equation model with exponential and constant growth, $x_{n+1} = ax_n + b$, it is a little harder to determine the long term behavior because their are two parameters, a and b, and the relative size and sign of each parameter will have a bearing on what happens to the values of x_n as $n \to \infty$. Suppose $x_0 > 0$. If $a > 1$ and $b > 0$ we would expect the population to grow, and if $a < 1$ and $b < 0$ we would expect the population to decline. And indeed, in both cases we are correct as shown in the cobweb diagrams depicted in Figures 4a and 4b. Notice that in the case where $a < 1$ and $b < 0$, the values of x_n quickly become negative despite the fact that $x_0 > 0$.



(a) $a > 1$ **(b)** $a < 1$

Figure 3: Cobweb diagrams for the discrete exponential growth model for (a) $a > 1$ and (b) $a < 1$.

Now, suppose $a < 1$ and $b > 0$. What do we expect to happen in this case. The value of a will cause a decrease while the value of b will cause an increase. If we examine the cobweb diagram in Figure 4c we see that the lines $x_{n+1} = x_n$ and $x_{n+1} = ax_n + b$ cross in the first quadrant at the point $\left(\frac{b}{1-a}, \frac{b}{1-a}\right)$. To determine where the lines intersect set the right-hand side of both equations equal to each other and solve for x_n, i.e.

$$ax_n + b = x_n \quad \Rightarrow \quad b = (1-a)x_n \quad \Rightarrow \quad x_n = \frac{b}{1-a}.$$

Notice if we choose $x_0 = \frac{b}{1-a}$, then

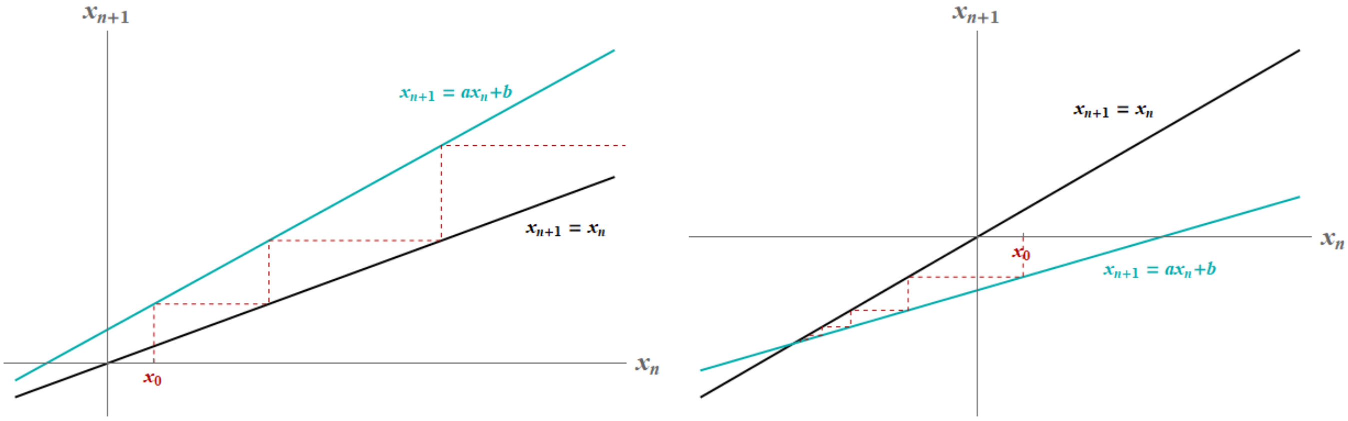
$$\begin{aligned} x_1 &= ax_0 + b = a\left(\frac{b}{1-a}\right) + b = \frac{ab + b - ab}{1-a} = \frac{b}{1-a} \\ x_2 &= ax_1 + b = a\left(\frac{b}{1-a}\right) + b = \frac{ab + b - ab}{1-a} = \frac{b}{1-a} \\ &\vdots \\ x_n &= ax_{n-1} + b = a\left(\frac{b}{1-a}\right) + b = \frac{ab + b - ab}{1-a} = \frac{b}{1-a}, \end{aligned}$$

and thus $x_n = \frac{b}{1-a}$ for all n which means $x^* = \frac{b}{1-a}$ is an equilibrium point. If we choose x_0 such that $0 < x_0 < \frac{b}{1-a}$, then the values of x_n increase towards $\frac{b}{1-a}$. If we choose $x_0 > \frac{b}{1-a}$, then the values of x_n decrease towards $\frac{b}{1-a}$. Thus, no matter what value of x_0 we choose, as $n \to \infty$ the values of x_n approach $\frac{b}{1-a}$. Since the sequence of values approach the equilibrium point, we classify the fixed point as a ***stable equilibrium***.

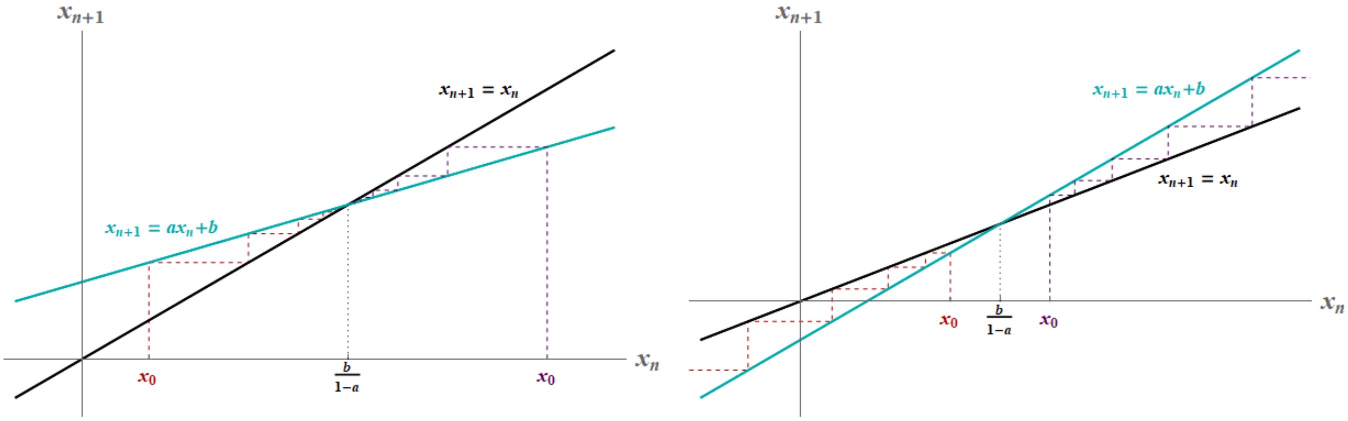
Lastly, suppose $a > 1$ and $b < 0$. If we examine the cobweb diagram in Figure 4d we see again that the lines $x_{n+1} = x_n$ and $x_{n+1} = ax_n + b$ cross in the first quadrant at the point $\left(\frac{b}{1-a}, \frac{b}{1-a}\right)$, and that if we choose $x_0 = \frac{b}{1-a}$, then $x_n = \frac{b}{1-a}$ for all n, and thus $x^* = \frac{b}{1-a}$ is an equilibrium point. However, if we choose x_0 such that $0 < x_0 < \frac{b}{1-a}$, then the values of x_n decrease towards $-\infty$. If we choose $x_0 > \frac{b}{1-a}$, then the values of x_n increase towards ∞. In this case, not matter what value of x_0 we choose, with the exception of $x_0 = \frac{b}{1-a}$, as $n \to \infty$ the values of x_n move away from $\frac{b}{1-a}$, and we classify this type of fix point as an ***unstable equilibrium***.

Example 5.

The half-life of caffeine in the human body is approximately 6 hours. Suppose you drink 2 cups of coffee (200 mg of caffeine) each morning



(a) $a > 1$, $b > 0$ **(b)** $a < 1$, $b < 0$



(c) $a < 1$, $b > 0$ **(d)** $a > 1$, $b < 0$

Figure 4: Cobweb diagrams for $x_{n+1} = ax_n + b$.

at 8 am. Let x_n be the amount of caffeine in the your body at 8 am on the n^{th} morning after you start drinking 2 cups of coffee each morning at 8 am. Then the difference equation model $x_{n+1} = 0.0625x_n + 200$, $x_0 = 200$ generates the terms of the sequence x_n.

What is the equilibrium point of this model? Is the equilibrium point stable or unstable?

The solutions to this example will be worked in class.

Example 6.

Consider a lake fish population which increases on average by 20% annually. Each year, fishing on the lake is allowed until 1200 fish are caught. Thereafter, fishing is banned. Currently, there are 12,230 fish in the lake.

(a) Write a difference equation for the lake fish population and find the general solution.

(b) How many fish are in the lake after 5 years?

(c) What is the equilibrium point of the model? Is it stable or unstable?

(d) If the resource managers of the lake wanted the population to remain constant each year, what level of harvesting should they allow? Note, this value is referred to as the ***maximum sustainable yield (MYS)*** .

The solutions to this example will be worked in class.

Exercise 5 – Stocking a Lake

A fish population is increasing by 80% per year. Each year, fishing on the lake is allowed until 1200 fish are caught, thereafter, fishing is banned. Currently, there are 12,230 fish in the lake. If the resource managers of the lake wish to keep the fish population constant from year to year, how many fish are needed to stock the lake each year?

Exercise 6 – Buffalo

A population of buffalo increases in size by about 10% each year. Let x_n be the population count after n years. Suppose that hunting allows h buffalo to be removed from the herd each year.

(a) Find an expression for x_n in terms of n if $x_0 = 1000$ and $h = 20$.

(b) Will terms in the sequence generated by the expression you found in (a) ever be negative?

(c) Determine a value for h such that the population with $x_0 = 1000$ remains at equilibrium.

Exercise 7 – Trout

Suppose that in a trout population increases its own numbers by 10% each year. After the births occur each year, 100 young trout are added to the population in an effort to build up the population. Let x_n denote the size of the population after n years, and assume $x_0 = 1000$.

(a) For what n does $x_n \geq 2000$ first occur?

(b) Suppose that after the time found in part (a) the lake is not longer stocked. How many trout should the lake managers allow to be caught each year if they want the trout population to remain stable (i.e., at equilibrium). *Hint: Let create a new first order linear difference equation where x_0 is the size of the population at the value of n found in part (a) rounded to the nearest whole fish.*

6 Drug Dosing & Pharmacokinetics

Pharmacokinetics is a branch of pharmacology which quantifies how the concentration of a substance administered to a living organism declines over time (from the moment it is introduced to the body/organism to the point at which it is entirely eliminated from the body). You have already seen one example of pharmacokinetics in Exercise 3 where you expressed the decay of digoxin (a drug used in treated heart disease) as a discrete exponential decay model. In that example, only once dose was given. In this section we examine what occurs to the concentration of a drug in the body when we give successive doses of the same drug over some period of time.

For many drugs administered intravenously, the amount of drug in the body after a dose of size b administered at time $t = 0$ is

$$x(t) = be^{-kt}, \tag{19}$$

where k is a decay rate specific to the drug being administered, and the larger the value of k the more quickly the drug is eliminated from the body (see Figure 5). More specifically, if t^* represents the time it take for the half of the initially administered dose to be eliminated from the body (i.e., $b/2$, also known as the

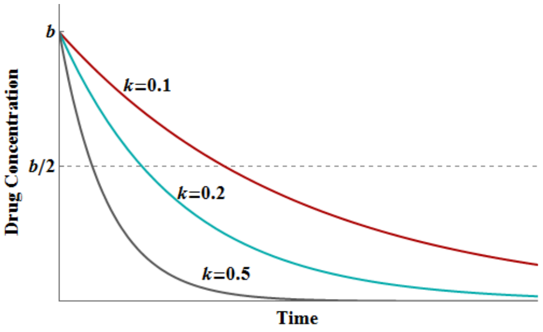


Figure 5: Graph of $x(t) = be^{-kt}$ for different drug specific decay rates k.

"half-life" of the drug), then

$$x(t^*) = \frac{b}{2} = be^{-kt^*} \quad \Rightarrow \quad \frac{1}{2} = e^{-kt^*} \quad \Rightarrow \quad \ln\left(\frac{1}{2}\right) = -kt^* \quad \Rightarrow \quad k = \frac{\ln 2}{t^*}.$$

Notice, as the value of the half-life t^* increases, the value of the decay rate k decreases, i.e., the longer the half-life, the slower the decay rate.

Now suppose an additional dose of size b is given every τ units of time. After τ units of time, the amount of the original dose left in the body is

$$x(\tau) = be^{-k\tau},$$

as given by our assumption in Equation (19). If we add another dose, then at time τ we will have $be^{-k\tau} + b$. If we let x_0 represent the concentration o the drug in the body at the time the initial dose is given, and x_1 represent the concentration of the drug in the body at the time the second dose is given then

$$x_0 = b \ \text{ and } \ x_1 = x_0e^{-k\tau} + b.$$

If we give a third dose after another τ units of time, then the concentration of drug in the body is

$$x_2 = x_1e^{-k\tau} + b.$$

In general, if we let x_n represent the concentration of the drug in the body at the time the n^{th} dose (after the initial dose) is given, then,

$$x_{n+1} = \underbrace{x_ne^{-k\tau}}_{\substack{\text{decay of} \\ n^{th} \text{ dose}}} + \underbrace{b}_{\substack{(n+1)^{th} \\ \text{dose}}} . \tag{20}$$

Since k and τ are constants, then $e^{-k\tau}$ is a constant. Let $a = e^{-k\tau}$, then we can represent Equation (20) as $x_{n+1} = ax_n + b$ which is a first order linear difference equation for which we have already found the general solution and analyzed the long-term behavior and equilibria. Notice that since $k > 0$ and $\tau > 0$, then $0 < e^{-k\tau} < 1$. This means we have $a < 1$ and $b > 0$ which corresponds to Figure 4c, and thus we should expect that the drug concentration approaches a stable equilibrium value as $n \to \infty$. The concentration of the drug over time is depicted in Figure 6. Note that the equilibrium value is depicted as x^*. What is the value of x^*?

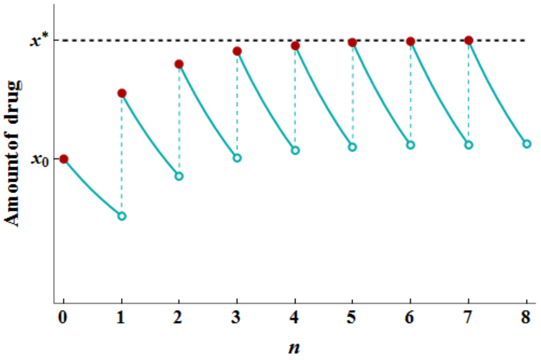


Figure 6: Amount of drug in the body where a dose of size b is administered very τ time units. The red points show the values of x_n where n represents $n\tau$ time units, the blue curves show the decay of the drug between doses, and x^* represents the equilibrium value.

We know that the equilibrium for the linear difference equation model is given by $\frac{b}{1-a}$, and that $a = e^{-k\tau}$, thus the equilibrium is

$$x^* = \frac{b}{1 - e^{-k\tau}}.$$

In this drug dosing model, the equilibrium represents the maximum amount of drug that could possibly accumulate in the body when a dose of size b of a drug with decay rate k is being given every τ units of time. This is useful information when you want to make sure the amount of drug in the body does not exceed a certain toxicity level. Notice that as the value of τ is increased the value of x^* decreases. This makes intuitive sense, since we would expect that as we allow more time to pass between doses, a patient should not be able to accumulate as much drug in their body.

Example 7.

Levetiracetam (sold under the name Keppra) is a drug used to treat epilepsy, and has a half-life of 6 – 8 hours in plasma. Plasma is the extracellular matrix of blood cells and accounts for about 55% of total blood volume.

(a) If the half-life of levetiracetam is 7 hours, what is the decay rate (k) of the drug? Given that the decay rate is estimated to be between 6 – 8 hours, what is the corresponding range in the decay rate?

(b) Suppose a patient is prescribed a daily dose of 1000 mg to be given in two 500 mg doses administered 12 hours apart. Let x_n represent the amount of levetiracetam in the body after then n^{th} dose is administered. If you wish to express the drug dosing scheme using Equation (20), what are the values of k, τ, and b? Assume that x_n is measured in mg. What is the length of one time step?

(c) What is the maximum amount of drug that will accumulate in the patient's body?

Suppose levetiracetam is most effective when the total amount in the body remains above 500 mg. However, it has been shown that a dosing regime of more than 3000 mg per day does not confer any additional benefit.

(d) If the patient were to use 1000 mg doses, what would be a suitable dosing strategy for the patient (i.e., how often should the patient take a dose)?

(d) What would be a suitable dose size if the patient takes two doses per day? What is the maximum amount of levetiracetam that accumulates in the patient's body?

(d) Could the patient take just one dose per day? If so, what should the dose size be? What is the maximum amount of levetiracetam that accumulates in the patient's body?

The solutions to this example will be worked in class.

Exercise 8 – Digoxin

Recall in Exercise 3 you constructed the difference equation $x_{n+1} = 0.69x_n$ where x_n represented the amount (in mg) of digoxin (a drug used in treating heart disease) in a patient after n days, and $x_0 = 0.5$ mg.

(a) If the amount of digoxin in the body is 69% of what is was one day prior, calculate the decay rate k.

(b) In a particular patient taking 0.5 mg doses, digoxin is effective if it remains above 0.01 mg, however digoxin toxicity occurs when the amount of digoxin in the bloodstream exceeds 0.6 mg. What would be a suitable dosing strategy for this patient (i.e., how often should the patient take a 0.5 mg dose)?

(c) Given the dosing schedule you proposed in (a), what is the maximum amount of drug that could possibly accumulate in the patient's body?

(d) Suppose it is easier for this patient to take digoxin daily. What drug dose size should be patient take each day in order for the digoxin to remain effective and non-toxic?

Exercise 9 – Cipro

Cipro is an antibiotic taken to combat many infections, including anthrax. Cipro is filtered from the blood by the kidneys at a rate of about 33% per 24-hours (that is, after 24 hours only one third of the Cipro initially in the blood remains). Suppose a patient is given a 500 mg dose each day for many days.

(a) If you wish to express the drug dosing scheme using Equation (20), what are the values of k, τ, and b? Assume that x_n is measured in mg.

(b) What is the maximum amount of drug that could possibly accumulate in the patient's body?

(c) Suppose the patient taking Cipro will experience detrimental side effects if the amount of Cirpo in her bloodstream exceeds 700 mg (this can be calculated based on the patient's body mass index). Will the dosing scheme need to be changed. If so, propose a new dosing scheme.

Exercise 10 – Fire that Neuron

The charge on a nerve cell (*neuron*) is increased by 1 millivolt every 2 milliseconds, but decays exponentially according to $x(t) = x_0 e^{-0.05t}$, where t is measured in milliseconds. Thus, if x_0 is the present charge on the cell, the charge remaining after 2 milliseconds is $x(2) = x_0 e^{-0.1}$ millivolts. Let x_n be the charge on the cell after $2n$ milliseconds.

(a) Write a difference equation representing x_{n+1} in terms of x_n, and find the general solution assuming $x_0 = 0$.

(b) The neuron will fire as soon as the total charge on the cell exceeds 4 millivolts. How frequently will the neuron fire?

7 Population Genetics

Population genetics is the study of how the genetic composition of a population changes over time. If you are unfamiliar with basic genetic theory please read through the Genetic Terminology box.

Genetic Terminology

Gene · Genes are genetic material on a chromosome that code for a trait. Often a trait is determined by multiple genes or even material on multiple chromosomes. For example, human eye color is determined by genes on two different chromosomes. A gene may be determined by the genetic information at a single location, it's *locus*, or by genetic material at several locations, or *loci*. Diploids are organisms with two copies copies of each chromosome (one from each parent), except for the sex chromosome, and thus two copies of each gene (one copy at each locus on each chromosome in a pair). Haploids are cells or organisms with one copy of each chromosome, and thus one copy of each gene.

Allele · An allele is a form of a gene at a single locus. For example, in humans there is an eye color gene on chromosome 15 with two possible alleles: B (brown) and b (blue). The gene for common blood type in humans has three alleles: A, B, and O. Some alleles are dominant over others, known as *dominant alleles*, while other alleles are not dominant, known as *recessive alleles*. For genes with only two alleles we typically use capital letters to denote dominant alleles and lower case letters to denote recessive alleles. For example, the dominant allele in the eye color gene is the allele for brown eyes, B, while the allele for blue eyes, b, is recessive. A gene typically contain pairs of alleles where one allele is inherited from the father and the other is inherited from the mother.

Genotype · A genotype is the set of alleles an organism carries. For diploid organisms like humans, each gene contains one or more pairs of alleles and may involve alleles at multiple loci. Genotype is denoted as a pair (or pairs) of letters that represent the pair (or pairs) of alleles for that particular gene. For example, the eye color gene at one locus can have alleles B (brown) and b (blue), resulting in four possible genotypes: BB, Bb, bB, and bb. The common blood type gene has three possible alleles, A, B, and O, resulting in nine possible genotypes: AA, BB, OO, AB, BA, AO, OA, BO, and OB. When it does not matter which of the alleles came from which parent, genotypes like Bb and bB are considered the same genotype. Genes with two dominant alleles are known as *homozygous dominant*, genes with two recessive alleles are called *homozygous recessive*, and genes with different alleles at a locus are known as *heterozygous*.

Phenotype · A phenotype is the physical expression of a trait, as determined by the interaction between genetics and developmental or environmental influences. For traits determined by a gene at a single locus, the gene will express the dominant allele if its genotype is homozygous dominant or heterozygous, and the recessive allele if its genotype is homozygous recessive. Under the simplest conception of the of human eye color phenotypes, the genotypes BB and Bb have the phenotype brown eyes, while genotype bb has the phenotype blue eyes. In the heterozygous case, the dominant allele (B) masks the recessive allele (b).[1] For the common blood type gene, the allele O is recessive while A and B are dominant. This arises because alleles A and B each result in the production of their own antigens while allele O is inactive. Thus, the phenotypes are determined by the antigens produced (see Table 4). In the case where the genotype is AB or BA, neither allele dominates the other, both antigens are produced, and an additional blood type is formed.

Table 4: Blood type genotypes and corresponding phenotypes.

Phenotype	Genotypes
A	AA AO, OA
B	BB BO, OB
AB	AB, BA
O	OO

Punnett Square · A Punnett square (named after Reginald C. Punnett) is a diagram used to show the potential genotypes resulting from a mating where the genotype of each of the parents is known. For example, if a woman with brown eyes (genotype Bb) and a man with blue eye (genotype bb) mated, the Punnett square in Figure 7 shows the possible genotypes of their offspring.

♀

		B	b
♂	b	bB	bb
	b	bB	bb

Figure 7: Example of a Punnett square with Bb(♀) × bb(♂).

Carriers of Recessive Alleles · Many genetic diseases, like sickle-cell anemia and albinism, are only expressed in the phenotype if an individual has the homozygous recessive genotype. In these cases, the individual is referred to as a "carrier" of the genetic disease if their genotype is heterozygous. For example, the hemoglobin gene has dominant allele S and recessive allele s. A person is a carrier for sickle-cell anemia if they have genotype Ss.

---

[1] This is a simplified view of human eye color control, since it is actually controlled by multiple genes and there are more eye colors than just brown and blue.

The Hardy-Weinberg Model

Suppose we want to keep track of the proportion of genotypes of one particular gene within a population over several generations, and suppose that gene has two possible alleles, A (dominant) and a (recessive). Let P be the frequency of the homozygous dominant genotype (AA), Q be the frequency of the heterozygous genotype (Aa), and R be the frequency of the homozygous recessive genotype (aa). Thus, if 20% of the individuals in the population have the homozygous dominant genotype, then $P = 0.20$. Since P, Q, and R are proportions, and they represent all possible genotypes (for this particular gene) within the population, $P + Q + R = 1$.

If there are N individuals in the population, then there are $2N$ alleles (for this particular gene). This means that

$$\begin{aligned}
\#\ A \text{ alleles in the population} &= \underbrace{P(2N)}_{\substack{A \text{ alleles from} \\ AA \text{ individuals}}} + \underbrace{\frac{1}{2}Q(2N)}_{\substack{A \text{ alleles from} \\ Aa \text{ individuals}}} = 2PN + QN, \quad \text{and} \\
\#\ a \text{ alleles in the population} &= \underbrace{R(2N)}_{\substack{a \text{ alleles from} \\ aa \text{ individuals}}} + \underbrace{\frac{1}{2}Q(2N)}_{\substack{a \text{ alleles from} \\ Aa \text{ individuals}}} = 2RN + QN.
\end{aligned}$$

Let p represent the proportion of alleles (for this particular gene) which are A, and let q represent the proportion of alleles that are a. Then,

$$p = \frac{2PN + QN}{2N} = P + \frac{1}{2}Q, \quad \text{and} \quad q = \frac{QN + 2RN}{2N} = \frac{1}{2}Q + R.$$

Since p and q represent proportions and A and a are the only two possible alleles (for this particular gene), the values of p and q should sum to 1. And indeed,

$$p + q = \left(P + \frac{1}{2}Q\right) + \left(\frac{1}{2}Q + R\right) = P + Q + R = 1.$$

A classic method for modeling the allele frequencies of a population from one generation to the next is known as the Hardy-Weinberg model. The model was developed independently by G. H. Hardy (mathematician) and Wilhelm Weinberg (physician) in 1908. The Hardy-Weinberg model has eight assumptions [2]:

1. The organism is diploid, sexual, and has discrete generations. Discrete generations refer to a life history like that of an annual plant, in which the parental generation has died by the time the offspring generation reproduces.
2. Allele frequencies are the same in both sexes.
3. Mendelian segregation occurs, which means that individuals with the heterozygous genotype produce equal numbers of gametes (haploid reproductive cells, e.g., eggs and sperm) containing each allele. For example, an Aa individual produces equal numbers of A and a gametes. There are a few genes that violate this assumption; this condition is known as meiotic drive or segregation disorder. When meiotic drive occurs, once allele in heterozygous individuals is overrepresented in the gametes.
4. Random mating occurs, meaning that mating is random with respect to the genotypes under consideration (it may be non-random with respect to genotypes at other loci).

5. There are no mutations (permanent change to the DNA molecule), or at least the mutation rate is negligible, i.e., very close to 0.
6. There is no migration (movement of individuals between populations). This assumption is also referred to as the population having no gene flow.
7. There is no random genetic drift which refers to fluctuations in allele frequencies that occur by chance, particularly in small subpopulations, as a result of random sampling error in the choice of gametes that form the next generation. For large populations with random mating, it is reasonable to assume there is no genetic drift.
8. There is no natural selection. Natural selection refers to a consistent (over multiple generations) relationship between fitness and phenotype, or differences in fitness among genotypes.

Let p_t be the allele frequency of A in generation t, then $q_t = 1 - p_t$ is the allele frequency of a in generation t. Since the allele frequencies are the same in both sexes (Assumption #2), we can assume that in both males and females that allele frequencies of A and a are p_t and q_t, respectively, and the genotype frequencies of AA, Aa, and aa are P_t, Q_t, and R_t, respectively. If we assume random matting occurs (Assumption #4), then we can assume the matings occur in proportion to the genotypic frequencies in the population. For example, the proportion of the male population in generation t that are genotype AA is P_t, and the proportion of the female population in generation t that are genotype AA is also P_t, thus probability of a $AA \times AA$ mating is $(P_t)(P_t) = (P_t)^2$. Similarly, the probability of a $AA \times Aa$ matings is the sum of the probability of a $AA(♂) \times Aa(♀)$ mating and a $AA(♀) \times Aa(♂)$ mating, i.e. $(P_t)(Q_t) + (P_t)(Q_t) = 2P_tQ_t$. The probability of each of the six types of mating occurring (denoted as the mating frequency) are given in Table 5.

Table 5: Hardy-Weinberg model mating frequencies and offspring genotype frequencies.

		Offspring Genotype Frequencies		
Mating Type	**Mating Frequency**	$\boldsymbol{AA}$	$\boldsymbol{Aa}$	$\boldsymbol{aa}$
$AA \times AA$	$(P_t)^2$	1	0	0
$AA \times Aa$	$2P_tQ_t$	½	½	0
$AA \times aa$	$2P_tR_t$	0	1	0
$Aa \times Aa$	$(Q_t)^2$	¼	½	¼
$Aa \times aa$	$2Q_tR_t$	0	½	½
$aa \times aa$	$(R_t)^2$	0	0	1

Using a Punnett square, the frequency of each genotype in the offspring can be determined for each mating type (see Table 5). The genotype frequency of AA in generation $t+1$ is

$$P_{t+1} = (1)(P_t)^2 + \frac{1}{2}(2P_tQ_t) + \frac{1}{4}(Q_t)^2 = (P_t)^2 + P_tQ_t + \frac{1}{4}(Q_t)^2 = \left(P_t + \frac{1}{2}Q_t\right)^2 = (p_t)^2. \tag{21}$$

The genotype frequency of Aa in generation $t+1$ is

$$Q_{t+1} = \frac{1}{2}(2P_tQ_t) + (1)(2P_tR_t) + \frac{1}{2}(Q_t)^2 + \frac{1}{2}(2Q_tR_t) = 2\left(P_t + \frac{1}{2}Q_t\right)\left(\frac{1}{2}Q_t + R_t\right) = 2p_tq_t. \tag{22}$$

The genotype frequency of aa in generation $t+1$ is

$$R_{t+1} = \frac{1}{4}(Q_t)^2 + \frac{1}{2}(2Q_tR_t) + (1)(R_t)^2 = \left(\frac{1}{2}Q_t + R_t\right)^2 = (q_t)^2. \tag{23}$$

Notice that

$$P_{t+1} + Q_{t+1} + R_{t+1} = (p_t)^2 + 2p_tq_t + (q_t)^2 = (p_t + q_t)^2 = (p_t + (1-p_t))^2 = 1,$$

as it should since P_{t+1}, Q_{t+1}, and R_{t+1} are the frequencies of the only three possible genotypes for this particular gene. Using the fact that $p_{t+1} = P_{t+1} + \frac{1}{2}Q_{t+1}$ and $q_t = 1 - p_t$, we find that

$$p_{t+1} = P_{t+1} + \frac{1}{2}Q_{t+1} = (p_t)^2 + p_tq_t = (p_t)^2 + p_t(1-p_t) = (p_t)^2 + p_t - (p_t)^2 = p_t.$$

Thus, we have just shown that the allele frequency of A in generation $t+1$ is equal to the allele frequency of A in generation t, i.e. the allele frequency of A (under the eight Hardy-Weinberg assumptions) does not change over time. Furthermore, since the allele frequency of A does not change over time, the allele frequency of a also does not change over time,

$$q_{t+1} = 1 - p_{t+1} = 1 - p_t = q_t.$$

The ***Hardy-Weinberg Principle*** encapsulates what we have just shown mathematically and states that the allele and genotype frequencies in a population will remain constant over time given the eight assumptions above (i.e. we are leaving out all the interesting and complicating factors). Note, the Hardy-Weinberg principle is sometimes also referred to as the Hardy-Weinberg equilibrium.

Gene Flow & the Continent-Island Model

The Hardy-Weinberg model is as simple as it get when it comes to modeling population genetics. To add some realism to this model, we can "relax" one of the eight assumptions and modify the model accordingly. We will start by relaxing Assumption #6 and assume that some migration is occurring. To keep things relatively simple, we will focus on one-way migration.

Suppose an island population receives migrants from a large source (continent) population. This migration causes gene flow in the direction of the island population (i.e. the influx of migrants to the island each generation changes the natural Hardy-Weinberg equilibrium that would occur in the absence of the migrants). Gene flow from the island population to the continent population may exist, but we will assume it to have a negligible effect on the continent population.

Let m be the proportion of the total number of island inhabitants in each generation which are migrants, and thus $(1-m)$ represents the proportion of native inhabitants. Let $\hat{q}$ be the frequency of allele a in the continent population (we will assume this to be constant over time), and q_t to be the frequency of allele a in the island population in generation t. Then in generation $t+1$,

$$q_{t+1} = (1-m)q_t + m\hat{q}. \tag{24}$$

Equation (24) is referred to as the Continent-Island model and was originally developed by Sewall Wright in 1931. Note this is different from the Hardy-Weinberg model which assumes $q_{t+1} = q_t$. Now, only a fraction of q_{t+1} come from the native island inhabitants, $(1-m)q_t$, with the rest coming from the migrant island inhabitants. Note, this model assumes that once migrants are on the island they will assimilate into the island population and mate randomly within that population.

Notice that Equation (24) can be rewritten as $q_{t+1} = aq_t + b$ where $a = (1-m)$ and $b = m\hat{q}$. Thus, the continent-island model is really just a first order linear difference equation (exponential and constant growth) in disguise. Thus, the closed form solution is

$$q_t \;=\; (1-m)^t q_0 + \frac{1-(1-m)^t}{1-(1-m)} m\hat{q} \;=\; (1-m)^t q_0 + \hat{q} - (1-m)^t \hat{q} \;=\; (1-m)^t (q_0 - \hat{q}) + \hat{q}.$$

What happens to the allele frequency of q_t over many generations? Note that $1-m < 1$, and thus $(1-m)^t \to 0$ as $t \to \infty$. Thus, $q_t \to \hat{q}$ as $t \to \infty$. This means the allele frequencies of A and a on the island approach the allele frequencies of A and a on the continent as successive generations of island inhabitants mate with migrants from the continent.

Example 8.

Red wolves (*Canis rufus*) historically occurred throughout southeastern North America, but by the 1960s, they were confined to a small population in Louisiana and Texas where there was hybridization with the much more abundant coyote (*Canis latrans*). Wolves from this population were captured to start a captive population in 1987, this captive population was used to establish a wild population in eastern North Carolina. However, over the next decade, coyotes colonized this area, and in the late 1990s, it was estimated that approximately 15% of the litters in the newly established population were hybrid.

To examine the impact of coyote introgression on red wolf ancestry, assume that the "island" red wolf population is receiving gene flow from the "continent" coyote population. Assume that in the initial generation the "island" red wolf population has 100% red wolf ancestry, and the "continent" coyote population has 0% red wolf ancestry. Additionally, in hybrid litters, half the ancestry (genes) are from the red wolves, and half are from the coyotes. Assume the generation length of the red wolves is approximately 5 years, and that $t = 0$ corresponds to 1990.

(a) To form the continent-island model, what are the values of m, q_0, and $\hat{q}$?

(b) Find the closed form solution of the continent-island model using the parameter values found in (a).

(c) What proportion of red wolf ancestry remains in the red wolf population after 50 years (i.e. in 2040)?

The solutions to this example will be worked in class.

Exercise 11 – Red Wolf Conservation

Recall the red wolf population discussed in Example 8. Suppose that coyote management actions are identified which can help reduce hybrid litters, thus significantly reducing the rate of gene flow from the coyote population to the red wolf population.

1. Suppose there is a 50% reduction in hybrid litters. What is the new value of m? What proportion of red wolf ancestry remains in the red wolf population in 2040 if the management actions are put into place in (i) 1990, (ii) 2000, and (iii) 2010?
2. Suppose there is a 90% reduction in hybrid litters. What is the new value of m? What proportion of red wolf ancestry remains in the red wolf population in 2040 if the management actions are put into place in (i) 1990, (ii) 2000, and (iii) 2010?

8 Nonlinear First Order Difference Equations

The last several sections have focused on biological examples which can be modeled using the linear first order difference equation $x_{n+1} = ax_n + b$, where a and b are constants. However, there are many biological systems whose dynamics cannot be captured by this linear model. In this section, we will explore some biological dynamics which are best modeled using nonlinear first order difference equations.

Resource Constrained Population Growth

Let's return to our first motivating example of the growth of a culture of *E. coli* cells. Table 1 shows the optical density of an *E. coli* cell culture measured every 30 minutes for 3 hours. However, Table 1 gives an incomplete data set. Data was actually collected every 30 minutes for 9 hours. Table 6 shows the full set of data and Figure 8 shows a graph of the full set of data (in gray, filled circles).

Table 6: Optical Density of *E. coli* cell culture.

Time (hours)	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
Cell Density (OD_{600})	0.055	0.120	0.231	0.360	0.516	0.821	1.300	1.752	2.102	2.683
Time (hours)	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	
Cell Density (OD_{600})	3.163	3.532	3.725	3.882	3.993	4.076	4.002	4.205	4.131	

What do you notice about the full set of data? Though it may have been reasonable to estimate the data with an exponential growth model for the first 3 hours, that model will certainly not be reasonable as time progresses past 3 hours. Notice that though the cell density continues to increase, around hour 6 it begins to level off. We often see this phenomenon when a population is growing in a resource constrained setting. The population will grow exponentially when the amount of resources (e.g., food or space) per individual are plentiful, but growth rates begin to slow and approach zero when the amount of resources per individual become scarce.

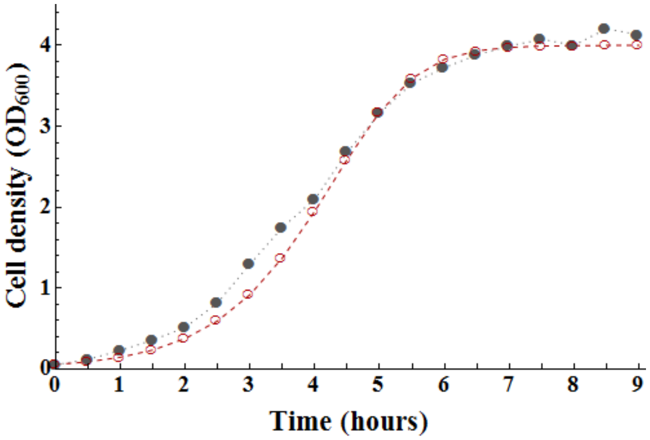


Figure 8: Optical density of the *E. coli* cell culture measured every 30 minutes over 9 hours where the actual data is shown in gray (filled circles) and the model, $x_{n+1} = x_n + 0.65x_n\,(1 - x_n/4)$, is shown in red (empty circles).

There are a variety of models which produce this type of S-shaped curve. We will begin by examining one called the ***discrete logisitic growth model*** which is defined by

$$x_{n+1} - x_n = rx_n\left(1 - \frac{x_n}{K}\right), \tag{25}$$

where r is the intrinsic growth rate, and K is the carrying capacity of the population's environment. Notice, the first term in the product on the right hand side of Equation (25) is the same as in the exponential growth model, rx_n, but it is being multiplied by the term $\left(1 - \frac{x_n}{K}\right)$. Notice that for $0 < x_n < K$, $\left(1 - \frac{x_n}{K}\right) < 1$, and $\left(1 - \frac{x_n}{K}\right) \to 0$ as $x_n \to K$. Thus, as the values of x_n approach K, the difference between x_{n+1} and x_n will go to zero, i.e., the growth rate is shrinking to 0. For the data given in Table 6 it can be estimated that $r = 0.64$ and $k = 4$. Figure 8 shows a graph of both the data and the model (in red, empty circles).

We can reduce the number of parameters in the discrete logistic growth model by "normalizing the model" with respect to the carrying capacity. Let $u_n = x_n/K$, then $x_n = Ku_n$, and Equation (25) becomes

$$u_{n+1} = u_n + ru_n(1 - u_n). \tag{26}$$

Not we refer to replacing x_n with Ku_n as either a *substitution* or *change of variables*. What is the equilibrium of the normalized discrete logistic model? Recall, the equilibrium occurs when $u_{n+1} = u_n$. Let u^* be such a value. Then,

$$u^* = u^* + ru^*(1-u^*) \quad \Rightarrow \quad 0 = ru^*(1-u^*),$$

which will hold true when $u^* = 0$ or $u^* = 1$. Thus, the normalized discrete logistic model has two equilibria: (1) when the population size is 0 (i.e., extinct), and (2) when the population size has reached carrying capacity.

Example 9.

Use cobweb diagrams and sequence plots to explore the dynamics of the normalized discrete logistic model, $u_{n+1} = u_n + au_n(1-u_n)$, as the parameter value r is increased from 0.2 to 3.1.

The solutions to this example will be worked in class.

Exercise 12 – Beverton-Holt Difference Equation

The Beverton-Holt difference equation is another commonly used for modeling the growth of a population in a resource constrained setting. The Beverton-Holt difference equation is

$$x_{t+1} = \frac{rx_t}{1 + \frac{r-1}{K}x_t}, \tag{27}$$

where r is the intrinsic growth rate, and K is the carrying capacity.

(a) Show that the substitution $u_t = 1/x_t$ transforms Equation (27) into a linear difference equation (i.e., it has the form $u_{t+1} = au_t + b$, where a and b are in terms of r and K).

(b) Given the linear difference equation you generated in part (a), find the closed form solution to the Beverton-Holt equation.

(c) Find the equilibrium for the transformed Beverton-Holt equation. Interpret this equilibrium in terms of the original Beverton-Holt equation.

(d) The original Beverton-Holt has an additional equilibrium. What is it?

(e) To determine the stability of each equilibria, consider two cases: (i) $r > 1$, and (2) $r < 1$. Use the closed form solution of the Beverton-Holt equation to determine the stability of each equilibria in each case.

(f) Use Matlab to plot the sequences generated by the Beverton-Holt difference equation and the logistic difference equation for $x_0 = 0.2$, $K = 1$, and (i) $r = 0.25$, (ii) $r = 0.75$, (iii) $r = 1.25$, (iv) $r = 1.90$, (v) $r = 2.20$, (vi) $r = 2.50$, and (vii) $r = 3.00$. Describe the ways in which the Beverton-Holt and logistic difference equations are similar and the ways in which they are different.

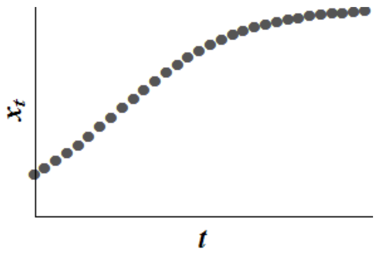


Figure 9: Plot of sequence produced by the Beverton-Holt difference equation.

Natural Selection

Recall in Section 7 we examined the Hardy-Weinberg model. The model had eight assumptions, and when we relaxed the "no migration" assumption (Assumption #6) we developed the continent-island model. In this section, we will develop a model for when we relax the "no natural selection" assumption (Assumption #8). Recall that ***natural selection*** refers to a consistent (i.e., over multiple generations) relationship between fitness and phenotype, or differences in fitness among genotypes.

Let A and a be the two possible alleles for a particular gene of interest where A is the dominant allele and a is the recessive allele. Let p_t be the allele frequency of A in generation t, and $q_t = 1 - p_t$ be the allele frequency of a in

Table 7: Genotype frequencies in generation $t+1$ with and without natural selection, where $\bar{w} = w_1 p_t^2 + 2w_2 p_t q_t + w_3 q_t^2$.

	AA	***Aa***	***aa***
Frequency in absence of natural selection	p_t^2	$2p_t q_t$	q_t^2
Relative fitness of genotype	w_1	w_2	w_3
Frequency with natural selection	$\frac{w_1 p_t^2}{\bar{w}}$	$\frac{2w_2 p_t q_t}{\bar{w}}$	$\frac{w_3 q_t^2}{\bar{w}}$

generation t. In the absences of natural selection, according to the Hardy-Weinberg model, the frequency of each possible genotype (AA, Aa, and aa) in generation $t+1$ are given in Equations (21) - (23), and shown again in Table 7. In the presence of natural selection, each genotype may confer a different relative fitness (relative with respect to the other genotypes). Let w_1 be the relative fitness of genotype AA, w_2 the relative fitness of genotype Aa, and w_3 the relative fitness of genotype w_3. We define the ***mean fitness*** of the population, $\bar{w}$, as the sum of the relative contributions of each genotype, that is,

$$\bar{w} = w_1 p_t^2 + 2w_2 p_t q_t + w_3 q_t^2. \tag{28}$$

Then, the frequency of each genotype in the presence of natural selection is the ratio of the relative contribution of each genotype to the mean fitness (see Table 7). Thus, in the presence of natural selection we have that

$$P_{t+1} = \frac{w_1 p_t^2}{\bar{w}}, \quad Q_{t+1} = \frac{2w_2 p_t q_t}{\bar{w}}, \quad \text{and} \quad R_{t+1} = \frac{w_3 q_t^2}{\bar{w}}. \tag{29}$$

Notice that, as with the Hardy-Weinberg model (without selection), $P_{t+1} + Q_{t+1} + R_{t+1} = 1$.

As with the continent-island model, let us form construct a difference equation to model the frequency of the recessive allele, a, over several generations. This time we will assume there is no gene flow, but that there is natural selection.

$$q_{t+1} = \frac{1}{2}Q_{t+1} + R_{t+1} = \frac{w_2 p_t q_t + w_3 q_t^2}{w_1 p_t^2 + 2w_2 p_t q_t + w_3 q_t^2} \tag{30}$$

Equation (30) expresses q_{t+1} as a function of p_t and q_t. To express q_{t+1} solely as a function of q_t we can use the fact that $p_t = 1 - q_t$, and thus,

$$q_{t+1} = \frac{w_2 (1 - q_t) q_t + w_3 q_t^2}{w_1 (1 - q_t)^2 + 2w_2 (1 - q_t) q_t + w_3 q_t^2}. \tag{31}$$

We will refer to Equation (31) as the ***natural selection difference equation model***.

We can change of the relative fitness values of w_1, w_2, and w_3 (mathematically described as *weight values* or *weight constants*) to reflect different "fitness relationships" between genotypes. Recall that the values of w_1, w_2, and w_3 reflect the relative fitness of each genotype. Thus, if two genotypes have the same relative fitness, their weight values should be the same. In this case, we typically set both weights to 1 for ease of computation. If one genotype has a greater fitness than another, than the relative fitness of the former genotype should be larger than that of the latter genotype. The relative fitnesses of each genotype for several different types of fitness relationships are shown in Table 8. The parameter s, called the ***selection coefficient***, is a measure of the amount of selection against the recessive allele a (or for the dominant allele A). The parameter h is called the ***level of dominance***, and when multiplied by s measures the amount of selection against the heterozygote. We will explore some of the fitness relationships in Table 8 through the following examples and exercises.

Table 8: Relative fitness values for different fitness relationships.

	Genotype		
	AA	Aa	aa
General relative fitness	w_1	w_2	w_3
Lethal recessive	1	1	0
Detrimental alleles (recessive)	1	1	$1-s$
Detrimental alleles (additive)	1	$1-{}^s/_2$	$1-s$
Dominance (purifying selection)	1	$1-hs$	$1-s$
Dominance (positive selection)	$1+s$	$1+hs$	1
Heterozygote advantage	$1-s_1$	1	$1-s_2$
Heterozygote disadvantage	$1+s_1$	1	$1+s_2$

Example 10.

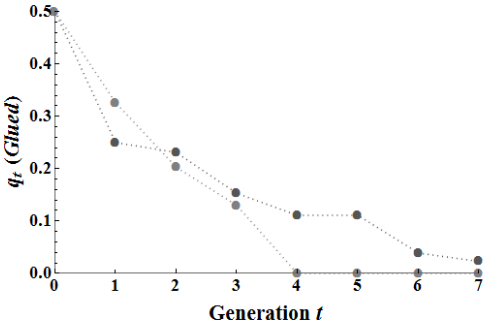


Figure 10: The frequency of the recessive (lethal) *glued* allele for two *D. melanogaster* population replicates each starting with all heterozygotes ($q_0 = 0.5$).

Lethal Recessive Submodel · There are some genes for which the homozygous recessive genotype has the phenotype of a lethal disease which causes pre-reproductive death, resulting in the fitness of the homozygous recessive genotype being 0, i.e., $w_3 = 0$. Some examples of such diseases in humans are Tays-Sachs disease and cystic fibrosis.

In *Drosophila melanogaster* (also known as the common fruit fly or vinegar fly), the homozygous recessive genotype of the gene *glued* causes pre-reproductive death, and the heterozygous genotype reduces eye size and affects eye appearance. Two replicate populations of *Drosophila melanogaster* were initiated with all heterozygotes so that $q_0 = 0.5$. The frequency of the recessive allele was recorded for 7 successive generations for each replicate population (data shown in Figure 10). We will compare two different lethal recessive submodels of the natural selection difference equation model and determine which more accurately represents the data shown.

(a) For the first model, we will assume that the recessive allele only causes a reduction in fitness in the homozygous recessive genotype, that is, the homozygous dominant and heterozygous genotypes have the same relative fitness: $w_1 = 1$, $w_2 = 1$, and $w_3 = 0$. Given the values of w_1, w_2, and w_3, simplify Equation (31), and determine a closed form solution. What happens to the values of q_t as $t \to \infty$?

(b) For the second model, we will assume that in additional to the lethality of the homozygous recessive genotype the defects in the heterozygotes (reduced eye size and altered eye appearance) cause a reduction in fitness in the heterozygotes such that $w_1 = 1$, $w_2 = {}^2/_3$, and $w_3 = 0$. Note, this would be a case of "dominance (purifying selection)" with $s = 1$ (lethal recessive) and $h = {}^2/_3$. Given the values of w_1, w_2, and w_3, use a cobweb diagram to determine what happens to the values of q_t as $t \to \infty$.

(c) Which model best describes the data shown in Figure 10?

(d) For each model calculate the number of generations needed for the recessive allele frequency to fall below 0.01. Which fitness relationship removes the recessive allele frequency from the population more quickly? Does this make intuitive sense?

The solutions to this example will be worked in class.

Example 11.

Detrimental Recessive Submodel · In many instances, there is no complete selection against the homozygous recessive genotype, and thus, the relative fitness of the homozygote is only partially reduced (when compared with other genotypes). For many human genetic diseases, such as albinism and sickle cell anemia, homozygous recessive individuals can survive and produce progeny, although the probability of this occurring is reduced compared with that of other individuals. In *Drosophila*, mice, corn, and other organisms that have been investigated in detail genetically, there are many examples of recessive morphologic mutants that reduce fitness of homozygotes but do not cause lethality. In the case of a detrimental (but non-lethal) recessive allele, $w_1 = 1$, $w_2 = 1$, $w_3 = 1 - s$, where $0 < s < 1$.

(a) Given the values of w_1, w_2, and w_3, simplify Equation (31) to form the detrimental recessive submodel.

(b) What are the equilibria of the detrimental recessive submodel?

(c) Use a cobweb diagram to determine what happens to the values of q_t as $t \to \infty$ for different values of q_0 and s. What affect does increasing the value of s have on the model?

The solutions to this example will be worked in class.

Exercise 13 – Heterozygous Advantage

A rare example of heterozygote advantage in a natural population is the plumage polymorphism in common buzzards (*Buteo buteo*) in Europe. Buzzard feathers vary from dark brown to almost pure white, and much of this variation is due to a single gene with two alleles: D (dark feathers, dominant) and d (light feathers, recessive). The relative fitness for the different genotypes are $w_1 = 0.45$ (DD), $w_2 = 1$ (Dd), and $w_3 = 0.54$ (dd).

(a) Given the values of w_1, w_2, and w_3, simplify (31).

(b) What are the equilibria of the model in part (a)?

(c) Use a cobweb diagram to determine what happens to the value of q_t as $t \to \infty$ for different values of q_0. Which equilibria are stable and which are unstable?

(d) Suppose the allele frequency reaches equilibrium within the population, which equilibrium would you expect to see in the population? If the allele frequency in the population is at equilibrium, then the genotype frequencies are also in equilibrium. Use Equations 29 to determine the genotype frequencies at equilibrium.

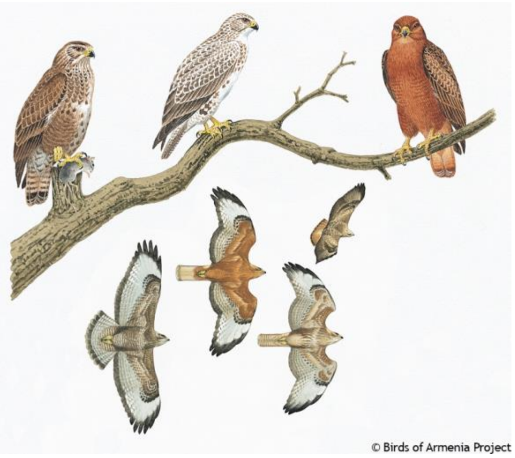


Figure 11: Polymorphs of the common buzzard.

Exercise 14 – Blood Type Rh Factor

In mammals, there are interactions between a mother and her fetus that influence the survival of offspring of particular genotypes. The best understood maternal-fetal incompatibility system is the Rh (rhesus) blood group system on the short arm of chromosome 1 in humans. The the *RHD* gene of the Rh system gives the + or – in the human blood type. The dominant allele R indicates the presence of antigen D from gene *RHD*, while the recessive allele r indicates the absence of antigen D. Fetal mortality occurs when the fetus is Rh positive (genotype RR or Rr) and the mother is Rh negative (genotype rr) and the maternal antibody production of D destroys the red blood cells of the fetus. Incompatible combinations only occur when the mother is Rh negative and the father is Rh positive. Approximately 10% to 15% of the matings in most Caucasian populations are incompatible.

Incompatible combinations can only occur when the mother is Rh negative and the father is Rh positive, that is matings $RR(♂) \times rr(♀)$ and $Rr(♂) \times rr(♀)$. In the matings $RR(♂) \times rr(♀)$, all offspring are incompatible with the mother (since all progeny have genotype Rr which is Rh positive). In the matings $Rr(♂) \times rr(♀)$, half of all offspring are Rr while the other half are rr. Thus, only half of the progeny are Rh positive and incompatible with the mother. Since a mother who is Rh negative will never produce progeny with genotype RR, selection is only occurring against heterozygous progeny.

(a) Assuming relative fitnesses of $w_1 = 1 + s_1$, $w_2 = 1$, and $w_3 = 1 + s_2$, simplify Equation (31) to form the heterozygous disadvantage model.

(b) Find all equilibria for the model you formed in (a). *Hint: There are 3 equilibria, and you will need to use the quadratic formula to find two of them.*

(c) Assume that $s_1 = 1$ and $s_2 = 0.5$ use a cobweb diagram to determine the stability of each of the equilibria.

(d) Most human populations have high frequencies of the R allele, but Basques, who live in the Pyrenees Mountains between Spain and France, have a high frequency ($q = 0.65$) of the r allele. Assuming that the allele frequency of r in Pyrenees Moutain Basque population is at equilibrium determine an expression for the relationship between s_1 and s_2 using the non-trivial equilibrium found in part (b). Is the advantage greater for the homozygous dominant or homozygous recessive genotype?

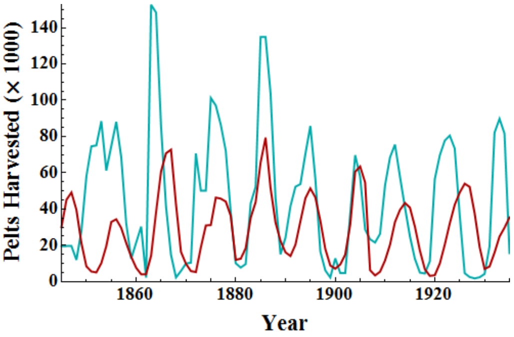
---

9 Systems of Difference Equations

Biological processes rarely happen in isolation. Interactions such as those between predators and prey, competitors for limited resources, and many biochemical processes at the cellular level are major drivers of biological complexity. These interactions are also rarely linear; that is, we typically cannot derive how one process affects another by assuming that one is proportional to the other. In this section we will explore how we can use discrete difference equations to model a biological system consisting of multiple species, classes, or organism groups.

Predator-Prey Systems

A classic example of predator-prey system dynamics can be found in the Canadian lynx (*Lynx canadensis*) and snowshoe hare (*Lepus americanus*) pelt-trading records of the Hudson Bay Company (see Figure 12a, thousands of lynx pelts shown in red and thousands of hare pelts shown in blue). The primary food source for the Canadian lynx is the snowshoe hare. If we think of the number of pelts of each species harvested each year by the Hudson Bay Company as being representative of the respective population sizes (obviously the total popualtion sizes are larger than the number of pelts caught), then we can use the pelt data to inform a model of the population sizes of both the lynx and the hares.



(a) Pelt-trading records of the Hudson Bay Company for Canadian lynx (red) and snowshoe hares (blue).

(b) Lynx catching a hare for his mid-afternoon snack.

Examine the data in Figure 12a. What do you notice? You should observe that both populations cycle over roughly a 10 year period. Additionally, the cycles in the lynx pelts harvested (on average) lags behind the cycles in the hare pelts harvested by 1 to 3 years.

In 1920 Alfred J. Lotka (a mathematician and physical chemist) proposed a set of ordinary differential equations to model the dynamics of an herbivorous animal species and the plant species it consumed. In 1926, Vito Volterra (a mathematician and physicist) independently investigate similar equations as applied to fish catches in the Adriatic. If you have taken a Calculus II course or a Differential Equations course, you have likely seen the classic Lotka-Volterra equations. In this course, we will examine the discrete difference equation analog of the Lotka-Volterra model,

$$
\begin{aligned}
x_{n+1} &= rx_n - \alpha x_n y_n \quad \text{(prey population size or density)} &(32)\\
y_{n+1} &= sy_n + \beta x_n y_n \quad \text{(predator population size or density)} &(33)
\end{aligned}
$$

where $r > 1$ is the growth rate of the prey in the absence of the predator, $0 < s < 1$ is the survival rate of the predator in the absence of its prey source, $\alpha > 0$ is the consumption rate of the predators (where αx_n is the average number of prey eaten per predator in time step n), and $\beta > 0$ is the growth rate of the predator population due to the consumption of prey. Notice the amount of prey consumed by the predator at each time step depends on not only on the numbers of predators present, but also the numbers of prey present. When the prey density is higher, it is easier for the predators to find prey and thus more prey will be consumed. Likewise, the growth rate the predators depends not only on the number of predators present (more predators should produce more offspring overall), but also on the number of prey present (a large food source often correlates with higher reproductive rates and/or larger litter or clutch sizes).

We will use this model, Equations (32) - (33), to demonstrate how to find the equilibria of a system of difference equations. However, the discrete Lotka-Volterra (like the discrete logistic) becomes unrealistic (negative population sizes, or unrealistically large population sizes) with many parameters sets (i.e., the values of r, s, α, and β). A simulation for one parameter set is shown in Figure 13, but in order for the model to remain realistic, the time step must be scaled to represent 14.6 hours (a little more than half a day). Notice also that population density (not population size) is shown. The initial conditions, $x_0 = 2$ and $y_0 = 3$, implies that the initial ratio of prey individuals to predator individuals (per unit area) is 2:3.

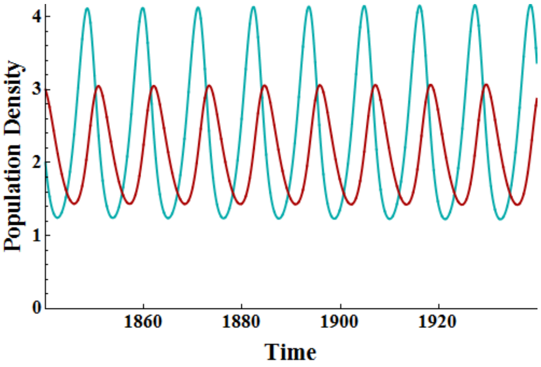


Figure 13: Plot of the sequences produced by the discrete Lotka-Volterra model, Equations (32) - (33), with $r = 1.0015$, $s = 0.9994$, $\alpha = 0.0006$, $\beta = 0.00025$, and each time step represents 14.6 hours (a little more than half a day).

To find the equilibria of the discrete Lotka-Volterra model, recall that an equation is at an equilibrium when the variable remains constant from one time step to the next. Thus, (x^*, y^*) is an equilibrium to the system $x_{t+1} = f(x_t, y_t)$ and $y_{t+1} = g(x_t, y_t)$, where f and g are functions, if $f(x^*, y^*) = x^*$ and $g(x^*, y^*) = y^*$. For Equations (32) - (33) this means

$$
\begin{aligned}
x^* &= rx^* - \alpha x^* y^*\\
x^* - rx^* + \alpha x^* y^* &= 0\\
x^*\left(1 - r + \alpha y^*\right) &= 0
\end{aligned}
$$

Thus, either $x^* = 0$, or $1 - r + \alpha y^* = 0$. The second equation implies $y^* = (r-1)/\alpha$.

$$
\begin{aligned}
y^* &= sy^* + \beta x^* y^*\\
y^* - sy^* - \beta x^* y^* &= 0\\
y^*\left(1 - s - \beta x^*\right) &= 0
\end{aligned}
$$

Thus, either $y^* = 0$, or $1 - s - \beta x^* = 0$. The second equation implies $x^* = (1-s)/\beta$.

Any equilibrium point of the system must satisfy both equation ($x^* = rx^* - \alpha x^* y^*$ and $y^* = sy^* + \beta x^* y^*$) simultaneously. The algebra above shows us we should consider four different cases: (1) $x^* = 0$ and $y^* = 0$, (2) $x^* = 0$ and $y^* \neq 0$ (i.e. $1 - s - \beta x^* = 0$), (3) $x^* \neq 0$ (i.e. $1 - r + \alpha y^* = 0$) and $y^* = 0$, and (4) $x^* \neq 0$ and $y^* \neq 0$.

Case 1: If $x^* = 0$ and $y^* = 0$, then both equations are satisfied simultaneously, and $(0, 0)$ is an equilibrium.

Case 2: Suppose $x^* = 0$ and $y^* \neq 0$. If $y^* \neq 0$, then $x^* = (1-s)/\beta$ which cannot be true since $x^* = 0$. Thus, we cannot have an equilibrium where $x^* = 0$ and $y^* \neq 0$.

Case 3: Suppose $x^* \neq 0$ and $y^* = 0$. If $x^* \neq 0$, then $y^* = (r-1)/\alpha$ which cannot be true since $y^* = 0$. Thus, we cannot have an equilibrium where $x^* \neq 0$ and $y^* = 0$.

Case 4: If $x^* \neq 0$ and $y^* \neq 0$, then $x^* = (1-s)/\beta$ and $y^* = (r-1)/\alpha$, and both equations are satisfied simultaneously. Thus, $((1-s)/\beta, (r-1)/\alpha)$ is an equilibrium.

Therefore, the two equilibria of the discrete Lotka-Volterra model are $(0,0)$, which we will refer to at the ***trivial equilibrium***, and $((1-s)/\beta, (r-1)/\alpha)$.

Example 12.

Consider a discrete-time predator-prey model in which the prey grows logistically in the absence of the predator, that is

$$\begin{aligned} x_{t+1} &= x_t + r x_t (1 - x_t) - \alpha x_t y_t \\ y_{t+1} &= s y_t + \beta x_t y_t, \end{aligned}$$

where $r > 1$, $0 < s < 1$, $\alpha > 0$, and $\beta > 0$ are defined the same as in the discrete Lotka-Volterra.

(a) Find all equilibria of this predator-prey model.

(b) Let $x_0 = 0.6$ and $y = 0.1$ (where x and y represent population densities). To answer each of the following make a time series plot of each population over at least 50 time steps. Additionally, make a plot of y_t vs. x_t over the same number of time steps. Plot the equilibria points on the second plot.
  (i) Describe the dynamics of each population when $r = 2$, $s = 0.8$, $\alpha = 0.5$, $\beta = 1.5$. What is the long-term behavior of the model?
  (ii) Describe the dynamics of each population when $r = 2$, $s = 0.8$, $\alpha = 0.5$, $\beta = 2.5$. What is the long-term behavior of the model?
  (iii) Describe the dynamics of each population when $r = 2$, $s = 0.8$, $\alpha = 1.8$, $\beta = 2.5$. What is the long-term behavior of the model?

The solutions to this example will be worked in class.

Example 13.

Related to the Lotka-Volterra model is the Nicholson-Bailey model, developed in the 1930s, to describe the population dynamics of a host-parasitoid system where both species have a number of life-cycle stages including eggs, larvae, pupae, and adult. Let p_t be the population of adult female parasitoids who will search for hosts uniformly distributed over an area A and who will oviposit c eggs in each host encountered, and each parasitoid searches an area A_S per unit time. Assuming that the hosts occurs in discrete generations, let h_t be the number of hosts in generation t. Each host lays enough eggs to produce λ larvae (the host net reproductive rate), and each larvae is vulnerable to parasitism for a period of time τ. The h_t hosts of generation t produce λh_t potential hosts in the $t+1$ generation, but only

$$e^{-\frac{A_S}{A}\tau p_t}$$

survive. The system dynamics are defined by

$$\begin{aligned} h_{t+1} &= \lambda h_n e^{-a p_t} \\ p_{t+1} &= c h_n \left(1 - e^{-a p_t}\right), \end{aligned}$$

where $a = A_S \tau / A$ is referred to as the "searching efficacy" of the host. Note that $e^{-a p_t}$ is a value in the range $[0,1]$ and represents the probability that a host will survive p_t parasitoids, while $1 - e^{-a p_t}$ is the probability that they will not survive.

1. Find all equilibria of the Nicholson-Bailey model.
2. Assume $h_0 = 15$, $p_0 = 10$ are measured as densities (#/hectare), and $\lambda = 2$, $c = 1$, and $a = 0.07$. Write a Matlab script to generate the terms for the first 30 generations after the initial generation. What happens to the populations over time?
3. In your Matlab script add a plot of the terms for the first 30 generations in the (h_t, p_t) space. Additionally, plot the location of any equilibrium points. Which equilibrium point is stable?

4. The Nicholson-Bailey model has been modified relax the assumption of the host population occurring in discrete generations. In this modification, we assume that some proportion s of the h_t host generation survives to the next generation. Thus, the host equation is modified to
$$h_{t+1} = \lambda h_n e^{-ap_t} + sh_t.$$
Find all the equilibria of this modified model.
5. Write a new Matlab script, similar to the one above, to generate graphs of he first 100 generations after the initial generation for this modified model for $s = 0.2$, 0.5, 0.7, 0.9. Plot the location of this model's equilibrium points. Which equilibrium point is stable for each value of s?
6. How did modifying the model impact the population dynamics? How did modifying the model impact the long-term behavior of the host-parasitoid system?

The solutions to this example will be worked in class.

Exercise 15 – A Competition Model

One way that multiple populations interact is through predator-prey interactions as we explored above. However, another way that multiple populations can interact is through competition for resources. For example, if two populations both feed upon a the same resource, or if two populations establish nests, burrows, dens, etc. in the same spatial range. Consider the Leslie-Gower difference equation competition model,
$$x_{t+1} = \frac{b_1 x_t}{1 + x_t + c_1 y_t} \quad \text{and} \quad y_{t+1} = \frac{b_2 y_t}{1 + c_2 x_t + y_t},$$
where x_t and y_t are the densities of two competing populations, and all the parameters (b_1, b_2, c_1, and c_2) are positive.

(a) Assume there is only one population, i.e. $y_t = 0$ for all t. Simplify the equation for the x population. What is the equilibrium of this population? Use cobweb diagramming to determine the stability of each equilibria (*Hint: Each equilibria will be stable under certain conditions*). Based on your examination of the model so far what does the remaining parameter in the equation represent?

(b) What happens to the x population if the parameter value c_1 is increased? What happens to the y population if the parameter value c_2 is increased? If $b_1 = b_2$ and $c_1 > c_2$ what do you think will happen to the two competing species?

(c) Find all equilibria of the competition model.

(d) Create a plot of y_t vs. x_t and additionally plot each of the equilibria points on this graph. For each of the following parameter sets, indicate which equilibria are in the first quadrant (i.e., biologically reasonable), and which equilibria is approached as $t \to \infty$. You should examine several different initial conditions for each parameter set (always choose $x_0 > 0$ and $y_0 > 0$).
  (i) $b_1 = b_2 = 2$, $c_1 = 0.4$, and $c_2 = 0.6$
  (ii) $b_1 = b_2 = 2$, $c_1 = 1.4$, and $c_2 = 0.6$
  (iii) $b_1 = b_2 = 2$, $c_1 = 0.4$, and $c_2 = 1.6$
  (iv) $b_1 = b_2 = 2$, $c_1 = 1.4$, and $c_2 = 1.6$

Allele Frequencies of Sex-Linked Genes

In dioecious organisms (those species with two sexes for which one member of each sex is required for reproduction), sex is determined by a difference in the chromosomes of each sex. The form that these difference take vary among different species. In many animals (humans included), sex is determined by a single chromosome which we will refer to as the sex-chromosome, the remaining chromosomes are referred to as autosomes.

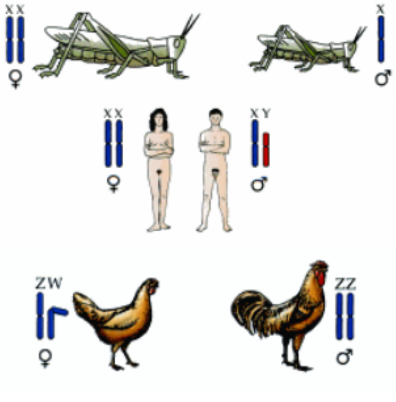


Figure 14: Sex chromosomes in different species.

In grasshoppers the females have two X chromosomes, while the males only have one. The eggs formed by the females have one copy of each autosome, plus one X chromosome. The males form two types of sperm (in roughly equal amounts): one which contains one copy of each autosome plus one X chromosome, and the other which only contains one copy of each autosome. When a sperm containing an X chromosome fertilizes an egg, the egg develops into a female grasshopper. When a sperm which does not contain an X chromosome fertilizes an egg, a male grasshopper develops. We will refer to the female grasshoppers as being XX, and the male grasshoppers as being XO.

In mammals (including humans), females have two X chromosomes and males have one X chromosome and one Y chromosome. Both males and females have two copies of each autosome. Like the grasshoppers, mammal males produce two types of sperm, but in this case one which contains one copy of each autosome plus one X chromosome, and the other which contains one copy of each autosome plus one Y chromosome. Here, we refer to female mammals as being XX, and male mammals as being XY.

In birds, butterflies, and moths, the females are XY and the males are XX. To avoid confusion, we use a different set of letters for these species. We will refer to females as ZW, and males as ZZ.

For the following discussion we will assume that females are XX and males are XO or XY. When the locus of a gene occurs on one of the sex chromosomes, we must modify the Hardy-Weinberg model. Recall that Assumption #4 of the Hardy-Weinberg model was, “Allele frequencies are the same in both sexes.” If the locus of a gene occurs on the X chromosome, then the females will have two alleles for that gene and the males will only have one allele for that gene. This will necessarily alter the allele frequencies between the sexes and we must account for this in the model.

Let P_t^f, H_t^f, and Q_t^f represent the diploid genotype frequencies AA, Aa, and aa in the females of the population in generation t, and P_t^m and Q_t^m represent the haploid genotype frequencies A and a in the males of the population in generation t. Then, the allele frequencies in each of the sexes in generation t are

$$(\female,\ A)\ \ p_t^f = P_t^f + \frac{1}{2}H_t^f \quad \text{and} \quad (\male,\ A)\ \ p_t^m = P_t^m,$$
$$(\female,\ a)\ \ q_t^f = Q_t^f + \frac{1}{2}H_t^f \quad \text{and} \quad (\male,\ a)\ \ q_t^m = Q_t^m.$$

If we assume that there are an equal number of males and females in each generation, then two-thirds of the alleles are in the females and one-third of the alleles are in the males, and thus the frequencies of alleles A and a in the population in generation t are

$$p_t = \frac{2}{3}p_t^f + \frac{1}{3}p_t^m \quad \text{and} \quad q_t = \frac{2}{3}q_t^f + \frac{1}{3}q_t^m,$$

respectively. To determine the allele frequencies in the next generation $(t+1)$, we start by examining the genotype frequencies in generation $t+1$. Table 9 shows the genotype frequencies in males and females of generation $t+1$ sort by mating type. For example, given a $AA(\female) \times a(\male)$ mating (row 2 in Table 9), the only possible genotype of the offspring is Aa in the females and A in the males (remember, the males get their X chromosome from their mother and their Y chromosome from their father). Since the genotype frequency of AA in females in generation t is P_t^f, and the genotype frequency of a in males in generation t is Q_t^m, the probability of a $AA(\female) \times a(\male)$ mating is $P_t^f Q_t^m$. The frequency of female (and male) offspring from this type of mating is the probability of this mating type, i.e. $P_t^f Q_t^m$. Since there is only one female genotype resulting from this type of mating (Aa), the frequency of the Aa genotype in female offspring resulting from a $AA(\female) \times a(\male)$ mating is $P_t^f Q_t^m$. Likewise, since there is only one male genotype resulting from this type of mating (A), the frequency of the A genotype in male offspring resulting from a $AA(♀) \times a(♂)$ mating is also $P_t^f Q_t^m$.

In contrast to this, consider a $Aa(♀) \times A(♂)$ mating (row 3 in Table 9). Since the genotype frequency of Aa in females in generation t is H_t^f, and the genotype frequency of A in males in generation t is P_t^m, the probability of a $Aa(♀) \times A(♂)$ mating is $H_t^f P_t^m$. From this type of mating (on average) half of the female offspring will have genotype AA and half will have genotype Aa. Thus, the frequencies of the AA and Aa genotypes in female offspring are both $\frac{1}{2} H_t^f P_t^m$. Likewise, from this type of mating (on average) half of the male offspring will have genotype A and half will have genotype a. Thus, the frequencies of each genotype in male offspring is $\frac{1}{2} H_t^f P_t^m$.

Table 9: Genotype frequencies of offspring for a gene on an X chromosome assuming random mating with ♀ XX and ♂ XY or XO.

Mating Type		**Female Offspring**			**Male Offspring**	
♀ ♂	*Frequency*	AA	Aa	aa	A	a
$AA \times A$	$P_t^f P_t^m$	$P_t^f P_t^m$	—	—	$P_t^f P_t^m$	—
$AA \times a$	$P_t^f Q_t^m$	—	$P_t^f Q_t^m$	—	$P_t^f Q_t^m$	—
$Aa \times A$	$H_t^f P_t^m$	$\frac{1}{2} H_t^f P_t^m$	$\frac{1}{2} H_t^f P_t^m$	—	$\frac{1}{2} H_t^f P_t^m$	$\frac{1}{2} H_t^f P_t^m$
$Aa \times a$	$H_t^f Q_t^m$	—	$\frac{1}{2} H_t^f Q_t^m$	$\frac{1}{2} H_t^f Q_t^m$	$\frac{1}{2} H_t^f Q_t^m$	$\frac{1}{2} H_t^f Q_t^m$
$aa \times A$	$Q_t^f P_t^m$	—	$Q_t^f P_t^m$	—	—	$Q_t^f P_t^m$
$aa \times a$	$Q_t^f Q_t^m$	—	—	$Q_t^f Q_t^m$	—	$Q_t^f Q_t^m$
Total	1	$p_t^f p_t^m$	$p_t^f q_t^m + q_t^f p_t^m$	$q_t^f q_t^m$	p_t^f	q_t^f

In the final row of Table 9 we calculate the frequency of each genotype in each of the sexes. For example, the genotype frequency of AA in female offspring (generation $t+1$) is found by summing all the frequencies in the $AA(♀)$ column.

$$P_{t+1}^f = P_t^f P_t^m + \frac{1}{2} H_t^f P_t^m = \left(P_t^f + \frac{1}{2} H_t^f \right) P_t^m = p_t^f p_t^m.$$

Similarly, the genotype frequency of A in male offspring is

$$P_{t+1}^m = P_t^f P_t^m + P_t^f Q_t^m + \frac{1}{2} H_t^f P_t^m + \frac{1}{2} H_t^f Q_t^m = \left(P_t^f + \frac{1}{2} H_t^f \right) (P_t^m + Q_t^m) = p_t^f,$$

since $P_t^m + Q_t^m = 1$. Using the fact that $p_t^f + q_t^f = 1$ and $p_t^m + q_t^m = 1$, we can now calculate the allele frequencies of a in both the females and males in generation $t+1$,

$$\begin{aligned} q_{t+1}^f &= Q_{t+1}^f + \frac{1}{2} H_{t+1}^f = q_t^f q_t^m + \frac{1}{2} \left(1 - q_t^f \right) q_t^m + \frac{1}{2} q_t^f \left(1 - q_t^m \right) = \frac{1}{2} \left(q_t^f + q_t^m \right), \\ q_{t+1}^m &= Q_t^m = q_t^f. \end{aligned}$$

As with the predator-prey system we now have a system of difference equations. The value of q^f in the next generation depends on the values of q^f and q^m in the current generation, and the value of q^m in the next generation depends on the value of q^f in the current generation.

Example 14.

Show that the system of difference equations

$$q_{t+1}^f = \frac{1}{2}\left(q_t^f + q_t^m\right) \quad \text{and} \quad q_{t+1}^m = q_t^f$$

has infinitely many points of equilibrium. Then, test many different initial conditions, (q_0^f, q_0^m), and devise a rule of thumb for the value of the equilibrium point given the initial conditions. *The solutions to this example will be worked in class.*

Example 15.

An interesting example of a gene residing on the X chromosome is the *orange* coat-color locus in cats. This gene has two alleles, O which codes for an orange coat, and b which codes for a black coat. Females with the heterozygous genotype Ob have tortoise-shell or calico coats. In a population of feral London cats, the sex and genotype of cats sent to a clinic for "destruction" were noted by a clinic worker. *PSA: If hearing about cats being sent to clinics for destruction (and yes, that means what you think it means) makes you sad, you should advocate for spaying and neutering cats. Most medium to large cities in the United States have free/low cost spaying and neutering programs to help reduce feral cat population sizes.*

Genotype	OO(♀)	Ob(♀)	bb(♀)	O(♂)	b(♂)
# Observed	7	54	277	42	311

(a) Determine the genotype frequencies of each genotype for both male and females.

(b) Let q_t^f represent the frequency of the O allele in the females in generation t, and q_t^m the frequency of the allele O in the males in generation t. If the data presented above were collected over a short period of time, we can assume it represents one generation (we will call it generation 0). Calculate the values of q_0^f and q_0^m.

(c) Given the values of q_0^f and q_0^m, and the system of difference equations

$$q_{t+1}^f = \frac{1}{2}\left(q_t^f + q_t^m\right) \quad \text{and} \quad q_{t+1}^m = q_t^f$$

estimate the equilibrium point.

(d) At the equilibrium point, what proportion of females have tortoise-shell coats.

The solutions to this example will be worked in class.

Exercise 16 – Natural Selection on X-Linked Gene

The model for the allele frequencies of a gene on an X chromosome (an X-linked gene) assumes that the relative fitness of each of the genotypes (for both males and females) are equal. However, this assumption is not always true. To model natural selection for an X-linked gene we use a variation of the natural selection model in Equation (31),

$$q_{t+1}^f = \frac{\frac{1}{2}w_{12}\left((1-q_t^f)q_t^m + q_t^f(1-q_t^m)\right) + w_{22}q_t^f q_t^m}{w_{11}(1-q_t^f)(1-q_t^m) + w_{12}\left((1-q_t^f)q_t^m + q_t^f(1-q_t^m)\right) + w_{22}q_t^f q_t^m} \quad \text{and} \quad q_{t+1}^m = \frac{w_2 q_t^f}{w_1(1-q_t^f) + w_2 q_t^f},$$

where w_{11}, w_{12}, w_{22}, w_1, and w_2 are the relative fitnesses of genotypes AA(♀), Aa(♀), aa(♀), A(♂), and A(♂), respectively. Let us consider a case where selection occurs against the homozygous recessive genotype in females and against the recessive genotype in males, i.e.,

$$w_{11} = 1, \quad w_{12} = 1, \quad w_{22} = 1 - s_f, \quad w_1 = 1, \quad \text{and} \quad w_2 = 1 - s_m.$$

(a) Given these relative fitnesses, simplify the natural selection model for an X-linked gene.

(b) Now, suppose the recessive allele is lethal, in that individuals with genotypes aa and a do not survive to reproduce. Then $s_f = s_m = 1$. Further simplify the model assuming a lethal recessive allele.

Figure 15: *Drosophila melanogaster* with red eyes (left) and white eye (right).

(c) Find the equilibria of the system assuming a lethal recessive allele.

(d) An example of X-linked selection occurs in *Drosophila melanogaster* (commonly known as fruit flies) on the white gene. The white gene controls eye color. The dominant allele A codes for red eyes, but the recessive allele a codes for white eyes. The diploid genotypes AA and Aa yield red-eyed females, while the aa genotype yields white-eyed females. The haploid genotype A yields red-eyed males, while the genotype a yields white-eyed males. The selection acts primarily against the white-eyed males by reducing their mating ability ($s_m = 0.9$), however there is also some reduction in the fertility of white-eyed females ($s_f = 0.5$). Use Matlab to generate a plot of q_t^f, q_t^m, and q_t for $t = 0, 1, \ldots, 7$ given (i) $q_0^f = q_0^m = 0.833$, and (ii) $q_0^f = 1$, $q_0^m = 0.5$.

References

[1] E.N. Bodine, S. Lenhart, and L.J. Gross. 2014. *Mathematics for the Life Sciences*. Princeton University Press.

[2] J.K. Conner and D.L. Hartl. 2004. *A Primer of Ecological Genetics*. Sinauer Associates, Inc.

[3] J.M. Cushing, S. Levarge, N. Chitnis, and S.M. Henson. 2004. Some discrete competition models and the competitive exclusion principle. *Journal of Difference Equations and Applications*, **10**(13–15): 1139–1151.

[4] F.R. Giordano, W.P. Fox, S.B. Horton, and M.D. Weir. 2009. *A First Course in Mathematical Modeling,* 4^{th} *Edition*. Brooks/Cole.

[5] P.W. Hedrick. 1976. Simulatoin of X-linked selection in *Drosophila*. *Genetics*, **83**: 551-571.

[6] P.W. Hedrick. 2011. *Genetics of Populations,* 4^{th} *Edition*. Jones & Bartlett.

[7] M. Kot. 2001. *Elements of Mathematical Ecology*. Cambridge University Press.

[8] B.E. Statland and T.J. Demas. 1980. Serum caffeine half-lives. Healthy subjects vs. patients having alcoholic hepatic disease. *American Journal of Clinical Pathology*, **73**(3): 390–393.

[9] `http://www.fwrc.msstate.edu/predator/bobSurvive.asp`

Answer Key to Exercises

Exercise 1 – Wild Hares

(a) $x_{n+1} = 1.12x_n$, $x_0 = 200$

(b) $x_n = (1.12)^n(200)$

(c) $x_6 \approx 395$ hares

(d) $x_n \to \infty$ as $n \to \infty$ which is not biologically reasonable. it *is* reasonable to assume that though growth may be exponential for a time, at some point the hare population will begin to exceed the capacity of its environment in terms of space or other resources (e.g., food), and thus limit reproduction or increase the death rate within the population.

Exercise 2 – Elk Herd

(a) $x_n = (1.1)^n(50)$; increasing sequence; $x_{20} \approx 336$ elk

(b) $x_n = (0.9)^n(50)$; decreasing sequence; $x_{20} \approx 6$ elk; If the average lifespan of a wild elk is 15 years, there certainly must be elk that live longer than 15 years. However, to truly test if the assumption in this case makes sense, we would need to know the ages of the 50 elk that were originally introduced into the wildlife management area, and an upper bound on the age of an elk in the wild. If most of the elk were already between over the age of 5, then this model would predict that 6 of the original elk lived to be at least 25 years old. This is 10 years older than the average and may or may not be reasonable.

Exercise 3 – Treating Heart Disease

(a) $a = 0.69$

(b) $x_n = 0.69^n(0.5)$

(c) Between 5 and 10 days

Exercise 4 – Dixogin Dosing

(a) $a = 0.69$, $b = 0.5$, one time step represents 1 day

(b) $x_0 = 0.5000$, $x_1 = 0.8450$, $x_2 = 1.0831$, $x_3 = 1.2473$, $x_4 = 1.3606$

(c) $x_n = \frac{0.5}{1-0.69}(1 - 0.69^{n+1})$

(d) $x_n \to \frac{0.5}{1-0.69} \approx 1.6129$ mg as $n \to \infty$

Exercise 5 – Stocking a Lake

$x_{n+1} = ax_n + b$, $x_0 = 12,230$, $a = 1.8$. With the current fishing policy $b = 1200$. To keep the population at equilibrium given the current population size, solve $x^* = x_0$ (or $b/(1-1.8) = 12230$) for b to get a maximum sustainable yield of $b = -9784$.

Exercise 6 – Buffalo

(a) $x_n = 800(1.1)^n + 200$

(b) No

(c) 100 buffalo

Exercise 7 – Trout

(a) $n = 5$

(b) 222 trout

Exercise 8 – Digoxin

(a) $k = 0.371$ (when time is measured in days)

(b) Using the difference equation $x_{n+1} = e^{-0.371\tau}x_n + 0.5$, solve $x^* < 0.6$ and $x^* - b < 0.01$ for τ to find $4.8 < \tau < 10.6$. Thus, a reasonable dosing schedule would be every seven days (i.e. $\tau = 7$).

(c) $x^* = \dfrac{0.5}{1 - e^{-0.371\times 7}} = 0.54$ mg

(d) Using the difference equation $x_{n+1} = e^{-0.371\times 1}x_n + b$, solve $x^* < 0.6$ and $x^* - b < 0.01$ for b to find $0.004 < b < 0.186$. Thus, a reasonable dose size given daily would be $b = 0.05$ mg (50 μg) or $b = 0.10$ mg (100 μg).

Exercise 9 – Cipro

(a) $k \approx 1.1$, $\tau = 1$, $b = 500$

(b) 750 mg

(c) Yes, the dosing scheme will need to be change. It can be changed by either lowering the dose size or increasing the time between doses. If lowing the dose size, b would need to be lowered to below 469 mg. A safe dose would be 450 mg. If increasing the time between doses, τ would need to be increased above 1.14 days ($\approx$ 1 day, 3 hours, 22 minutes). This might be a difficult regime for the patient to follow. It may be reasonable to increase the time between doses to 2 days, but it is unknown if this would allow the amount of Cipro in the bloodstream to fall below its level of effectiveness.

Exercise 10 – Fire that Neuron

(a) $x_{n+1} = \left(e^{-0.1}\right)x_n + 1$; $x_n = \dfrac{1 - e^{-0.1n}}{1 - e^{0.1}}$

(b) 10 milliseconds

Exercise 11 – Red Wolf Conservation

(a) $m = 0.0375$

  (i) $q_t = (1-m)^t q_0$ with $q_0 = 1$; in 2040 ($t = 10$) 68% red wolf ancestry remains

  (ii) $q_t = (1-m)^t q_0$ with $q_0 = (1-0.075)^2$; in 2040 ($t = 8$) 63% red wolf ancestry remains

  (iii) $q_t = (1-m)^t q_0$ with $q_0 = (1-0.075)^4$; in 2040 ($t = 6$) 58% red wolf ancestry remains

(b) $m = 0.0075$

  (i) $q_t = (1-m)^t q_0$ with $q_0 = 1$; in 2040 ($t = 10$) 93% red wolf ancestry remains

  (ii) $q_t = (1-m)^t q_0$ with $q_0 = (1-0.075)^2$; in 2040 ($t = 8$) 81% red wolf ancestry remains

  (iii) $q_t = (1-m)^t q_0$ with $q_0 = (1-0.075)^4$; in 2040 ($t = 6$) 70% red wolf ancestry remains

Exercise 12 – Beverton-Holt Difference Equation

(a) Let $u_t = 1/x_t$, then $x_t = 1/u_t$. Substitute $x_t = 1/u_t$ into Equation (27) and simplify to obtain $u_{t+1} = \frac{1}{r}u_t + \frac{r-1}{rK}$.

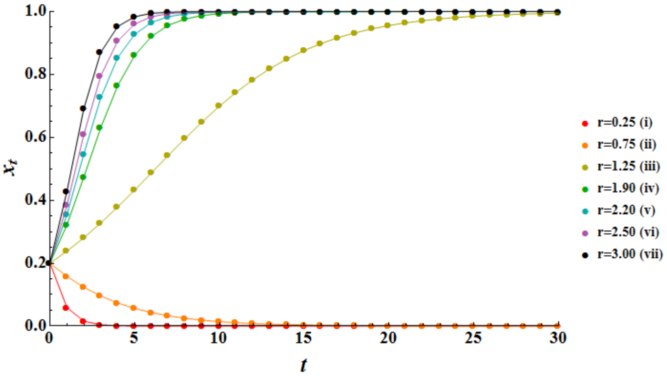
(b) $u_t = \left(u_0 - \frac{1}{K}\right)\left(\frac{1}{r}\right)^t + \frac{1}{K} \quad \Rightarrow \quad x_t = \frac{K}{1+(u_0K-1)\left(\frac{1}{r}\right)^t}$

(c) $u^* = \frac{1}{K}$ (transformed Beverton-Holt), thus $x^* = K$ (original Beverton-Holt). This means that the equilibrium of the Beverton-Holt model is the carrying capacity of the model.

(d) $x^* = 0$; This is often called the trivial equilibrium.

(e) (i) If $r > 1$, then $\frac{1}{r} < 1$ and $\left(\frac{1}{r}\right)^t \to 0$ as $t \to \infty$, and thus $x_t \to K$ as $t \to \infty$. (ii) If $r < 1$, then $\frac{1}{r} > 1$ and $\left(\frac{1}{r}\right)^t \to \infty$ as $t \to \infty$, and thus $x_t \to 0$ as $t \to \infty$.

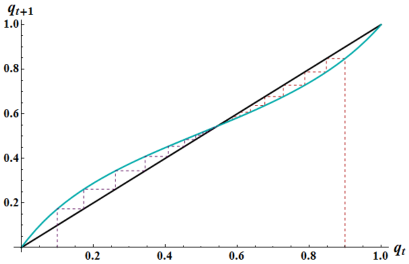
(f)



Exercise 13 – Heterozygous Advantage

(a) $q_{t+1} = \dfrac{q_t\,(1-0.46q_t)}{0.45 + 1.1q_t - 1.01q_t^2}$

(b) $q^* = 0$, ${}^{55}\!/_{101}$ (≈ 0.545), 1

(c) 

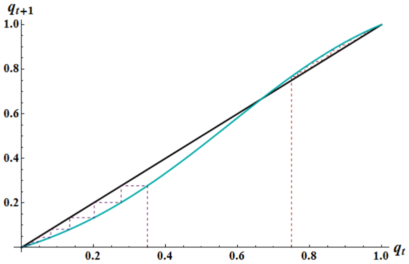
$q_t \to {}^{55}/_{101}$ as $t \to \infty$, $q^* = 0$ unstable, $q^* = {}^{55}/_{101}$ stable, $q^* = 1$ unstable

(d) Expect to see $q^* = {}^{55}/_{101}$ in the population since it is the stable equilibrium. Genotype frequencies at equilibrium: $P^* = 0.12454$ (DD), $Q^* = 0.66181$ (Dd), $R^* = 0.21365$ (dd)

Exercise 14 – Blood Type Rh Factor

(a) $q_{t+1} = \dfrac{q_t + s_2 q_t^2}{1 + s_1 (1 - q_t)^2 + s_2 q_t^2}$

(b) $q^* = 0, \ \dfrac{s_1}{s_1 + s_2}, \ 1$

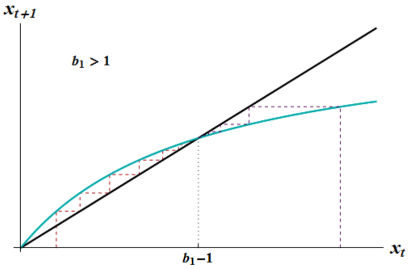
(c) 

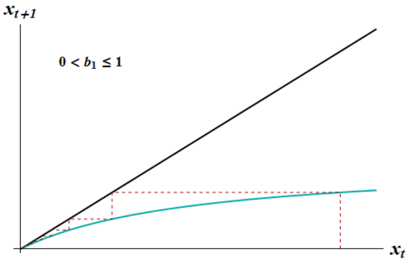
$q^* = 0$ stable, $q^* = {}^{2}/_{3}$ unstable, $q^* = 1$ stable

(d) $s_2 = 7s_1/13$, the advantage is greater for the homozygous dominant genotype

Exercise 15 – A Competition Model

(a) $x_{t+1} = \frac{b_1 x_t}{1 + x_t}$; the equilibria are $x^* = 0, \ b_1 - 1$;





$x^* = 0$ is stable when $0 < b_1 < 1$, $x^* = b_1 - 1$ is stable when $b_1 > 1$, and when $b = 1$ the only equilibrium is $x^* = 0$ which is stable in this case; The parameter b_1 acts like a growth rate, and when $b_1 > 1$ the populations is growing and when $0 < b_1 \leq 1$ the population is declining to extinction.

(b) As the value of c_1 is increased the value of x_{t+1} decreases. The parameter c_1 is a measure of the severity with which x's competition with y reduces the growth of x. Likewise, as the value of c_2 is increased, the value of y_{t+1} decreases, and the parameter c_2 is a measure of the severity with which y's competition with x reduces the growth of y. If the growth rates of the two populations are the same ($b_1 = b_2$), then the fate of the populations comes down to who has a competitive advantage, which is population y when $c_1 > c_2$. We would expect population y

to grow more quickly than population x, and possibly even drive population x to extinction.

(c) $(x^*,y^*) = (0,0)\,,(0,b_2-1),\;(b_1-1,0),\;\left(\dfrac{b_1-b_2c_1+c_1-1}{1-c_1c_2},\dfrac{b_2-b_1c_2+c_2-1}{1-c_1c_2}\right)$; Notice that if $b_1<1$, then the equilibrium $(b_1-1,0)$ is not biologically reasonable as it implies $x^*<0$. Likewise, if $b_2<1$, then the equilibrium $(0,b_2-1)$ is not biologically reasonable. Notice that under certain parameter sets, the fourth equilibrium does not reside in the first quadrant and is thus biologically unreasonable. If you want an extra exercise, show that when $b_1=b_2$, the fourth equilibrium will be in the first quadrant when $c_1<1<{}^1\!/_{c_2}$.

(d) EQ1 $=(0,0)$, EQ2 $=(0,b_2-1)$, EQ3 $=(b_1-1,0)$, and EQ4 $=\left(\dfrac{b_1-b_2c_1+c_1-1}{1-c_1c_2},\dfrac{b_2-b_1c_2+c_2-1}{1-c_1c_2}\right)$

- (i) All four equilibria are in the first quadrant. All initial conditions lead to EQ4.
- (ii) EQ4 is no longer in the first quadrant. All initial conditions lead to EQ2.
- (iii) EQ4 is no longer in the first quadrant. All initial conditions lead to EQ3.
- (iv) All four equilibria are in the first quadrant. The equilibria that is approaches as $t\to\infty$ is determined by the relative position of (x_0,y_0) with respect to the line $y_0=\frac{b_2-b_1c_2+c_2-1}{b_1-b_2c_1+c_1-1}x_0=1.5x_0$. When $y_0<1.5x_0$, then solutions approach EQ3; When $y_0=1.5x_0$, then solutions approach EQ4; When $y_0>1.5x_0$, then solutions approach EQ2.

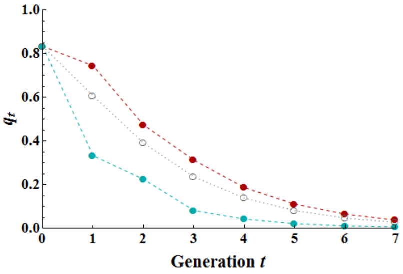
Exercise 16 – Natural Selection on X-Linked Gene

(a) $q^f_{t+1}=\dfrac{\frac{1}{2}(q^f_t+q^m_t)-s_fq^f_tq^m_t}{1-s_fq^f_tq^m_t}$ and $q^m_{t+1}=\dfrac{(1-s_m)q^f_t}{1-s_mq^f_t}$

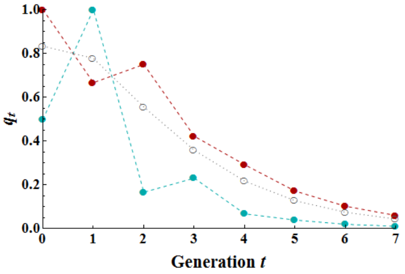
(b) $q^f_{t+1}=\frac{1}{2}q^f_t$ and $q^m_t=0$

(c) $(0,0)$ is the only equilibria

(d) (i)



(ii)



q^f_t (red), q^m_t (blue), and q_t (gray)