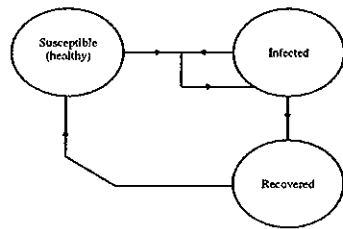


Chapter 1

Introduction



1.1 The Modeling Process

Mathematical biology dwells at the interface of two fields: applied/computational mathematics and biology. Individually, these fields are growing quickly due to rapidly changing technology and newly emerging subdisciplines. Coupled together, the fields provide the basis for the emerging scientific discipline of mathematical biology, whose focus is interdisciplinary scientific problems in quantitative life sciences.

What can biology offer mathematics and computation? Biological models offer a seemingly endless supply of challenging and interesting nonlinear problems to solve. These nonlinear problems can provide a testing ground for applied mathematical and computational methods, and generate the impetus to develop new mathematical and computational methods and approaches.

What can mathematics and computation offer biology? Mathematics and computation can help solve a growing problem in biological research. Data collection, varying from gene sequencing to remote sensing via satellites, is now inundating biologists with complex patterns of observations. The ability to collect new data outstrips our ability to heuristically reason mechanisms of cause and effect in complex systems. It is the analysis of mathematical models that allows us to formalize the cause and effect process and tie it to the biological observations.

The mathematical model describes interactions between biological components. Analysis of the model, via computational and applied mathematical methods, allows us to deduce the consequences of the interactions. For example, voltage-dependent data on movement of electrically charged ions across a nerve membrane are inputs for models of electrophysiology. The output is a prediction of the dynamics of electrical activity in nerves. The behavior and survival of newly infected individuals are inputs to disease models. The output is a prediction of when and where the disease will outbreak, and how it can be controlled.

To become a successful modeler, modeling tools are required. The first part of this book gives an introduction to some of the more powerful modeling tools, such as discrete models, ordinary differential equations, partial differential equations, stochastic models,

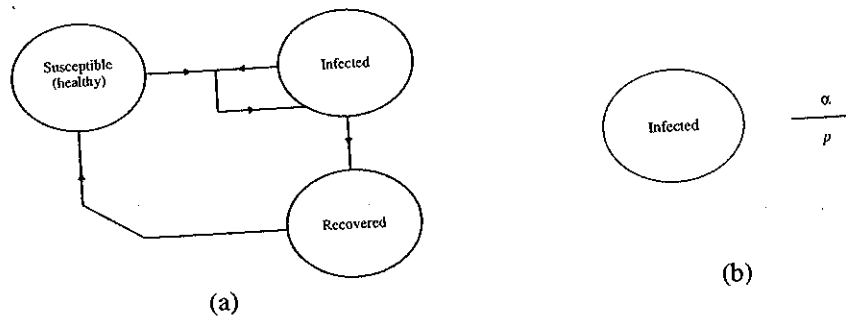


Figure 1.1. (a) Arrow diagram for a simple epidemic model, showing the relationships between the classes of susceptible, infected, and recovered individuals. (b) Subgraph of the arrow diagram in (a) representing the recovery of infected individuals, with probability p or rate α .

cellular automata, and parameter estimation techniques. The second and third parts of the book apply the modeling tools to biological problems.

1.2 Probabilities and Rates

We start with the derivation of a simple epidemic model for the spread of an infectious disease, such as influenza, through a population of healthy individuals. Assume that one infected individual is introduced into the population. In addition, assume that the infection is spread from individual to individual through contact, and that the infected recover after a certain period of time (two weeks for influenza). Recovered individuals are not available to catch the disease again.

Thus, after some time, the population consists of three types of individuals, namely, susceptible (healthy), infected, and recovered individuals. The relationships between these three classes are shown in Figure 1.1 (a). Note that in the diagram, recovered individuals can become susceptible again. In this case, we can think of recovered individuals being temporarily immune to the disease. Individuals return to the susceptible class when the immunity wears off.

In order to create a model for this situation, we need to quantify this diagram. To do that, we follow these three steps:

1. First, we identify the important quantities (the *dependent variables*) to keep track of. In our example, there are three classes of individuals. Let S be the number of susceptibles in the population, I the number infected, and R the number recovered.
2. Second, we identify the *independent variables*, such as time t , space x , or age a , and so on. For our example, we write $S(x, t)$, $I(x, t)$, and $R(x, t)$ if we wish to include time and space dependence, but not age dependence.
3. Finally, we quantify the transitions and/or interactions between these classes, as indicated by the arrows in Figure 1.1 (a). To do this, we use either *probabilities* or *rates*, as explained below.

To explain the use of probabilities versus rates, we consider a subgraph of Figure 1.1 (a), concerning only the recovery of infected individuals, shown in Figure 1.1 (b). In the discussion below, note that we ignore the generation of infected individuals through contact between infected and susceptible individuals (the full epidemic model will be treated in Section 3.3.3). In order to create a model representing this particular process, we apply the three steps outlined above:

1. The dependent variable is the number of infected individuals, I .
2. As time progresses, infected individuals recover. Thus, the independent variable is time, t .
3. If we assume that 2 out of every 100 infected individuals recover per day, then the probability of recovery in a single day is $p_1 = \frac{2}{100}$. The corresponding rate, α_1 , is defined as the probability per unit of time, that is,

$$\alpha_1 = \frac{p_1}{\text{unit of time}} = \frac{2}{100} \cdot \frac{1}{\text{day}} = \frac{1}{50} \frac{1}{\text{day}}.$$

Similarly, the probability of recovery in two days is $p_2 = \frac{4}{100}$ (we use p_n to denote the probability of recovering in n days). The corresponding rate, α_2 , is then

$$\alpha_2 = \frac{4}{100} \cdot \frac{1}{2 \text{ days}} = \frac{1}{50} \frac{1}{\text{day}}.$$

For a time unit of $\frac{1}{2}$ of a day, we get $p_{\frac{1}{2}} = \frac{1}{100}$ and

$$\alpha_{\frac{1}{2}} = \frac{1}{100} \cdot 2 \frac{1}{\text{day}} = \frac{1}{50} \frac{1}{\text{day}}.$$

We find that the rate α is independent of the time unit chosen, whereas the probability depends on the chosen time unit. Since the rate is independent of the chosen time unit, we can generalize. Let Δt denote a general unit of time, and let $p_{\Delta t}$ be the probability of recovering in Δt . Then the number of infectives after one unit of time is given as

$$I(t + \Delta t) = I(t) - p_{\Delta t} I(t).$$

With some rearrangements, we get

$$\begin{aligned} I(t + \Delta t) - I(t) &= -p_{\Delta t} I(t), \\ \frac{I(t + \Delta t) - I(t)}{\Delta t} &= -\frac{p_{\Delta t}}{\Delta t} I(t), \\ \frac{I(t + \Delta t) - I(t)}{\Delta t} &= -\alpha I(t), \end{aligned}$$

where now the rate $\alpha = \frac{p_{\Delta t}}{\Delta t}$ appears.

Since α is constant for all values of Δt , we can take the limit as $\Delta t \rightarrow 0$. On the left, we obtain the differential quotient, and we obtain the following equation governing the dynamics of $I(t)$:

$$\frac{d}{dt} I(t) = -\alpha I(t).$$

To summarize, for the simple subgraph shown in Figure 1.1 (b), we found two models, namely, a *discrete-time model* with probabilities,

$$I(t + \Delta t) = I(t) - p_{\Delta t} I(t), \quad (1.1)$$

and a *continuous-time model* with rates (a *differential equation*),

$$\frac{d}{dt} I(t) = -\alpha I(t). \quad (1.2)$$

Both models can be solved, analyzed, and simulated. For the discrete-time model, (1.1), we have to specify a time unit, say $\Delta t = \frac{1}{2}$ day. Then $p_{\Delta t} = p_{\frac{1}{2}} = \alpha \cdot \frac{1}{2}$ day. If we define $I_n := I(n \cdot \Delta t)$, then we obtain the simple *difference equation*

$$I_{n+1} = \left(1 - P_{\frac{1}{2}}\right) I_n,$$

which has the solution

$$I_n = \left(1 - P_{\frac{1}{2}}\right)^n I_0, \quad n \geq 1,$$

where I_0 denoted the initial number of infected individuals. The differential equation, (1.2), is solved by an exponential, $I(t) = I(0) e^{-\alpha t}$. The latter solution indicates that the number of infected individuals decreases with time, as expected intuitively (recall that the generation of new infected individuals has been ignored).

In Figure 1.2, we compare the solutions of the discrete-time and continuous-time models over a time period of 15 days, starting with 100 infected individuals ($I(0) = I_0 = 100$)

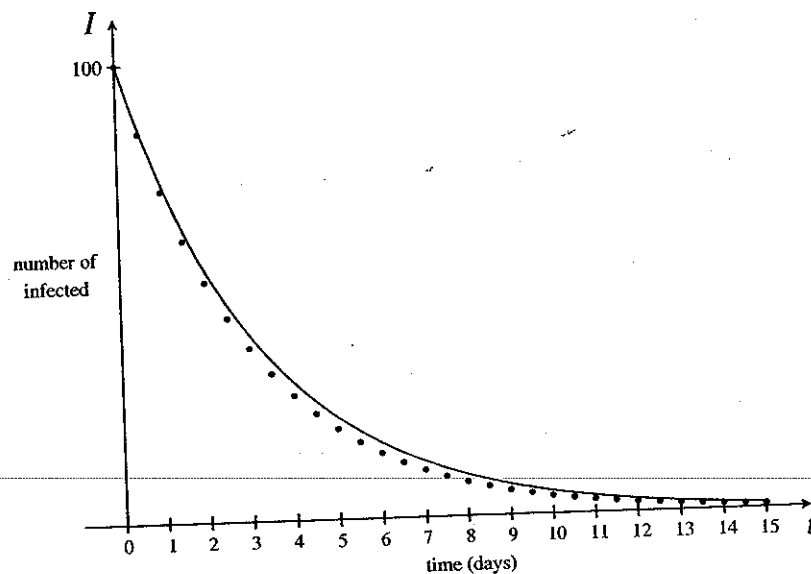


Figure 1.2. Comparison of the solutions to the discrete-time model, (1.1), and the continuous-time model, (1.2), starting with 100 infected individuals ($I(0) = I_0 = 100$) and using a recovery rate of $\alpha = 0.3$.

(1.1)

(1.2)

ne model,
day. If weation, (1.2),
the number
generationuous-time
(0) = $I_0 =$ →
15 t, (1.1), and the
= $I_0 = 100$,

100) and using a recovery rate of $\alpha = 0.3$ per day and a time increment of $\Delta t = 1/2$ day. The full epidemic model corresponding to the arrow diagram shown in Figure 1.1 (a) will be discussed in detail in Section 3.3.3.

In Figure 1.2, the agreement of the discrete and the continuous models is quite convincing. However, this is not always the case. In Exercise 1.4.2, the reader is asked to vary Δt and to investigate if the agreement is still good.

1.3 Model Classes

In the previous section, we derived two models for the recovery of infected individuals, namely, a discrete-time or difference equation, (1.1), and a differential equation, (1.2). The difference between these models is that the time variable is discrete for the difference equation, whereas it is continuous for the differential equation. So far, both models appear suitable. The final choice of model depends on the scientific question asked, the purpose of the model, the available data, etc.

The independent or state variables also can be chosen to be either discrete or continuous. For example, a discrete state variable may represent the number of individuals in a population, whereas a continuous variable may represent a density or a concentration.

Both of the above models are called *deterministic*. That means that if you know the state of the system at a certain point in time t , you can *determine* all future states by solving the corresponding model. Sometimes, however, *stochastic* effects play a dominant role. For example, in a laboratory setting you can predict that a pair of healthy rabbits will produce offspring. Outside the laboratory, life is less predictable, and the same pair of rabbits may not reproduce. In general, stochastic variations are more important for small population sizes. A model for small populations and unpredictable environments should include the uncertainty via a stochastic formulation. Large populations in constant environments (such as an aggregate of cellular slime molds, which contains about 100,000 cells) usually are modeled by deterministic models.

The number of choices presented above generates many types of models. A discussion of all types of models is beyond the scope of this book. We have chosen to restrict the material in this book to the most common model classes, summarized in the following list:

Difference Equations: The state (or dependent variable) can be discrete or continuous but the time is always discrete. Discrete models are suitable for seasonal events. We treat deterministic difference equations in Chapter 2 and stochastic difference equations in Chapter 5.

Ordinary Differential Equations (ODEs): ODEs are used to describe population evolution over a continuous time period. Deterministic ODEs are one of the major modeling tools and are discussed in detail in Chapter 3. The theory of stochastic differential equations is quite involved and is not covered in this book.

Partial Differential Equations (PDEs): PDEs are used if two or more continuous independent variables are used, for example, time and space, or time and age. We discuss age-structured models and reaction-diffusion equations for spatial spread in Chapter 4. Historically, stochastic PDEs were used primarily in the context of statistical

physics. Only recently have such models been considered to describe population dynamics (see, e.g., [76]).

Stochastic Processes: Stochastic processes and *Markov chains* are completely stochastic model classes. They are particularly useful for small populations. We treat them in detail in Chapter 5.

Cellular Automata: Cellular automata and related models are fully discrete models. All independent variables (such as time and space) and all dependent variables (such as population sizes) are discrete. The analysis of cellular automata is mainly restricted to computer analysis and numerical simulation. We give an introduction in Chapter 6. Cellular automata can be either deterministic or stochastic, using a random number generator.

1.4 Exercises for Modeling

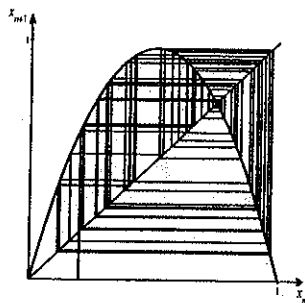
Exercise 1.4.1: Discrete-time versus continuous-time models. Assume you have a culture of bacteria growing in a petri dish, and each cell divides into two identical copies of itself every 10 minutes.

- Choose a unit of time, and find the corresponding probability of cell division.
- Write down a discrete-time model which balances the amount of cells at time t and at time $t + \Delta t$.
- Define the growth rate, and derive the corresponding continuous-time model.
- Solve both the discrete-time and continuous-time models, and compare the solutions.
- When is a discrete-time model appropriate? When is a continuous-time model appropriate?

Exercise 1.4.2: Comparison of discrete and continuous models. Study the two models (1.2), (1.1) which lead to Figure 1.2 and vary the time increment Δt (e.g., try $\Delta t = \frac{1}{4}$ day, $\frac{1}{8}$ day, 1 day, 2 days, 10 days). What do you observe? Which choice of Δt gives the best, and which gives the worst agreement? Can you explain why?

Exercise 1.4.3: Structured populations.

- Give examples of spatially structured problems. What kind of effects cannot be understood without spatial structure?
- Give an example of an age-structured problem.
- Give an example of a size-structured problem.



Chapter 2

Discrete-Time Models

2.1 Introduction to Discrete-Time Models

In this chapter, we use *discrete-time models* to describe dynamical phenomena in biology. Discrete-time models are appropriate when one can think about the phenomenon in terms of discrete time steps or when one wishes to describe experimental measurements that have been collected at fixed time intervals.

In general, we are concerned with a sequence of quantities,

$$x_0, x_1, x_2, x_3, x_4, \dots,$$

where x_i denotes the quantity at the i th measurement or after i time steps. For example, x_i may represent

- the size of a population of mosquitoes in year i ;
- the proportion of individuals in a population carrying a particular allele of a gene in the i th generation;
- the number of cells in a bacterial culture on day i ;
- the concentration of oxygen in the lung after the i th breath;
- the concentration in the blood of a drug after the i th dose.

You can undoubtedly think of many more such examples. Note that the time step may or may not be constant. In the example of the bacterial culture, the time step is fixed to be a day, but in the example of the oxygen concentration in the lung, the time step is variable from breath to breath. Also, time steps can be anywhere from milliseconds to years, depending on the biological problem at hand.

We can now ask ourselves, what does it mean to build a discrete-time model? In the context of our sequence of quantities x_i , a discrete model is a rule describing how the quantities change. In particular, a discrete model describes how x_{n+1} depends on x_n (and

perhaps x_{n-1}, x_{n-2}, \dots). Restricting ourselves to the case where x_{n+1} depends on x_n alone, a model can then be thought of as an updating function (Adler [1]) of the form

$$x_{n+1} = f(x_n). \quad (2.1)$$

Equation (2.1) is often referred to as a discrete-time equation or difference equation, and f is called a *map*.

Given some initial condition x_0 , the updating function can be *iterated* to give $x_1 = f(x_0), x_2 = f(x_1), x_3 = f(x_2)$, and so on. The resulting simulated sequence x_0, x_1, x_2, \dots is called an *orbit* of the map. A good model should be able to produce orbits that are in close agreement with observed experimental data.

Finding the precise function f that describes experimental data well or that gives a certain desired type of behavior is not always straightforward. It is often said that modeling (here, finding the right function f) is more of an art than a science. One starts with a particular function f , and then makes adjustments. Insight into how a function f should be adjusted to get a better model can often be obtained from knowledge of the behavior of the current model.

Simple but powerful analytical tools are available to help determine possible types of behavior of a given model. In this chapter, we will give an introduction to some commonly used tools. We divide the chapter into two main sections. Section 2.2 deals with scalar discrete-time equations of the form (2.1), while Section 2.3 deals with systems of discrete-time equations. Throughout the chapter, applications of discrete-time equations to real biological systems, such as population growth and genetics, are discussed. More applications can be found in the exercises at the end of this chapter.

2.2 Scalar Discrete-Time Models

2.2.1 Growth of a Population and the Discrete Logistic Equation

In this section, we build a simple model describing the growth of a population of *Paramecium aurelia*. A paramecium is a unicellular organism found in large numbers in freshwater ponds. It is a member of the group of organisms called protozoa and feeds on small organisms such as bacteria and other protozoa.

We will build the model based on a classic data set collected by Gause [63]. In Table 2.1, the mean density of *Paramecium aurelia*, measured in individuals per 0.5 cm^3 , is tabulated as a function of time, measured in days. The corresponding graph of the data is shown in Figure 2.1. The population was grown in isolation and provided with a constant level of nutrients.

Let p_n be the mean density of this population on day n . A good starting point for building a model for p_n is to think of the word equation

$$\text{future value} = \text{present value} + \text{change},$$

which readily translates to the following mathematical equation:

$$p_{n+1} = p_n + \Delta p_n. \quad (2.2)$$

The goal of the modeling process, then, is to find a reasonable approximation for Δp_n that more or less reproduces the given set of data.

Table 2.1. Growth of *Paramecium aurelia* in isolation. Here, density is the number of individuals per 0.5 cm³. Data taken from Gause [63].

Day (n)	Mean density of <i>P. Aurelia</i> (p_n)	Change in density ($\Delta p_n = p_{n+1} - p_n$)
0	2	—
1	—	—
2	14	20
3	34	32
4	56	38
5	94	95
6	189	77
7	266	64
8	330	86
9	416	91
10	507	73
11	580	30
12	610	-97
13	513	80
14	593	-36
15	557	3
16	560	-38
17	522	43
18	565	-48
19	517	-17
20	500	85
21	585	-85
22	500	-5
23	495	30
24	525	-15
25	510	—

Finding a suitable form for Δp_n is not always easy. But let's examine the data more closely. Initially, the population increases slowly; values of Δp_n , tabulated in the third column of Table 2.1, are relatively small. As time progresses, values of Δp_n increase and reach a maximum approximately halfway through the experiment. After that, they decrease again. We can attribute the decrease in the growth rate to intraspecific competition for nutrients and space. At the end of the experiment, the population appears to be leveling off when it reaches a mean density of approximately 540 individuals per 0.5 cm³ (roughly, Δp_n is negative when $p_n > 540$ and Δp_n is positive when $p_n < 540$). To keep things relatively simple, we will ignore the fluctuations in the population. Note that the choice of using 540

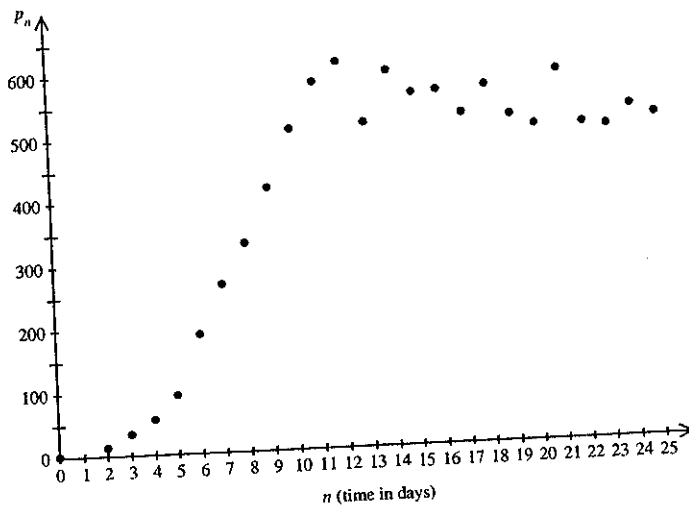


Figure 2.1. Graph of the data shown in Table 2.1.

for the limiting density may not be the best. However, it will do for a first approximative model. For clarity, then, we continue to use this number in the following discussion, but keep in mind that there may be room for improvement.

Based on the above observations, we must find a suitable form for Δp_n that is small when p_n is close to 0 and 540, and positive for intermediate values of p_n . The following quadratic expression fits the bill:

$$\Delta p_n = k(540 - p_n)p_n. \quad (2.3)$$

Note that this expression ensures that $\Delta p_n < 0$ when $p_n > 540$ and $\Delta p_n > 0$ when $p_n < 540$. Substituting (2.3) into (2.2), we obtain the following model for the population:

$$p_{n+1} = p_n + k(540 - p_n)p_n, \quad (2.4)$$

where the value of the parameter k remains to be determined.

The experimental data contains enough information to allow us to obtain an estimate for k from the data set. In particular, note that we have hypothesized that $\Delta p_n = p_{n+1} - p_n$ is proportional to the product $(540 - p_n)p_n$, with the parameter k being the constant of proportionality. To test our hypothesis, we plot $\Delta p_n = p_{n+1} - p_n$ versus $p_n(540 - p_n)$ and check whether there is reasonable proportionality. The graph is shown in Figure 2.2.

Although the data looks scattered, we can fit it nicely with a straight line passing through the origin, consistent with our hypothesis. The line of best fit has slope approximately 0.00145 (in Chapter 8, you will learn how to obtain lines of best fit using Maple). Thus, setting $k = 0.00145$, we obtain the following model for the growth of the population:

$$p_{n+1} = p_n + 0.00145(540 - p_n)p_n. \quad (2.5)$$

Last but not least, we compare the behavior of our model with the observed initial data. We start our simulation with $p_2 = 14$ since there is no data point for $n = 1$. As we

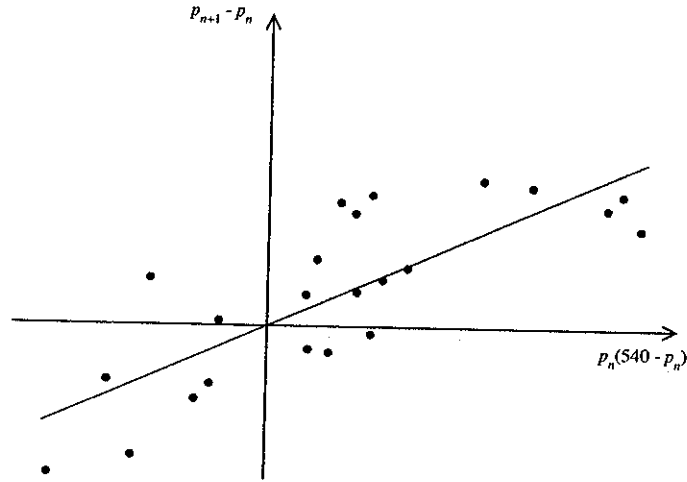


Figure 2.2. Testing (2.4) against the data shown in Table 2.1. The slope of the line of best fit is approximately 0.00145.

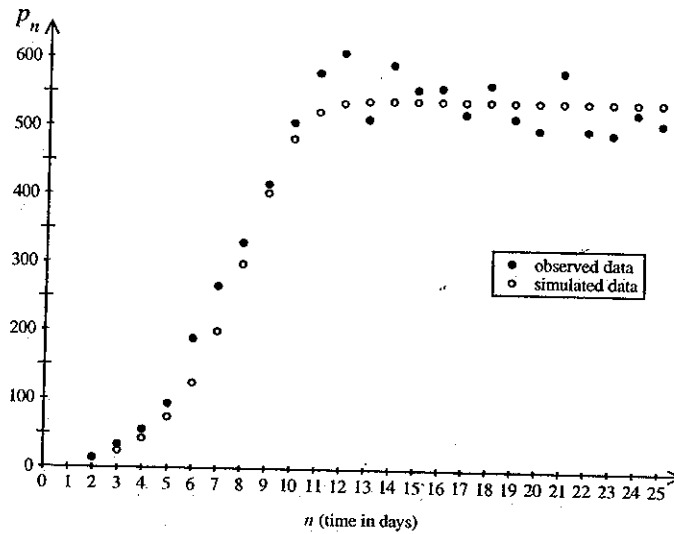


Figure 2.3. Comparison of the data simulated with (2.5) and the data observed by Gauss, from Table 2.1.

Using (2.5), we obtain the sequence $\{p_2, p_3, p_4, \dots\}$. The simulated data and the observed data are shown together in Figure 2.3. We see that the agreement looks good. Recall that the choice to use 540 in the model was rather arbitrary, and improvement in the fit may be possible by adjusting this number (see the exercises).

native
n, but

small
owing

(2.3)

when
ation:

(2.4)

time

$1 - p_n$
tant of
 p_n) and

assing
proxi-
mple).
lation:

(2.5)

initial
As we

In general, we can write the model just developed as

$$x_{n+1} = x_n + k(N - x_n)x_n, \quad (2.6)$$

where N is the maximum population that can be sustained by the environment. N often is referred to as the *carrying capacity* of the population. As we have just seen, this model can be used to describe the growth of a population in an environment with limited resources. The model can be used for other purposes as well. In particular, it can be used to describe the spread of an infectious disease, such as the flu or the common cold, through a small, closed population of size N . Here, x_n is the number of infected individuals after n time steps (e.g., days). Then $(N - x_n)$ is the number of individuals who have not yet become ill. The parameter k is a measure of the infectivity of the disease, as well as the contact rate between healthy and infected individuals. Similarly, the model can be used to describe the spread of a rumor through a population of size N . In this case, x_n is the number of individuals who have heard the rumor, and $N - x_n$ is the number of individuals who have not yet heard the rumor. The parameter k measures how juicy the rumor is. The larger k , the juicier the rumor, and the faster its spread through the population.

Equation (2.6) generally is rewritten as follows:

$$\begin{aligned} x_{n+1} &= x_n + k(N - x_n)x_n \\ &= (1 + kN)x_n - kx_n^2 \\ &= (1 + kN) \left(1 - \frac{k}{1 + kN} x_n \right) x_n \\ &= (1 + kN) \left(1 - \frac{x_n}{(1 + kN)/k} \right) x_n \\ &= r \left(1 - \frac{x_n}{K} \right) x_n, \end{aligned} \quad (2.7)$$

where

$$r = 1 + kN, \quad (2.8)$$

$$K = \frac{1 + kN}{k}. \quad (2.9)$$

Since this model is similar in appearance to the continuous-time model known as the logistic model or the Verhulst model (you will encounter this model in Section 3.1), the model here is known as the *discrete logistic model*.

Although the discrete logistic model provides a nice fit to Gause's data, it has the unfortunate (but mathematically interesting) property that it does not exhibit logistic growth (exponential growth initially, after which growth levels off until the population's carrying capacity is reached) for all choices of the model parameters.

In Section 2.2.4, we will discuss possible alternatives to the discrete logistic model. We first make a mathematical detour. In Section 2.2.2, we introduce techniques that can be used to analyze discrete-time equations of the form

$$x_{n+1} = f(x_n).$$

In Section 2.2.3, we use these techniques to explore in some detail the dynamical behavior of the discrete logistic equation in various parameter regimes.

 x_{n+1}

Figure
used are $r = 2$

2.2.2 Cob

With the fast
conditions and
it is easy to mi
dynamical from
a few of these

We begin
quickly visually
and every iterat

We demon
the function x_n

the form $x_n = A$
 $f(x_n) = f(A)$, wh

the form $x_n = 0$
can be obtained
previously have

the form $x_n = 1$
the form $x_n = 0$
the form $x_n = 1$

the form $x_n = 0$
the form $x_n = 1$
the form $x_n = 0$

(2.6)
 N often is
 model can
 resources.
 o describe
 h a small,
 ter n time
 et become
 he contact
 o describe
 number of
 ; who have
 e larger k ,

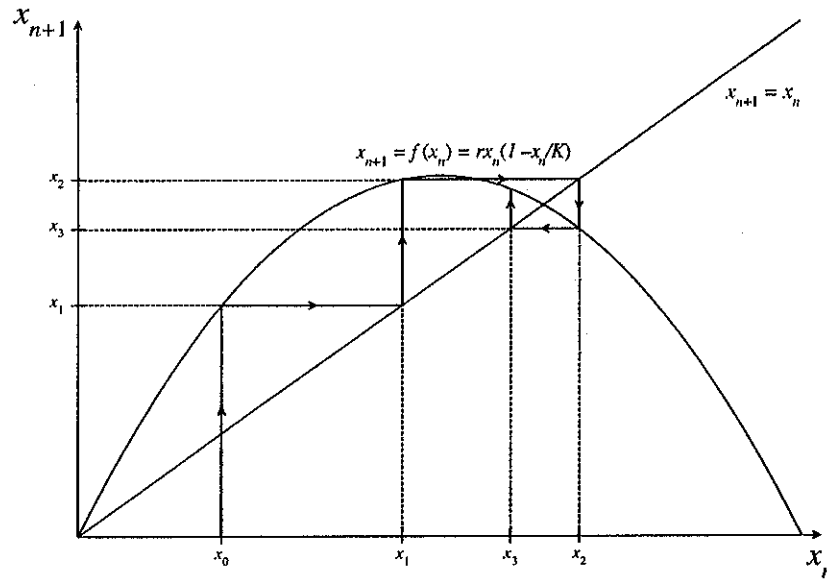


Figure 2.4. Cobwebbing for the discrete logistic model (2.7). Parameter values used are $r = 2.8$ and $K = 1$.

2.2.2 Cobwebbing, Fixed Points, and Linear Stability Analysis

With the fast computers of today, it is easy to generate many orbits by varying initial conditions and model parameters, and get a feel for the dynamics of the model. However, it is easy to miss some subtle behavior. We often can gain valuable insight into the model dynamics from sophisticated, but easy-to-learn, mathematical techniques. We will examine a few of these techniques in this section.

We begin with *cobwebbing*, which is a graphical solution method allowing one to quickly visualize the orbits and their long-term behavior without explicitly calculating each and every iterate along the way.

We demonstrate the cobwebbing technique in Figure 2.4, which shows the graphs of the function $x_{n+1} = f(x_n) = rx_n(1 - \frac{x_n}{K})$, using $r = 2.8$ and $K = 1$, and the straight line $x_{n+1} = x_n$. We choose our first iterate, x_0 , on the horizontal axis. The next iterate is $x_1 = f(x_0)$, which we can just read off the parabola. Visually, this is shown by a vertical line from x_0 on the horizontal axis to the point (x_0, x_1) on the parabola. The next iterate, x_2 , can be obtained in a similar way from x_1 . We first need to locate x_1 on the horizontal axis. We already have x_1 on the vertical axis, and the easiest way to get it onto the horizontal axis is to reflect it through the diagonal line $x_{n+1} = x_n$. Visually, this is shown by a horizontal line from x_1 on the vertical axis to point (x_1, x_1) on the diagonal line, and then a vertical line from point (x_1, x_1) on the diagonal line to x_1 on the horizontal axis. This process is repeated for subsequent iterates.

In summary, one starts by traveling from x_0 vertically to the parabola, then horizontally to the diagonal line, vertically to the parabola, and so on, as indicated by the solid portion of

is the logistic
e model here

ta, it has the
gistic growth
on's carrying

gistic model.
es that can be

ical behavior

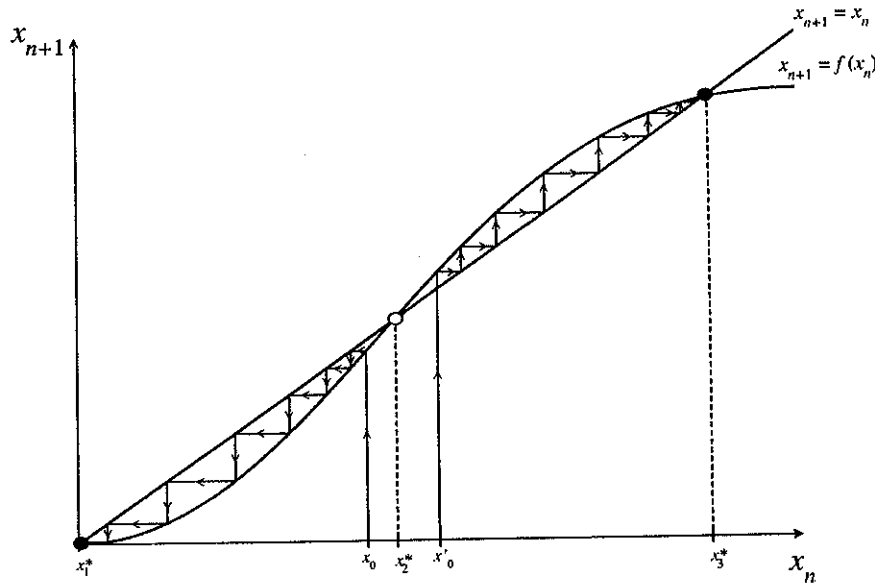


Figure 2.5. Illustration of stable and unstable fixed points of the difference equation $x_{n+1} = f(x_n)$. The fixed points x_1^* and x_3^* are stable (indicated by a filled circle), and the fixed point x_2^* is unstable (indicated by an open circle).

the vertical and horizontal lines on the cobwebbing diagram in Figure 2.4. In this particular case, the orbit converges to the rightmost intersection of the parabola and the diagonal line.

Any intersection of the parabola and the diagonal line represents a special point. Let x^* be such a point. Then $f(x^*) = x^*$. We call any such point a *fixed point* (or an *equilibrium point* or a *steady state*) of the model. If any iterate is x^* , then all subsequent iterates also are x^* . A question of interest is, what happens when an iterate is close to, but not exactly at, a fixed point? Do subsequent iterates move closer to the fixed point or further away? In the former case, the fixed point is said to be *stable*, whereas in the latter case, the fixed point is said to be *unstable*.

Examples of both stable and unstable fixed points are shown in Figure 2.5. The three fixed points shown are x_1^* , x_2^* , and x_3^* . Choosing an initial condition x_0 just to the left of x_2^* , we see that the orbit moves away from x_2^* , and towards x_1^* . Similarly, choosing the initial condition x_0 just to the right of x_2^* , we see that the orbit again moves away from x_2^* , but now towards x_3^* . Choosing the initial condition x_0 near x_1^* or x_3^* results in the orbit moving towards x_1^* or x_3^* , respectively. We say that x_2^* is an unstable fixed point of the model $x_{n+1} = f(x_n)$, and x_1^* and x_3^* are stable fixed points.

From Figure 2.5, note that the slope of f at the stable fixed points x_1^* and x_3^* is less than 1 (the slope of the straight diagonal line), whereas the slope of f at the unstable fixed point x_2^* is greater than 1. We can formalize these ideas via a *linear stability analysis*.

We choose the n th iterate to be close to a fixed point x^* of (2.1),

$$x_n = x^* + y_n, \quad (2.10)$$

with y_n small, now is, what happens to the deviation grows or shrinks. We can

We expand the function $R_2(y_n)$, to obtain

Since x^* is a fixed point, we neglect all terms of order y_n^2 or higher, then we are left with

We recognize that this is a linear difference equation

We can write y_n as

The behavior of the fixed point is determined by the eigenvalue λ .

$$\begin{aligned} \lambda &> 1 \\ 0 &< \lambda < 1 \\ -1 &< \lambda < 0 \\ \lambda &< -1 \end{aligned}$$

The four cases are: 1) $\lambda > 1$: instability of the fixed point; 2) $0 < \lambda < 1$: stability of the fixed point; 3) $-1 < \lambda < 0$: stability of the fixed point; 4) $\lambda < -1$: instability of the fixed point. This is referred to as the *Koopman operator*.

More generally,

Theorem 2.1. Let x^* be a fixed point of (2.1). Then

- 1) x^* is stable if $|\lambda| < 1$.
- 2) x^* is unstable if $|\lambda| > 1$.
- 3) there is no stability if $|\lambda| = 1$.

with y_n small, so that x_n can be thought of as a perturbation of x^* . The question of interest now is, what happens to y_n , the deviation of x_n from x^* , as the map is iterated? If the deviation grows, then the fixed point x^* is unstable, and if the deviation decays, then it is stable. We can find the equation for the deviation by substituting (2.10) into (2.1) to obtain

$$x^* + y_{n+1} = f(x^* + y_n). \quad (2.11)$$

We expand the right-hand side using a Taylor series about x^* , with a remainder term of $R_2(y_n)$, to obtain

$$x^* + y_{n+1} = f(x^*) + f'(x^*)y_n + R_2(y_n). \quad (2.12)$$

Since x^* is a fixed point, we can replace $f(x^*)$ on the right-hand side by x^* . If, in addition, we neglect all the terms in the Taylor series that have been collected in the term $R_2(y_n)$, then we are left with the following equation for the deviation:

$$y_{n+1} = f'(x^*)y_n. \quad (2.13)$$

We recognize that $f'(x^*)$ is some constant, say λ . The equation for the deviation is thus the linear difference equation

$$y_{n+1} = \lambda y_n. \quad (2.14)$$

We can write y_{n+1} explicitly in terms of λ and the initial condition y_0 :

$$\begin{aligned} y_1 &= \lambda y_0, \\ y_2 &= \lambda y_1 = \lambda(\lambda y_0) = \lambda^2 y_0, \\ &\vdots \\ y_n &= \lambda^n y_0. \end{aligned}$$

The behavior of the deviation y_n , and the subsequent conclusion regarding the stability of the fixed point x^* , can be summarized as follows:

- $\lambda > 1$: geometric growth; fixed point x^* is unstable;
- $0 < \lambda < 1$: geometric decay; fixed point x^* is stable;
- $-1 < \lambda < 0$: geometric decay with sign switch; fixed point x^* is stable;
- $\lambda < -1$: geometric growth with sign switch; fixed point x^* is unstable.

The four cases are illustrated in Figure 2.6. Note that no conclusion can be reached about the stability of the fixed point x^* when $\lambda = \pm 1$. These two cases require advanced treatment, involving a careful examination of the neglected terms that were collected in the term $R_2(y_n)$ in (2.12), which is beyond the scope of this book. For treatment of these cases, the reader is referred to Kuznetsov [104].

More generally, we can summarize the results of the analysis in the following theorem.

Theorem 2.1. *Let x^* be a fixed point of $x_{n+1} = f(x_n)$. Then,*

- x^* is stable when $|f'(x^*)| < 1$;
- x^* is unstable when $|f'(x^*)| > 1$;
- there is no conclusion about the stability of x^* when $|f'(x^*)| = 1$.

(2.10)

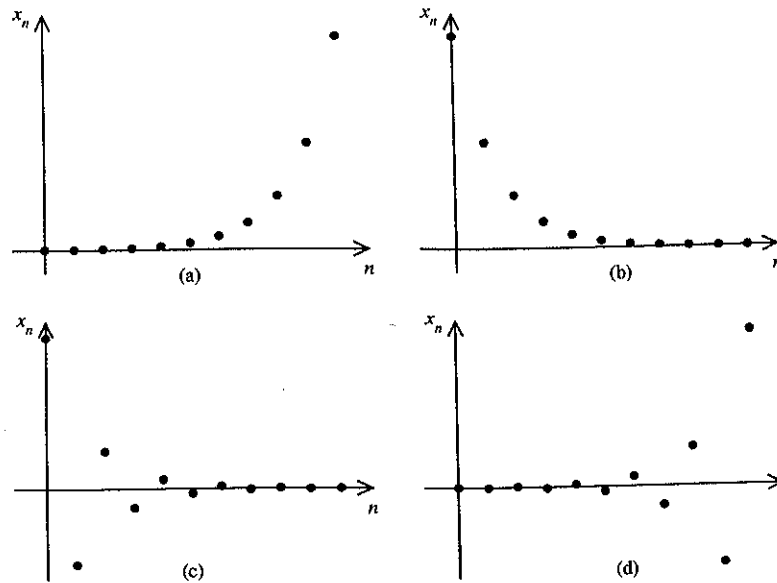


Figure 2.6. Behavior of the general linear difference equation, (2.14), as a function of the iterates n , for the cases (a) $\lambda > 1$; (b) $0 < \lambda < 1$; (c) $-1 < \lambda < 0$; (d) $\lambda < -1$.

That is, the linear stability of a fixed point x^* is determined by the slope of the map $f(x)$ at the fixed point, as intuited earlier. The parameter $\lambda = f'(x^*)$ generally is referred to as the *eigenvalue* of the map at x^* .

2.2.3 Analysis of the Discrete Logistic Equation

We now return to the discrete logistic equation, (2.7), and apply the tools discussed in the previous section. We begin with eliminating the parameter K by using the transformation $\bar{x}_n = \frac{x_n}{K}$ to obtain, after dropping the overbars,

$$x_{n+1} = f(x_n) = rx_n(1 - x_n). \quad (2.15)$$

Note that if we have $x_n > 1$, then $x_{n+1} < 0$. To avoid such situations, we impose the restriction $0 \leq r \leq 4$ (can you think of the reason why this should be so?), so that $x_n \in [0, 1]$ for all n provided $x_0 \in [0, 1]$.

The fixed points of the map can be found exactly by setting $f(x^*) = x^*$ and solving for x^* . There are two fixed points. The trivial fixed point, $x^* = 0$, always exists, while the nontrivial fixed point, $x^* = \frac{r-1}{r}$, is positive only when $r > 1$.

To determine the stability of the fixed points, we need $f'(x)$, which is

$$f'(x) = r(1 - 2x). \quad (2.16)$$

At the trivial fixed point, $x^* = 0$, the eigenvalue is $f'(0) = r$. That is, the trivial fixed point is stable for $0 \leq r < 1$ and unstable for $1 < r \leq 4$. At the nontrivial fixed point, $x^* = \frac{r-1}{r}$,

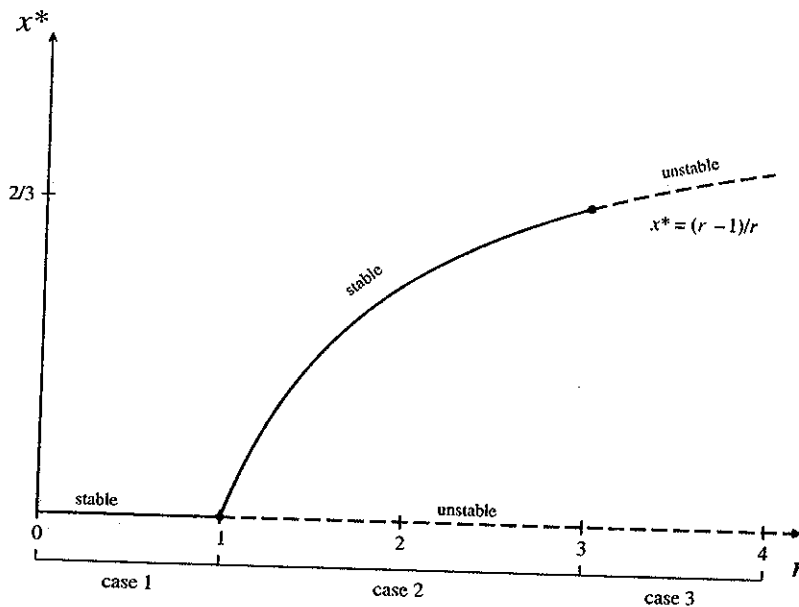


Figure 2.7. Partial bifurcation diagram for the rescaled discrete logistic equation, (2.15). Shown are the fixed points and their stability as a function of the model parameter r . Solid lines indicate stability of the fixed point, and dashed lines indicate instability. The filled circles represent bifurcation points.

the eigenvalue is $f'(\frac{r-1}{r}) = 2 - r$. That is, the nontrivial fixed point is stable for $1 < r < 3$, and unstable for $3 < r \leq 4$.

The existence and stability of the fixed points is summarized in the bifurcation diagram of the fixed points versus the parameter r , shown in Figure 2.7. Reading the diagram from left to right, note that the trivial fixed point becomes unstable as soon as the nontrivial fixed points come onto the scene at $r = 1$, when the eigenvalue moves through $+1$. The nontrivial fixed point is stable initially, but loses its stability at $r = 3$, when the eigenvalue moves through -1 .

The two points $r = 1$ and $r = 3$ are known as *bifurcation points*. A bifurcation point is a parameter value at which there is a qualitative change in the dynamics of the map. The bifurcation at $r = 1$ is called a *transcritical bifurcation*, referring to an exchange of stability when two branches of fixed points meet (the two branches meeting here are $x^* = 0$ and $x^* = \frac{r-1}{r}$). The bifurcation at $r = 3$ is called a *flip bifurcation* or a *period-doubling bifurcation*. We will see shortly how the dynamics of the map changes at this flip bifurcation.

There are many other types of bifurcations. A detailed discussion of bifurcation theory is beyond the scope of this book, and the interested reader is referred to Alligood, Sauer, and Yorke [4], Kuznetsov [104], and Strogatz [152].

We can easily read the long-term behavior of the logistic map from the bifurcation diagram. As before, let us think of x_n as the size of a population (now scaled by the factor K). We can distinguish three cases (indicated along the bottom of Figure 2.7). In the first case, for $0 \leq r < 1$, the population goes extinct, no matter what the size of the initial population,

function
-1.
ie map
ferred

l in the
mation

(2.15)

ose the
[0, 1]

olving
ile the

(2.16)

d point
= $\frac{r-1}{r}$,

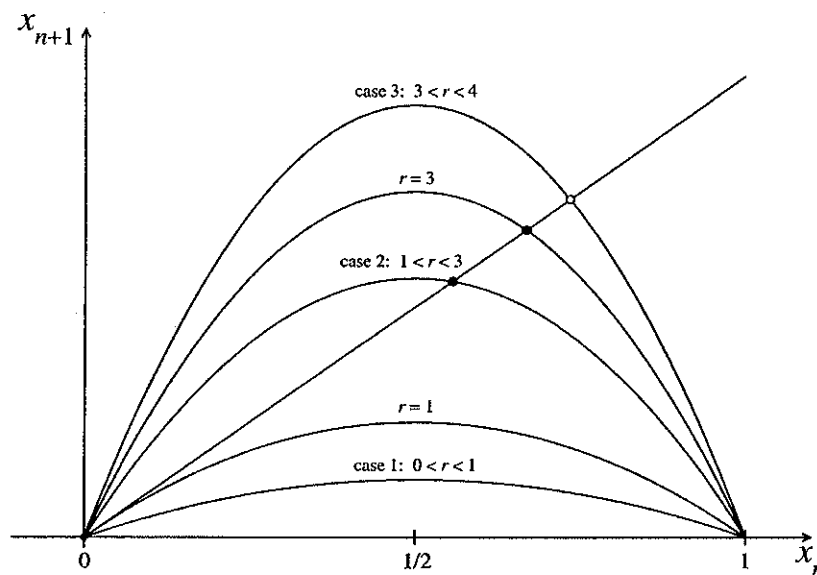


Figure 2.8. Dependence of the shape of the parabola $f(x_n) = rx_n(1 - x_n)$ on the value of the model parameter r .

x_0 , is. In the second case, for values of r between 1 and 3, the population reaches a nonzero steady state. The larger the value of r , the larger the steady-state population. What happens when the parameter exceeds 3 is not clear. Before investigating this third case, however, it pays to perform a graphical analysis complementing the results from the linear stability analysis.

Figure 2.8 shows how the shape of the parabola $f(x_n) = rx_n(1 - x_n)$ depends on the value of the model parameter r . Note that the roots remain fixed at $x_n = 0$ and $x_n = 1$. However, the maximum of the parabola is $\frac{r}{4}$, and thus increases with r .

For the first case, $0 < r < 1$, the parabola lies entirely below the diagonal line $x_{n+1} = x_n$, and the only point of intersection is at the origin. That is, the only fixed point is the trivial fixed point. Since the slope of f at the origin clearly is positive but less than 1, the trivial fixed point is stable. Any population will go extinct, eventually. This situation is illustrated in Figures 2.9 (a) and (b).

When $r = 1$, the parabola is tangent to the diagonal line $x_{n+1} = x_n$ at the origin. This event marks the transition to the second case. As soon as $r > 1$, the slope of f at the origin exceeds 1 (i.e., the fixed point at the origin has switched from being stable to unstable) and there is an additional point of intersection, namely, the nontrivial fixed point, $x^* = \frac{r-1}{r}$. The slope of f at the nontrivial fixed point is always less than 1. Initially, for $1 < r < 3$ (case 2), the slope of f at the nontrivial fixed point is greater than -1 , and so the fixed point is stable. Any population will eventually reach a steady-state size. This situation is illustrated in Figures 2.9 (c) and (d).

When $r = 3$, the slope of f at the nontrivial fixed point is -1 , and this event marks the transition to the third case. When $3 < r \leq 4$ (case 3), the slope of f at the nontrivial

$\bar{x} = 0$ is
The fixed
size sta

fixed po
from lin
W
pens in
for three
of the li
diagram
shown in

In
oscillate
(d), for
size. Va
16-cyc
an aperio
chaotic, a
mathemat
book, and
(192) for
the simpl

We
converge

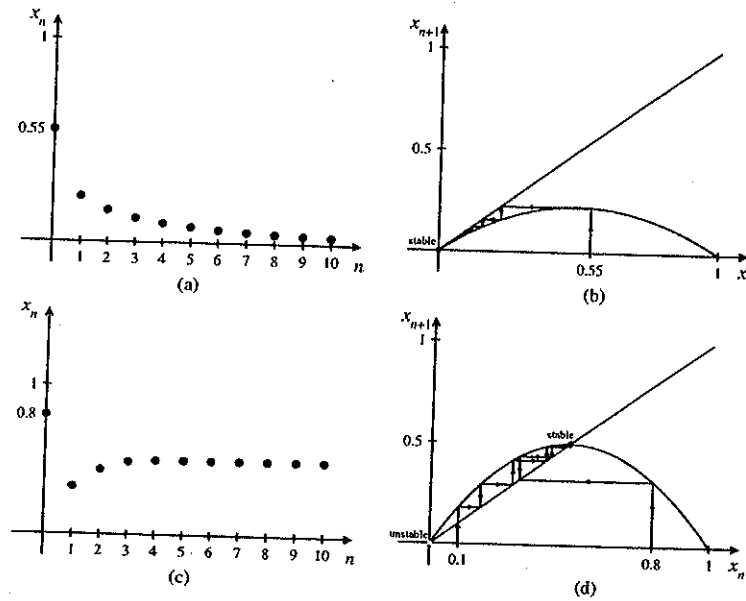


Figure 2.9. (a) and (b) Case 1 ($0 < r < 1$), for $r = 0.9$. The only fixed point $\bar{x} = 0$ is stable, and the population goes extinct. (c) and (d) Case 2 ($1 < r < 3$), for $r = 2$. The fixed point $\bar{x} = 0$ is unstable, the nontrivial fixed point is stable, and the population size stabilizes.

fixed point is less than -1 , and so the fixed point now is unstable, as we had inferred earlier from linear stability analysis (Theorem 2.1).

We now continue with the graphical analysis and cobwebbing to determine what happens in the third case. In Figure 2.10, we show the dynamics of the discrete logistic equation for three values of r between 3 and 4. The plots in the left column of Figure 2.10 show values of the iterates x_n as a function of n for various values of r . Corresponding cobwebbing diagrams are shown in the right column of Figure 2.10 (to clarify the cobwebbing diagrams shown in (b) and (d), only the last few iterates are used).

In Figures 2.10 (a) and (b), for $r = 3.2$, we observe that the population eventually oscillates between two values. We refer to the oscillation as a *2-cycle*. In Figures 2.10 (c) and (d), for $r = 3.55$, we eventually observe a *4-cycle*, or an oscillation between four population sizes. Values of r can be found at which the discrete logistic equation exhibits an *8-cycle*, a *16-cycle*, and so on. But not all values of $r > 3$ give periodic oscillations. An example of an aperiodic oscillation is shown in Figures 2.10 (e) and (f), for $r = 3.88$. The orbit appears chaotic, and indeed, it can be shown that the discrete logistic equation exhibits *chaos* in the mathematical sense. A careful mathematical definition of chaos is beyond the scope of this book, and the interested reader is referred to Alligood, Sauer, and Yorke [4] and Strogatz [152] for more information. For the purposes of our discussion, it suffices to observe that the simple model under investigation can exhibit some very complicated dynamics.

We can broaden our analysis to determine the origin of the 2-cycle. When an orbit converges to a 2-cycle, it oscillates between two values, say u and v (see Figure 2.10 (b)),

), on the
 nonzero
 happens
 however,
 stability
 is on the
 $x_n = 1$.
 onal line
 l point is
 s than 1,
 ation is
 gin. This
 ne origin
 ble) and
 $= \frac{r-1}{r}$.
 $< r < 3$
 the fixed
 ation is
 nt marks
 ontrivial

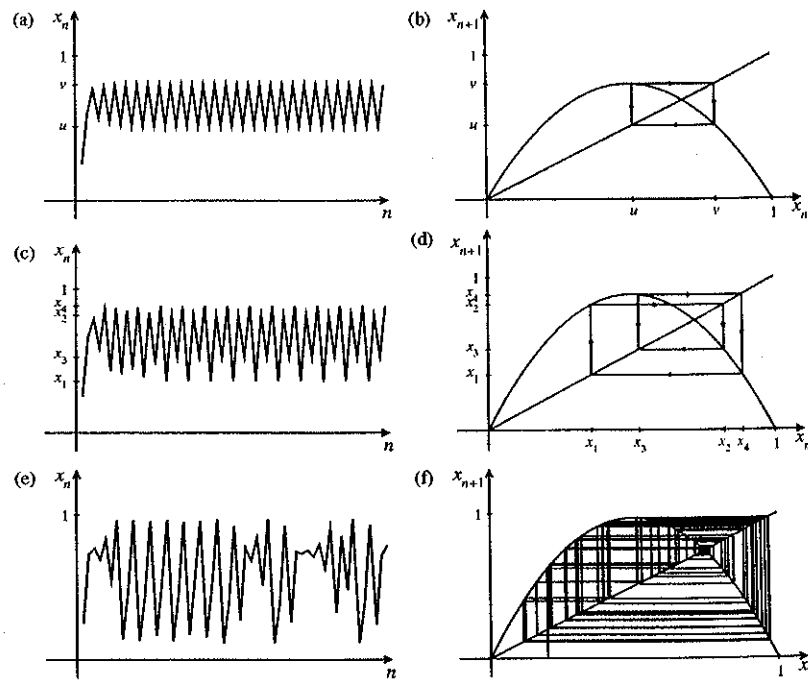


Figure 2.10. Illustration of the various types of dynamical behavior of the discrete logistic equation in case 3, when $3 < r \leq 4$. (a) and (b) Two-cycle with $r = 3.2$. (c) and (d) Four-cycle with $r = 3.55$. (e) and (f) Chaos with $r = 3.88$.

with

$$f(u) = v, \quad (2.17)$$

$$f(v) = u, \quad (2.18)$$

or, equivalently,

$$f(f(u)) = u, \quad (2.19)$$

$$f(f(v)) = v. \quad (2.20)$$

Recalling the definition of a fixed point (x is a fixed point of $f(x)$ if $f(x) = x$), we see that the above equations imply that u and v are fixed points of the *second-iterate map*, $f(f(x)) = f^2(x)$.

The graph of the second-iterate map f^2 is shown in Figure 2.11 for various values of the parameter r . For values of $r < 3$ (Figure 2.11 (a)), the second-iterate map has two fixed points, namely, the origin, which is unstable, and the nontrivial fixed point, $x^* = \frac{r-1}{r+2}$ of the original logistic map, which is stable (note that any fixed point of the logistic map automatically also is a fixed point of the second-iterate map). That is, no interesting 2-cycles exist for these values of r . As r increases, the maxima of the second-iterate map rise and

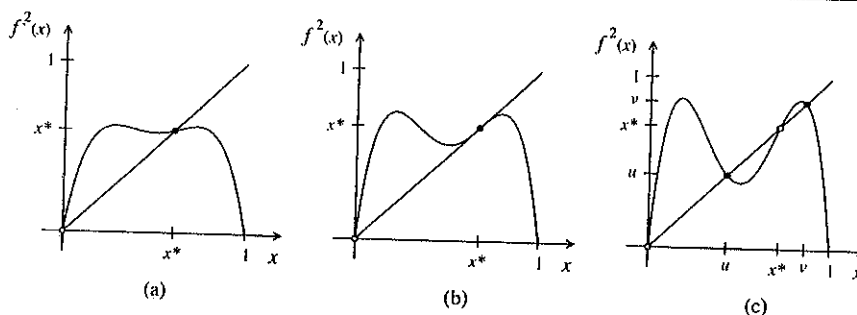


Figure 2.11. The second-iterate function, $x_{n+2} = f^2(x_n)$, for the logistic map, for various values of r . (a) $r < 3$; (b) $r = 3$; (c) $r > 3$.

the local minimum descends, until at $r = 3$ (Figure 2.11 (b)), the local minimum is tangent to the diagonal line $x_{n+2} = x_n$. At this point, two new stable fixed points of the second-iterate map emerge, namely, u and v , corresponding to the 2-cycle (Figure 2.11 (c)). At the same time, the nontrivial fixed point x^* becomes unstable (this is in accordance with our findings from the analysis of the original logistic map). The bifurcation at $r = 3$ is called a *period-doubling* or *flip bifurcation*.

Initially, u and v are close together, so the 2-cycle is barely noticeable. But as r increases, u and v move away from each other, and the 2-cycle becomes more pronounced. The stability of u and v corresponds to the stability of the 2-cycle. That is, the 2-cycle is stable initially, since the graph of f^2 at u and v is shallow. As r increases beyond $1 + \sqrt{6}$ (see Exercise 2.4.6), the slope of f^2 at u and v becomes less than -1 , indicating that the 2-cycle becomes unstable. At this point (another flip bifurcation), the 4-cycle arises. We could continue the analysis by graphing f^4 for various values of r , but this is left as an exercise for the reader.

We can update the bifurcation diagram shown in Figure 2.7 by including information about the 2-cycle, as shown in Figure 2.12.

Ideally, we should also include information about the 4-cycle, the 8-cycle, and so on. The algebra to do so becomes unwieldy rather quickly. However, we can use the computer to create a similar diagram. The idea is to let the computer program determine the long-term behavior of the map for many values of the parameter r . For example, for $r = 2$, the iterates converge to $\bar{x} = \frac{1}{2}$, the stable fixed point of the map for this value of r . If we had computed 2000 iterates, say, from an arbitrary initial condition, then the last 100 or so iterates will all have a value virtually indistinguishable from $\frac{1}{2}$. So, plotting these last 100 iterates above $r = 2$ on a diagram of \bar{x} versus r just gives a point, $(r, \bar{x}) = (2, \frac{1}{2})$. If we choose $r = 3.2$ (cf. Figure 2.10 (a)), the last 100 iterates or so will jump back and forth between the values of the corresponding 2-cycle. Plotting these iterates above $r = 3.2$ gives two points, and so on. A lot of computation, using many values of r close together, eventually leads to the *orbital bifurcation diagram*, also known as the *Feigenbaum diagram*, shown in Figure 2.13. Note that since the computations only detect stable behavior (stable fixed points, stable 2-cycles, and so on), the orbital bifurcation diagram differs from the bifurcation diagrams shown in Figures 2.7 and 2.12 in that the branches of unstable behavior, indicated by dashed lines, no longer are shown.

discrete
(c) and

(2.17)

(2.18)

(2.19)

(2.20)

we see
the map,

values
has two
 $= \frac{r-1}{r}$,
tic map
2-cycles
rise and

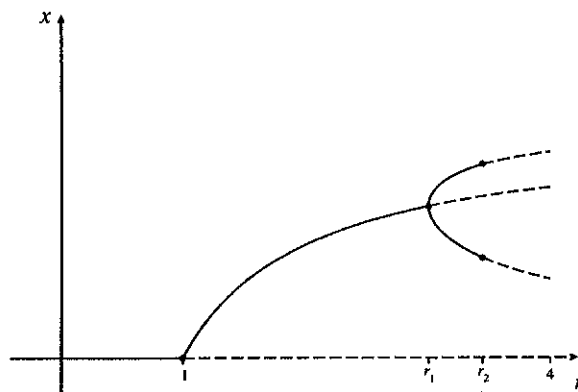


Figure 2.12. Updated bifurcation diagram for the discrete logistic equation shown earlier in Figure 2.7. Shown are the fixed points, as well as the 2-cycle for values of $r > r_1 = 3$. The 2-cycle is stable up to $r_2 = 1 + \sqrt{6}$, and unstable thereafter.

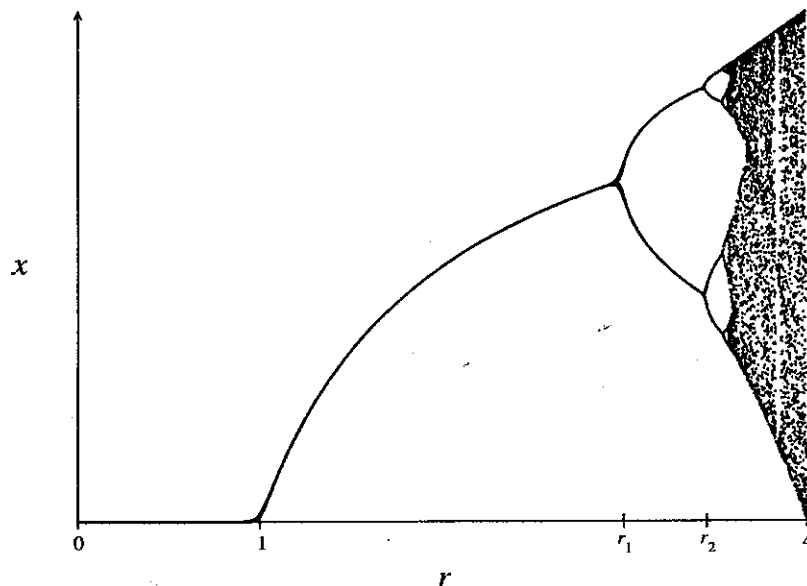


Figure 2.13. Orbital bifurcation diagram for the discrete logistic equation.

By examining the orbital bifurcation diagram, it can be seen that the 4-cycle exists only over a small range of r , the 8-cycle over an even smaller range of r , etc. It can be shown (see, e.g., Holmgren [89]) that the bifurcation points leading to higher-order cycles converge at $r \approx 3.57$. Beyond $r \approx 3.57$, the logistic map becomes chaotic, that is, the iterates no longer appear to follow a predictable pattern, although they are confined to take on only certain values (e.g., when $r = 3.6$, the iterates never take on values below 0.14).

or above 0.900, but they can take on any intermediate value). We say that the attractor of the map is a *strange attractor*. By computing the orbital bifurcation diagram over a smaller range of r , say for $3 < r < 4$, with higher resolution, many interesting features of the map can be observed, such as *periodic windows* surrounded by chaos. A periodic window is a small range of r where the attractor is periodic again. For example, near $r = 3.83$, one can find 3-cycles. From this periodic window, the transition back to chaos occurs through a series of period-doubling or flip bifurcations, leading to 6-cycles, 12-cycles, and so on.

The discrete logistic equation is a well-studied difference equation, and there are many interesting mathematical investigations that can be pursued. We will stop here and refer the interested reader to Devaney [43] and Strogatz [152].

In the 1970s, May [113] noticed that simple difference equations can give rise to very complicated dynamics. He hypothesized that the wild fluctuations observed in some natural populations might reflect chaotic orbits of low-dimensional systems of difference equations. J. M. Cushing and his colleagues [41] have followed up on this hypothesis. They conducted controlled experiments on laboratory populations of flour beetles living under constant environmental conditions. They showed that the population dynamics can be described and predicted accurately by a relatively simple model of difference equations that reflect well-understood facts about the life cycle of the flour beetle. In addition, they were able to explain observed dynamics of the population and demonstrate nonlinear phenomena such as bifurcations, periodic orbits, and chaos in real biological populations. Thus, their work has lent credibility to the use of models such as those described in this chapter. It should be no surprise, then, that these models continue to be used on a regular basis.

2.2.4 Alternatives to the Discrete Logistic Equation

In the previous section, we saw that the behavior of the discrete logistic equation, (2.7), is quite complex. For many choices of the model parameters, the solution does not exhibit logistic growth (exponential growth initially, followed by a leveling off of the growth rate, until the population reaches a steady state). Also, for some choices of parameter values, the model gives unrealistic results. For example, if the population $x_n > K$ in any year, then the population is extinct (negative) the next year. For these reasons, it is worthwhile to examine alternative models that do not have these problems and that are widely used.

In particular, we examine the Beverton–Holt and Ricker models. Before discussing these models in detail, we note that these two models and the discrete logistic equation belong to a class of models that can be written in the following general form:

$$x_{n+1} = f(x_n) = g(x_n)x_n. \quad (2.21)$$

Many other population models can be written in this form. Of course, the simplest model belonging to this class is the *geometric growth model*, using $g(x_n) = r$, so that

$$x_{n+1} = rx_n, \quad (2.22)$$

where $r > 0$. We have encountered this model previously, as (2.14) in the discussion of linear stability analysis. When $r > 1$, $x_n \rightarrow \infty$ as $n \rightarrow \infty$; when $r < 1$, $x_n \rightarrow 0$ as $n \rightarrow \infty$. In this case, the growth rate $g(x_n) = r$ is constant; that is, the number of offspring

wn
of

exists
an be
cycles
is, the
o take
0.324