

# Mathematics for the Life Sciences: Calculus, Modeling, Probability, and Dynamical Systems

## Instructor Guide

As described in the preface, this book does not fit into a standard slot in the mathematics curriculum. The choice of topics is unique, and some of these topics are presented in a way intended to sacrifice thoroughness for accessibility. There are also a number of topics in mathematical modeling that have not been given a place in the standard curriculum; hence, many instructors who venture to use this book will find themselves having to teach topics that they have not seen as students or instructors. The purpose of this guide is to assist instructors who are new to mathematical modeling or skeptical about the possibility of teaching some advanced topics at a relatively elementary level.

## 1 A BRIEF SUMMARY OF CALCULUS

This chapter requires some flexibility on the part of any mathematician who teaches from it. I have shunned a number of standard practices in the teaching of calculus. Perhaps the most significant of these are my decisions not to cover limits, except in the most informal way, and to teach the definite integral without a formal presentation of Riemann sums.<sup>1</sup> These choices fly in the face of the way calculus has been taught for longer than anyone alive today can remember. But that does not mean they lack historical precedent. The first paper to introduce calculus was published by Leibniz in 1684, but the first paper to use Riemann sums was only published (by Riemann, of course) in 1854. Definite integrals were known and used for applications for 170 years before the modern definition. This long delay indicates two things: 1) the integral concept is easier to understand than its development via Riemann sums, and 2) a lot of use can be made of the integral by biologists who have never seen a Riemann sum. The situation with limits is similar. Leibniz and Newton had no difficulty doing calculus (as opposed to real analysis, about which neither knew anything), but 133 years passed before Bolzano first defined continuity, with the definition of the limit by Cauchy following in 1821.

Although it may sound as though I am altogether opposed to rigorous mathematics, this is not the case. I quote myself in the preface as saying that mathematicians are people who believe that you should not drive a car until you have built one yourself. To extend the metaphor, I think biologists will be better drivers if we spend all of our time giving them driving lessons rather than teaching them to build cars. This does not mean that I believe in teaching unmotivated mathematical techniques. The text includes informal justifications and plausibility arguments for almost everything.

The secret to preparing students to do advanced work in linear algebra and dynamical systems with a minimal calculus background is for that background to focus on concepts and modeling while deemphasizing rigor and technique. Throughout this chapter, instructors will

---

<sup>1</sup>Riemann sums are a much more reasonable topic than the formal definition of the limit; I have included a development of the definite integral with Riemann sums as Appendix B.

find opportunities for students to practice mathematical modeling in the problem sets. Extra time should be allowed for these; for example, they can be used as group activities in the classroom. One possibility is to spend an hour in the classroom to start a group project and then have the students do the remainder of the work on their own. The extended problems in some of the exercise sets make it a challenge for instructors to plan the amount of time needed for each section. It is important to be realistic about what students can do. This does not mean that we can't assign challenging problems, but we do need to slow the pace of the course to accommodate them. We can cover a section each day in lecture, but we can't expect students to be able to do both the routine problems and the challenging problems before the next class.

## 1.1 Working with Parameters

Rigor, which I have argued against, is not the only characteristic that sets mathematics apart from other disciplines. The essence of mathematics also includes generalization and abstraction, and both of these features are as essential in mathematical modeling as they are in mathematics proper. The standard mathematics curriculum does not give these features the attention that they deserve, a flaw that this section is intended to partially ameliorate. One key requirement of mathematical modeling is an understanding of the different roles symbols play in a given calculation. These roles can change with the context. As an example, the quantities  $x$  and  $y$  are variables in the general formula  $y = b + mx$ , while  $m$  and  $b$  are parameters. We fix the parameters and plot the variables, not the other way around. However, when we use the least squares method to fit the general formula to data, we do these calculations with  $m$  and  $b$  as variables and  $x_i$  and  $y_i$  as parameters. It is a mistake to underestimate the extent to which the distinction between variables and parameters is confused in most students' thinking. This is one of the reasons why many students cannot do even the simplest of word problems.

The purpose of this opening section is to introduce the variable-parameter distinction and give students lots of practice working with parameters. In both the examples and the problem sets, one must pay attention to the context in order to determine how a quantity should be treated, and it is sometimes necessary for a quantity to be treated as both a variable and a parameter in the same problem. Time spent reaching a solid understanding of the material in this section will be repaid many times over throughout the rest of the book. Resist the temptation to complete this section too quickly.

None of the problems require a lot of calculation. Several require extensive reading, and a number of them require graphs. In some cases, the viewing windows for the graphs are specified, but in other cases these are left for students to choose. Being able to select an appropriate viewing window for a graph is a critical modeling skill. If, for example, a variable can only take on nonnegative values, then the interval used for that variable in a graph should not include negative values. Calculator graphs are centered about the origin unless that default is overridden, so many students simply do what their calculator seems to require. If the independent variable is a probability, then the domain used for it in a graph should be at most  $[0, 1]$ . These things are so obvious to mathematicians that it is hard to appreciate the need to teach them to students. Interestingly, this is not a problem I have with graduate students in biology, who are generally much better at understanding

and using graphs than undergraduate mathematics students. Another common problem, for both undergraduates and biology graduate students, is using the range of data to determine the viewing window in conditions where this is not appropriate. Suppose, for example, the data for  $y$  falls in the interval  $[5, 6]$ . Using this interval for the viewing window may be appropriate in some circumstances, but usually it is far more meaningful to use the interval  $[0, 6]$ , which has the advantage of providing a visual display of relative differences in values as well as absolute differences. This first section is the best context in which to teach students the skill of obtaining a useful graph.

## 1.2 Rates of Change and the Derivative

This section presents the definition of the derivative in the broader context of rates of change. In practice, we can define average rates of change for functions of discrete time and both average and continuous rates of change for functions of continuous time. While the former is merely a review of basic algebra, it is included here partly to help students understand difference quotients for functions given via formulas and partly because discrete rates of change are important for the study of discrete dynamical systems that comes in Chapter 5.

## 1.3 Computing Derivatives

My treatment of derivative computation is essentially the same as that found in calculus books, but with a few differences. In keeping with the emphasis on using calculus rather than deriving results, the derivative formulas for elementary functions are presented with very little fanfare. The main focus of the section is on the product, quotient, and chain rules, which provide a means to reduce derivative computations for complicated functions to elementary derivative formulas. Where the typical calculus book has the better part of a chapter for differentiation techniques, the topic occupies just one section of this book. Students should know the basic derivative formulas so that they can work problems without getting stuck on the minor details, but they do not need to be experts at differentiation. None of the modeling problems in the remainder of the text require differentiation of anything more complicated than the examples in this section. The one unusual feature of my treatment of differentiation is that I introduce partial derivatives along with ordinary derivatives. Multivariable calculus is beyond the scope of this book; however, stability analysis in Chapter 7 requires students to be able to compute Jacobian matrices, which consist of partial derivatives. While the partial derivative concept is more advanced than the ordinary derivative concept, the computation of partial derivatives merely requires variables to be treated as parameters. The calculation in Example 1.3.9 is no more difficult than that of Example 1.3.5.

## 1.4 Local Behavior and Linear Approximation

Tangent lines, local extrema, and linear approximation are combined in this section. I link these topics by thinking of local behavior as whatever one sees when zooming in on a point on a graph, which will be the first nonconstant term in the Taylor series. Of course a full discussion of this point is beyond the scope of this book; however, it does provide a conceptual framework broad enough to incorporate concavity when the tangent line is

horizontal while ignoring it in other cases. Local extrema is a topic that is important on its own, but its use in the rest of the book is limited; in contrast, linear approximation is not particularly important on its own while being essential for analytical study of dynamical systems. Instructors will want to emphasize local extrema for courses that emphasize calculus and linear approximation for courses that will continue with dynamical systems. (Of course these are not mutually exclusive.)

## 1.5 Optimization

The amount of time spent on this section should depend strongly on the overall purpose of the course. For courses that emphasize calculus with modeling, it is worth spending some time on some of the more challenging word problems. Problems 1.5.4, 1.5.5, 1.5.8, and 1.5.9 are particularly difficult; instructors are cautioned not to assign very many of these. Each one could stand alone as a group project. Courses that emphasize probability or dynamical systems and are only doing calculus to prepare for those subjects should treat this section sparingly.

## 1.6 Related Rates

This section differs in significant ways from any section with the same name in a standard calculus book. It treats two related topics, neither including the usual geometric problems about sliding ladders and the like, but both of which are critical for mathematical modeling. The first deals with differentiation of equations that include more than one function, such as equations that relate time-dependent volume to time-dependent length. Many mathematical models are constructed in a similar fashion to the one in Example 1.6.3. There is a differential equation that describes the rate of change of one variable in terms of a different variable, augmented by an algebraic relation between the two variables. These can sometimes be combined by simple substitution, but it is often easier to differentiate the algebraic relation first.

The second topic is one that I use whenever I teach basic calculus, including standard courses for physical scientists and engineers. As it is generally presented, the chain rule is merely a mechanical exercise of differentiating a composition of functions. While that is certainly important, it misses the conceptual sense of the chain rule, which is critical for some applications. This conceptual sense is brought out by the insect-on-a-heated-rod scenario presented in Subsection 1.6.2. In this scenario and the following discussion, I have tried to supplement the mathematician's perspective of formulas-as-functions with an alternative perspective of formulas as relationships between variables. This is helpful in contexts where it is not clear which variable(s) should be independent and which dependent, such as in chemical thermodynamics, where a gas is characterized by a number of variables indicating its physical state, and any two of these can be considered as the independent variables for the rest. There are cases in biology where it is not immediately clear which of two variables should be taken as independent, particularly in modeling investigations in which the role of variable is played by model parameters.

Aside from the practical value of a treatment of the chain rule in terms of variables, it is important to make a connection between the chain rule as correctly written and the chain

rule in the form

$$\frac{du}{dt} = \frac{du}{dx} \frac{dx}{dt}$$

of (1.6.5). The latter version of the chain rule is much easier to remember and much more intuitive than the more formal version; however, it lacks the essential information that the derivative  $du/dx$  must be evaluated at the particular  $x(t)$  corresponding to the second factor. This is one of the more universal conceptual errors of calculus students, one which I feel needs to be addressed explicitly.

Problems 1.6.3 and 1.6.7 require quite a bit of work, but I particularly like Problem 1.6.7 because of the modeling issues it addresses. Problems 1.6.4 through 1.6.6 are much shorter, but students will find these to be difficult as well. Problem 1.6.6 anticipates nondimensionalization, a key topic in mathematical modeling that is treated in some detail in Section 2.6.

## 1.7 Accumulation and the Definite Integral

This section focuses on the concept of the definite integral, with a particular emphasis on the ideas of integrals of rates as total accumulation and integrals of gradients as total aggregation. I emphasize these applications as constituting the meaning of the integral and the approximation with finite sums as a method for estimating its value rather than the more mathematical approach of using the definition in terms of sums to confer meaning and thinking of the applications as consequences of the definition. The traditional geometric applications such as areas between curves and volumes are omitted here. These geometric problems are nice if the goal is to present applications of integration to a general audience. For biologists, these geometric applications have much less value than applications in population dynamics and other biological areas. The exercises for this section include some numerical approximation and some derivation of integrals to calculate specific quantities.

## 1.8 Properties of the Definite Integral

This section presents the properties of the definite integral and computation of the definite integral via antiderivatives. It is similar to the treatment that appears in standard calculus books, with one major exception. Rather than proving the fundamental theorem, I prefer to derive it by setting the integral of the rate of change equal to the total change. This derivation is simple and intuitively clear.

## 1.9 Computing Antiderivatives and Definite Integrals

This section combines substitution with the part of the fundamental theorem that addresses differentiation of integrals. The substitution method is necessary if any integrals are going to be done by hand. Given the audience for this book, I have chosen not to include integration by parts or other integration techniques. In today's world, biologists can use Maple or Mathematica for these tasks. I feel that it is important to understand the ideas of definite integrals and to be able to do the elementary calculations, but not the more complicated calculations. Time saved by not studying these topics is then available to do more modeling.

In this section, as elsewhere in this chapter, technical topics have been deemphasized to make room for modeling topics. One example of this is Problem 1.9.12, which consists of an optimization problem with an embedded definite integral and concludes with questions about the biological claims inherent in the results. This exercise uses a result from Problem 1.9.11, but that result is included in the statement of Problem 1.9.12; hence, it is not necessary to assign the former in order to be able to do the latter.

## 2 Mathematical Modeling

This chapter presents a broad overview of the modeling process along with specific topics in mechanistic and empirical modeling. There is a tendency among modelers to focus on just one of these two types, but I have tried to present a balanced view. As far as I am aware, the treatment here is the only comprehensive presentation of mathematical modeling in print. This actually represents a deviation from my personal proclivities, which come down hard on the mechanistic side. I recognize, however, that students get a much more useful and complete view of modeling from a serious study of both. Mechanistic modeling, particularly Section 2.6, is necessary as background for the dynamical systems portion of the book. Empirical modeling is the only alternative for settings in which we do not know enough to construct a mechanistic model. One reason why mechanistic and empirical modeling are not generally considered together is that they lie on opposite sides of a standard boundary in the mathematical sciences, that between mathematics and statistics. Empirical modeling topics, such as least squares analysis, appear primarily in statistics books. In my view, it is an error for mathematicians to allow empirical modeling to be classified as statistics rather than mathematics. Indeed, one can sometimes find a fully mathematical treatment of least squares in linear algebra books. Additionally, statisticians do not give parameterization and selection of models much emphasis in their courses. At the University of Nebraska, the introductory statistics course barely touches least squares analysis; parameterization of nonlinear models and AIC for selecting among models are not found at all. Nor have I seen AIC in any elementary statistics books, for that matter. For the most part, statistics is concerned with the limited goal of identifying relationships rather than the more ambitious goal of quantifying relationships. My view is that the statisticians are right about this emphasis, and that we mathematicians should take on the quantification of relationships as part of our job.

### 2.1 Mathematics in Biology

This introductory section presents readers with the central problem of modeling in biology: the difficulty of finding trends that are largely obscured by scatter in data. The combination of demographic stochasticity with limited sample sizes mean that biological data is almost always problematic. Some areas in biology, such as ecology, are also subject to significant measurement error, but even areas such as medicine, where measurement error is small, are significantly affected by demographic stochasticity. It is difficult to understand the full importance of this principle. For me, there are two ways of seeing its significance. One is to compare the subjects of biology, which are living organisms, with the subjects of chemistry,

which are molecules. If we measure radioactivity using a sample size of atoms that is as small as the sample size for a medical study, we can see the effect of stochasticity in a case where there is no systematic variation among individuals. The systematic individual variation of biological organisms only magnifies the effect. We don't see the problem of stochasticity in radioactivity, however, because the sample sizes are unimaginably large. One *microgram* of any type of atom consists of more than  $2 \times 10^{15}$  individual atoms. Thus, macroscopic levels of demographic stochasticity are unknown in laboratory physical science. Advanced biology students and practicing biologists are already fully aware of this issue, but most introductory biology students have little experience in collecting biological data. It is easy for them to take accurate biological measurements too seriously, giving them a degree of respect that should be reserved for accurate physical measurements. The only real way to engender a proper mistrust of numerical data is to collect some, and it is particularly helpful to do so in a setting in which measurement error is eliminated. This is the point of the BUGBOX software. These applications are platforms for biological experiments in a virtual world that is so simple as to allow for exact measurement and so quick that full data sets can be collected in an hour, as compared to the time scales of days, weeks, months, and years needed for collecting real biological data. The data collected from a BUGBOX experiment can be used for modeling activities later in the chapter, but the primary reason for doing these experiments is to see that exact measurement does not yield a perfect data set. As the students are collecting this data, the instructor should frequently remind them that the conclusions that can be drawn from data are no more reliable than the individual data points.

Given the limitations of biological data, it makes sense to ask whether there is any value at all in mathematical modeling. Historically, most biologists have believed that modeling has no value beyond statistical analysis of data. This has changed to the point where most biologists will probably agree that mathematical modeling has some value and regret lacking the background needed for it. This background is supplied in the rest of this chapter and put to use in Chapters 5 through 7. In the present section, I indicate two goals for mathematics in biology. The first of these is the determination of patterns, which is the program of descriptive statistics. While not mathematical modeling, descriptive statistics does at least count as an application of mathematics to biology. The other goal is the determination of relationships, by which I mean the search for deterministic patterns that underly some stochastic data sets. This is the topic of both empirical modeling, in which we try to quantify deterministic patterns in data, and mechanistic modeling, in which we try to explain and predict such patterns. Two distinct issues raised by data sets from the BUGBOX-predator experiments are addressed later in the chapter:

- How can we quantify deterministic patterns that are partly obscured by scatter in data?
- How can we use biological principles to explain and predict systematic differences such as those between the predation data sets shown in Figure 2.1.1?

It is important to raise these questions in the context of this section and to identify the purpose of the empirical modeling section of the chapter as providing some answers.

## 2.2 Basic Concepts of Modeling

An algebraist colleague of mine once introduced me as a mathematician who worked in “modeling theory.” I appreciated that this error in terminology was an attempt to show respect—surely an intellectual endeavor is more worthy if it is supported by some theory!

In spite of the fact that modeling is almost entirely based on practice, there actually is a small amount of theory that is worth examining. These theoretical aspects of modeling are not much considered, and it has taken me a large part of a career to identify them and write them in the form that appears in this section. Nevertheless, I am confident that most mathematical modelers who read this section will say, “This is really right, although I’ve never thought about it like this myself.”

The key concepts of mathematical modeling are encapsulated in the two figures of this section, and classroom discussion with students should consist of explanation of these figures. Figure 2.2.1 shows the relationships between the real world and the mathematical model for the two types of modeling, along with the various processes that comprise mathematical modeling. Figure 2.2.2 illustrates the two types of thinking that mathematical modelers need to balance.

Mechanistic and empirical modeling share some common features, as can be seen in Figure 2.2.1. In both, we have a mathematical model that corresponds to a conceptual model rather than the real world. Whatever we learn from a study of the mathematical model is unerringly true for the corresponding conceptual model and may or may not be true for the real world that the conceptual model is trying to represent. It is this recognition of the correspondence between mathematical model and *conceptual* model, rather than real world, that determines how we think of models. This recognition is absent from the “applications” of mathematics that usually appear in mathematics books. This conceptual error has two unfortunate results: standard applications can give mathematics students a belief in false biological facts; more importantly, they go a long way toward creating the common view among biologists that mathematics is useless. What good is mathematics to a biologist when it claims to prove that ecological policies, such as hunting of coyotes, cannot cause local extinction of predators? The only way to salvage mathematical modeling from this debacle is to recognize that mathematical results from models are true only for the corresponding conceptual model. Not only does the competent mathematical modeler refrain from claiming mathematical results to be biologically true; she also determines what is wrong with the conceptual model, replaces it with a better one, provides more meaningful modeling results, and suggests further experiments that could be used to confirm new predictions.

Mechanistic models have explanatory and predictive value that empirical models lack. In empirical modeling, we merely use mathematical methods to analyze data. In mechanistic modeling, we use scientific principles to create a conceptual model, derive a mathematical model from the conceptual model, and tie the results of analysis to the assumptions in the conceptual model. When the results are valid, we gain additional explanatory value as compared to empirical models, and we gain predictive value because we can apply the models to scenarios for which no data is yet available. When the results are not valid, we can sometimes identify and correct flaws in a conceptual model, leading to a better mathematical model and an improved biological understanding.

Figure 2.2.2 illustrates what I call the narrow and broad views of mathematical models.



This distinction ties in with the material of Section 1.1 on parameters in models. Application problems typically work only on the narrow view. The essence of modeling, however, lies in the broad view, where we examine the effects of parameters on features of solutions. A discussion of the broad view would be enhanced by references to topics in Section 1.1.

Problem 2.2.1 is essential, as it is the only *simple* example I know of where an extremely elementary mathematical analysis can quickly identify the flaw in a conceptual model.<sup>2</sup> The crucial point here is that the erroneous prediction made by the model is a direct result of a flaw that should have been immediately obvious. When a careful study of the Lotka-Volterra model results in a startling prediction, this should not be a surprise if we understand that the most elementary preliminary analysis shows that no other prediction was possible. While this may seem like an obscure case, the actual fact is that a surefire way to find undergraduate research projects in mathematical biology is to identify a recently published paper that makes a startling biological claim based on a mathematical model. Such papers almost always have a concealed conceptual flaw that would easily have been determined had the authors done an elementary analysis such as the one in this problem.

Problems 2.2.3 through 2.2.8 are intended to get students to examine models with the broad view of modeling. It is difficult to find good problems for students with minimal modeling background, and I hope at least some of these will serve this purpose. Problem 2.2.7 requires quite a bit of calculus and algebra and should only be used with students who have brought a significant knowledge of calculus to the course.

## 2.3 Empirical Modeling I: Fitting Linear Models to Data

All of the treatments of linear regression that I have seen begin with the full linear model  $y = b + mx$ . This seems to be the logical approach on mathematical grounds—why *not* do the most general linear model? From a modeling perspective, however, starting with the full linear model is not the logical approach. In Section 2.7 on model selection, we will invoke a modern interpretation of the scientific principle known as Occam’s Razor, which is generally used to argue that simple explanations should be preferred. The original statement of this principle translates directly into English as “Entities must not be multiplied beyond necessity,” a version that translates very nicely to mathematical modeling as “Parameters should not be added to models unless the increased accuracy justifies the increase in complexity (measured as the number of parameters).” This principle suggests that we should assume a linear model of the form  $y = mx$  unless the extra complexity of the full linear model is justified by increased accuracy. The final arbiter of the inclusion or exclusion of the parameter  $b$  is the Akaike Information Criterion (AIC) for model selection, which quantifies Occam’s Razor. For many linear data sets, it is obvious that the  $y$  intercept is non-zero, but there are also a significant number of cases where the data suggests the plausibility of a zero  $y$  intercept. In most of these cases, there are mechanistic reasons why the  $y$  intercept should be zero, and generally the data supports this choice through application of AIC. This section is not the right place to address the idea that optimal accuracy does not mean optimal choice

---

<sup>2</sup>Given the ubiquity of this model and the simplicity of the analysis, it is very telling that the analysis given in this problem does not appear in any mathematics books to my knowledge. This is strong evidence for my contention that instruction in mathematical modeling is almost completely absent from standard academic practice.

of model, so the issue is best left unmentioned here. We do not need to explain to students our reason for starting with a one-parameter model because they have no expectation that we should do otherwise.

Starting with the one-parameter model  $y = mx$  has many pedagogical advantages. Conceptually, having just one parameter allows us to demonstrate as in Figure 2.3.1 that there is clearly an optimal value of the slope, which we can estimate by visual inspection. Mathematically, having just one parameter allows us to set up a one-variable calculus problem to determine the optimum value for that parameter. Decoupling the problem of finding the best slope from that of finding the best intercept makes the second problem one-dimensional as well. My approach to data sets requiring a full linear model is to shift the coordinate system so that the mean data point is the origin and then fit the model  $y = mx$ . Clearly the shift of coordinate system does not affect the optimal slope, so we can approach the problem of finding the optimal intercept after the optimal slope is already known. This overall approach is an example of my efforts to use the most elementary mathematics possible when developing new results. Whereas the standard approach to parameterizing the full linear model requires multivariable calculus to derive the correct formulas, my approach requires only one-variable calculus.

Because the emphasis on this book is modeling, we cannot simply assume that the optimal  $m$  should be the one that minimizes the sum of squares of vertical discrepancies. Instead, we must think about how to assign a quantitative value to the discrepancy between a function  $y = mx$  and a set of data points. It is important to explicitly connect the standard choice with the assumption that the  $x$  values are known exactly and all the uncertainty is in the  $y$  values. A full treatment of this issue is far beyond the scope of this book, but a recognition of it is necessary for good modeling practice.

I see little value in simply doing linear least squares with a computer and writing down the results. For this reasons, all the problems ask the reader to do more than that. Problem 2.3.3 examines the question of how much of a difference an error in the data makes on the result, showing that the results are more sensitive to values at the edges of the data than values in the interior. Problems 2.3.4 and 2.3.5, which should be done with no computational aids other than a calculator to compute squares, products, and sums, show the importance of having some measure of accuracy for the overall fit and leaves open the question of whether a better function might give a better fit for some data sets. Problem 2.3.8 uses a parameterization problem as a setting in which to consider the sort of modeling issues that are addressed in Section 2.2. This problem is worth the extra time required to have students work it in detail, particularly if modeling is a major emphasis of the course.

## 2.4 Empirical Modeling II: Fitting Semilinear Models to Data

Most books that discuss the problem of fitting a nonlinear function to a data set operate under the tacit assumption that it is always good to transform the model into a linear form so as to be able to use linear least squares. Spreadsheets and software such as Maple have built-in routines for doing this with common functions such as exponential and power functions. From a mathematical point of view, it is interesting to see how to do this. From a modeling point of view, this practice raises the important question of whether the results are useful. The central issue is that any least squares procedure will find the parameters that minimize

the total discrepancy on the graph of the variables used for the computation. When fitting a power function, for example, there is a difference between minimizing the total discrepancy on a graph of  $y = Ax^p$  and minimizing the total discrepancy on a graph of  $\ln y = \ln A + p \ln x$ . Given scattered data, the parameter values that solve these two problems are not the same. Which is better is a modeling question rather than a mathematical question. Since the chapter is about modeling, I take a nuanced approach to the problem of fitting nonlinear models to data.

The first part of Section 2.4 develops the standard methods of converting exponential and power function models into a linear form so as to use linear least squares. The standard practice for fitting data directly to a nonlinear model is to use a fully numerical method. In keeping with my overall philosophy of preferring methods that are simple enough for readers of the book to fully understand, I instead present what I call the “semilinear” method. This method does not work for all models, but it does work for any two-parameter models in which one of the parameters is merely a scaling parameter. This applies to exponential functions, power functions, and Briggs-Haldane/Holling functions, which are among the most common nonlinear models in biology. The idea is simple enough to present to students at the level of this book, and it provides an illustration of an important problem-solving technique, that of iteration. The materials provided along with this book include simple programs in R and Matlab that use built-in one-variable optimization routines to implement the semilinear method. Since the optimization problem has just one parameter, it is also possible for students to solve it by graphing the objective function, thereby confirming that the computer program is giving the right answer. As a mathematician, I consider this to be far superior to the alternative of using the fully nonlinear method on a computer and having no way to confirm that the results are correct.

For a course that emphasizes modeling, it is worth spending a class period in a computer laboratory writing the program for semilinear regression from scratch. I generally work with the students to write out a generic algorithm that is independent of the specific syntax for any programming language. Then I present the appropriate programming statements for each step in the algorithm. Students acquire some facility with programming by the simple expedient of starting with a program that they have written with this kind of guidance and then making the minor modifications needed to adapt the program to work on similar problems. For example, it takes much less programming skill to modify a program that already works for a different nonlinear model than it does to write a program from scratch. Program statements that deal with the inner workings of the algorithm are unchanged, so the students do not need to master the syntax for these statements, although they should have a general idea of what these statements do. Similarly, I do not expect students to learn the details of producing good plots in R; I provide them with examples of plotting commands and a little information about how to modify line thickness, color, axis limits, and other relevant minor details.

The Lineweaver-Burk method, which fits parameters for the Briggs-Haldane equation for Michaelis-Menten reaction rates using a particular linearized form of the model, consistently produces terrible results with real data. Nevertheless, a simple Google search attests to its continuing popularity among biochemists; as of this writing, only one of the top 30 hits is about the woeful inaccuracy of the method while the rest are about how to apply it. This is a great example of the institutionalization of bad quantitative techniques. Other

linearizations have been proposed; some of these are better, but still not very good. Before modern computers, there was justification for wanting to find a model linearization that could approximate the best fit. In the modern world, these model linearizations make no more sense than the “qwerty” computer keyboard, which was originally designed to make people type slowly and yet remains in use because of institutional inertia.

While my diatribes about the unjustifiable use of model linearizations might carry some weight as an expert opinion, Problem 2.4.8 allows students to see for themselves just how bad Lineweaver-Burk is. The data provided in this problem is measured data reported in the literature. Although not recent, the data is very good in the sense that the model fits it very well with optimal parameter values. The Lineweaver-Burk method fails to come close to the optimal parameter values, however, and a graph that shows its results compared to the semilinear method is dramatic. There is nothing special about this data set; the same general features are found in any real data set that comes close to fitting the Briggs-Haldane/Holling model. The only way to get good results with Lineweaver-Burk is to use simulated data based on assumed values of the parameters with very little added noise. Independent of the specific lesson this problem teaches about a particular method, it also serves to point out that not everything found in standard practice is correct. Mathematical modelers need to do their own thinking and not just copy what others have done before.

## 2.5 Mechanistic Modeling I: Creating Models From Biological Principles

This is a long difficult section, but it need not be mastered by students. The point of the section is to show students how mathematical models arise from experiment and observation and how faulty models can be improved by a critique of the conceptual model. It is unrealistic to expect students at the level of this book to be able to construct mathematical models. Instead, the goal is for them to learn to interpret mathematical models. In terms of Figure 2.2.1, the goal is to understand the derivation process that links the conceptual model and the mathematical model. In particular, being able to reconstruct the conceptual model from a mathematical model is an essential skill for anyone who wants to be able to read biology papers that feature mathematical models. There are too many instances of poorly designed or poorly chosen models in biology papers to allow a reader to accept a model uncritically.<sup>3</sup> The primary focus of the problems for this section is on describing the conceptual model implied by a mathematical model; a secondary focus is on making small modifications to existing mathematical models to account for changes in the conceptual model.

---

<sup>3</sup>In saying that there are many instances of poorly designed or poorly chosen models, I do not mean that the conceptual models are unrealistic. Almost all conceptual models, radioactive decay being a rare exception, are necessarily unrealistic. The Holling type 2 model only accounts for the most elementary features that can be observed in consumer-resource or predator-prey interactions. What I mean is that the conceptual model is so bad that the results cannot even be considered approximately valid. The Lotka-Volterra model of Problem 2.2.1 is the most obvious example. I have also seen a structured model for a cannibalistic population in which the growth of juveniles through eating some other species was absent. Most errors are more subtle than this, but many are easily caught by careful reading of the differential equations.

## 2.6 Mechanistic Modeling II: Equivalent Forms

Most students find nondimensionalization to be difficult. Because of this difficulty, instructors will no doubt be tempted to skip this section. This would be a serious error.

Nondimensionalization is essential in mechanistic modeling for two reasons. One is that fewer parameters make the analysis easier. The model analyses that appear in Chapters 5 and 7 would not be possible without first scaling the models, by which I mean nondimensionalizing them using reference quantities that represent meaningful or convenient values. The mechanics of nondimensionalization is developed in this section, while the choice of reference quantities is done only in the context of specific models. The other reason is that careful attention to nondimensionalization can identify flaws in a conceptual model. I have pointed out in several places that modeling errors are disturbingly common in mathematical biology, but they are far less common in models that have been properly scaled. Readers of mathematical biology will encounter nondimensionalization in good modeling papers, and it is important that they be able to make sense of what is really a rather simple idea.

I have put a lot of thought into the question of how to teach nondimensionalization to students who are at or just beyond the level of one-variable calculus. The key is to understand why students find the topic to be difficult. There are two reasons: weak algebra skills and a lack of understanding of how symbols are used in mathematics. Given that the algebra of nondimensionalization is trivial, I believe that lack of understanding of symbol use is the key problem. This is the reason for the opening section of the book that focuses on parameters. It is also why this section embeds nondimensionalization in the broader context of algebraic equivalence of models.

To those of us with a good understanding of how symbols work, it is hard to see why students cannot immediately identify models that differ only in the choice of symbols. We know that the meaning of a symbol is contextual and we have the ability to absorb the context of a symbol when it is presented in a model. Naive students take symbols literally and are uncomfortable with, for example, defining a dependent variable  $x$  as a function of an independent variable  $y$ . This section takes students very gently from examples where only the symbols are different to examples where some algebra is needed to establish equivalence, and finally to examples of nondimensionalization. Nondimensionalization is developed in two stages, beginning with the detailed development via graphs leading up to Figure 2.6.2b and proceeding to the algebraic technique for nondimensionalization.

Readers not planning to study the dynamical systems part of this book can manage without nondimensionalization; however, they should not skip this section altogether. The material through Subsection 2.6.3 is important in empirical modeling and probability as well.

## 2.7 Empirical Modeling III: Choosing Among Models

This is not a section for mathematicians who believe that we should only present mathematical methods that we can prove. The Akaike Information Criterion is a valuable tool that has caught on in ecology. It is based on the simple idea that model selection should be based on simplicity as well as accuracy of fit. It is easy to quantify each of these aspects individually: residual sum of squares quantifies accuracy and number of model parameters quantifies complexity. How these should be weighted is by no means obvious. At the level

of this book, we can explain the idea of AIC but we can't justify the specific formula. This is an example of what one of the reviewers of the book referred as a "gray box" method, a phrase that makes an appealing comparison with black box methods.

Some AIC practitioners advocate the practice of model weighting, in which models are averaged using AIC values to determine weight factors. I do not address this issue in the book, but instructors may want to address it briefly in class. Shane Richards, who I consider to be the leading expert on AIC, dislikes model averaging because it does not tend to produce better results than the best model identified by AIC. My argument against model averaging is theoretical. If I average a model with two parameters and a model with three parameters, then I am really using a model with six parameters, counting the weight factors as one additional parameter. Accordingly, it seems wrong to me to use AIC values for the separate models. If we compute an AIC for the averaged model, the resulting value will be much higher than the AIC values for either of the individual models. In other words, a model created by averaging two models together ought to have a much larger complexity penalty than either of the models used in the average. I also dislike model averaging because it explicitly devalues the contribution of mechanistic justification to model selection. Even empirical models often have some mechanistic justification, which averaged models necessarily lack.

### 3 Probability Distributions

Instructors who have taught a course in elementary statistics will find that my treatment is very different from that in statistics books. This is not due to sudden inspiration on my part but grew out of my experience of trying to teach the material to biology students. Probability and statistics are difficult subjects. This is partly because some of the ideas run counter to normal intuition, but it is also because of the way these topics are usually taught. Some books use an axiomatic approach to probability, which is far too formal and sophisticated for beginning students while also being unnecessary for applied students. Other books present an ad hoc approach to statistics that makes the subject appear as a collection of enshrined rules for obtaining results that the student is incapable of converting into meaningful verbal statements. In my experience, relatively few papers that use statistics for analysis also include a thoughtful discussion of what the analysis means.

My coverage, approach, and organization is based on discussions with biologists. In particular, my colleague Drew Tyre told me that he finds it to be a relatively easy matter to teach statistics to students who understand probability distributions. At the same time, I had learned from teaching experience that students have a difficult time with the standard combinatoric calculations of probability with uniform distributions, such as probabilities of poker hands. Since combinatorial probability is unimportant to biology, I have omitted it entirely. I do include conditional probability, but I have placed it at the end of the two chapters on probability so as not to distract from the principal goal of teaching probability distributions.

Chapter 3 begins with a section on descriptive statistics. The rest of the chapter is based on the overall theme of probability distributions as models for populations of data. In keeping with the general theme of modeling, a model population should be infinite, so that its characteristics do not change when individuals are removed, and simple, so that

probabilities can be calculated from basic formulas.

I characterize statistics as the use of arbitrary rules of thumb to draw conclusions from probability results. This should be interpreted less as a criticism of statistics than as an argument for why a mathematics book should omit many standard statistics topics while including significant coverage of probability. I am happy to appropriate those topics in statistics that are arguably mathematical while leaving those that are not, such as hypothesis testing, to the statisticians. There is also the matter of what questions modelers should address. Questions about whether a hypothesis should be rejected do not fit with a mathematical modeling theme. Questions about which model should be chosen for a relationship, which distribution should be used for a data set, and how likely it is that a hypothesis is true do fit within an overall modeling theme.

### **3.1 Characterizing Data**

This section provides a basic introduction to descriptive statistics, focusing on the use of histograms to display data and the mean and standard deviation to characterize it.

### **3.2 Concepts of Probability**

This section provides a basic introduction to probability, focusing on the key terminology, concepts, and properties of probability and the basic rules for sequences and complementary events. Two minor features distinguish my presentation from that found in a typical probability and statistics book. One is that I do not use an axiomatic approach to probability. Only a mathematician would define probability as a function with certain properties and make no reference to what probability is supposed to do. To everyone else, probability is a measure of how likely something is believed to be. The properties used by mathematicians to define probability are necessary because of what probability is supposed to do. Defining a vector space as a set of objects with certain properties makes sense because that definition generalizes the familiar concept of vectors to other settings; the axiomatic definition of probability does not do this. The second feature is that I use a biological setting in which to introduce probability. One of the important problems of bioinformatics is to distinguish between DNA sequences that are used by an organism and those that are merely artifacts of evolutionary history. Identifying meaningful sequences can in principle be based on probability because certain words in the genetic code cannot occur in the middle of a meaningful sequence; hence, long stretches that lack these words are almost certainly meaningful. In practice, this is not as good as it sounds because genes do not always consist of a single sequence. Just as programs on a hard drive can consist of portions stored in more than one location, genes can have fragments that are isolated from other fragments. Nevertheless, the limited occasions where gene identification can be based on an absence of stop codons make this example ideal for showing students how important conclusions can sometimes be drawn from simple probability results.

The problem set is short and simple, in keeping with a section whose purpose is to provide background for later material.

### 3.3 Discrete Probability Distributions

With the preliminaries out of the way, this section begins the study of probability distributions by introducing the basic concept and some simple examples with finite sample spaces, including the uniform distribution and distributions that are easy to define without a complicated formula. It makes sense to look at these simple distributions before working on the more complicated ones that play a central role in biology. As with the first two sections, this one should be relatively quick reading and the problem set is limited. Problems 3.3.4, 3.3.5, and 3.3.7 are best done with the aid of trees.

### 3.4 The Binomial Distribution

The binomial distribution is introduced using Mendel's famous genetics experiments to provide context. We start with an example having just three Bernoulli trials, worked out using sequence arguments. Details in the exposition, such as a histogram for Mendel's actual sample size of 24, are intended to build intuition. In particular, some of the discussion anticipates the study of sample distributions that occupies the first half of Chapter 4. It is well worth noting the high probability that one out of twenty biologists repeating Mendel's experiments will get results that do not suggest Mendelian genetics; this is not through any experimental error, but merely because of demographic stochasticity. There is a tendency to think that a random sample must be representative, and we need to take every opportunity to find ways to illustrate that this is not correct. This theme will be central in Section 4.1.

Problem 3.4.8 has students define a negative binomial distribution on their own, program a formula for it, and use the program to determine the mean and standard deviation. I do not identify the distribution as a negative binomial because I want the focus to be on the ideas of probability distributions rather than the mathematical formulas for them. Simply reading about probability distributions and doing individual calculations is not sufficient for good conceptual understanding. Rather than looking up the formulas for mean and standard deviation of the negative binomial distribution, it is much more meaningful to compute these quantities by writing a computer program. This example returns in Chapter 4 as a source population for sampling experiments.

### 3.5 Continuous Probability Distributions

Continuous probability distributions are usually defined in terms of the probability density function. This is yet another reason why probability is such a difficult topic for students. The problem is that it is very difficult to say just what the probability density function means. It is used in the same way that a mass density function is used to compute mass, but there is a big conceptual difference. The differential in an integral for mass of a wire is clearly length, so mass density is mass per unit length. Similarly the differential in a probability integral is interval width, so probability density is probability per interval width. The difference is that interval width is much harder to get hold of than physical length.

Continuous probability calculations can be done as either an integral of a probability density function or as a difference of a cumulative distribution function. The former has the advantage of generally being easier to write down, as in the normal distribution for which



the cumulative distribution function is usually given as an integral. This distinction matters less in the age of modern computers than it did in the past. We could write the cumulative distribution for the normal distribution as an error function, but it is easier simply to assign the name  $N(z)$  to the integral definition and then write a computer code to evaluate it, at which point it becomes like other transcendental functions that we understand by their properties and calculate with a computer. While we may prefer the definition by probability density function as mathematicians, as modelers we use cumulative distribution functions to calculate probabilities. Given my preference for working with cumulative distribution functions, the exponential distribution is the obvious place to start. In keeping with a section focused on general themes, I do not name this distribution here or describe its use; these are deferred to Section 3.7.

Preferring to define continuous distributions with a cumulative distribution function does not mean choosing to ignore the probability density function altogether. One of the key points of this section, and the primary point of the homework, is to explore the connections between the probability density function of continuous probability and the distribution function of discrete probability. The key point is that collecting values from a distribution into bins and creating a histogram is essentially the same for both discrete and continuous distributions, with the histograms for the latter serving to approximate the probability density function. The key difference is that the graininess of the data limits the width of bins for discrete distributions while bin widths for continuous distributions can be made arbitrarily small. Thus, we can get from histogram to probability density function by taking a limit of a histogram as bin width goes to zero. The difficulty in this limiting process is that the frequencies of the intervals also goes to zero as the interval width goes to zero. The solution to this difficulty is, of course, to label the heights of the bars with frequency per interval width rather than frequency, thus providing a limiting process in which the vertical axis range converges to some interval and the dimension of the vertical axis matches that of the probability density function. Figure 3.5.3 illustrates this idea and is the key to the section. Similarly, Problem 3.5.7 is essential for students to do themselves, even if doing so requires use of class time.

## 3.6 The Normal Distribution

I define the normal distribution using parameters  $\mu$  and  $\sigma$  without asserting that these parameters are actually the mean and standard deviation of the distribution. These facts emerge after the introduction of the standard normal distribution and calculation of its mean and standard deviation. I do this because I think it is important to make a distinction between distribution parameters and distribution properties. These may be the same, as in the normal distribution, but not necessarily so. Symbols can take on any meaning, so the mere fact that  $\mu$  is used as a distribution parameter begs the question of whether that parameter is actually the mean of the distribution.

In keeping with the emphasis on modeling, it makes sense to raise the question of how we can use a normal distribution, which by definition admits all possible outcomes, for a setting in which negative values are impossible. This is addressed by considering the probability of negative values arising in a normal distribution, given the numerical values of mean and standard deviation.

At the end of the section, I define standard probability intervals, which correspond roughly to the more common notion of confidence intervals. One has to use the term “confidence interval” carefully, as it properly refers to the question of whether the true mean lies within a given interval rather than the simpler questions of what fraction of measurements lie within an interval for a given mean and standard deviation. The use of the term “standard probability interval” avoids this common confusion.

The problem set includes only a small number of problems for finding probability of a normally-distributed random variable falling within a given interval. These problems were once done with tables and required practice, but now they are done trivially with built-in functions in computer software and calculators. Most of the problems use histograms and probability density functions to build intuition about normal distributions. Problems 3.6.12 through 3.6.15 address the modeling question of how to tell the difference between data that represents one normally-distributed variable and data that represents a combination of two normally-distributed variables with different means. This is a practical matter, as it is often necessary to decide whether two subsets of data can be combined into one set. Indeed, the BUTTER.\* data file used for several of the problems originally distinguishes two ages of cattle, but computation of means makes it very clear that this distinction is unnecessary and I combined them together. Whether distinctions must be made by breed is a more complicated questions addressed in Problems 3.6.12 through 3.6.14. Problem 3.6.15 considers the same question from a theoretical viewpoint. This problem is very helpful for building intuition, as it shows a gradual transition from the case where the subpopulations are similar enough that they can clearly be considered together to the case where they are so different that the probability density functions show little overlap.

One additional issue is raised in the text and problems but is not addressed here: how does one decide whether a normal distribution is a good model for a data set? This important question is largely ignored by elementary statistics books, which focus on theoretical issues rather than modeling issues. It is addressed in Section 4.5 along with the related question of how to decide whether to discard outliers in a data set.

### 3.7 The Poisson and Exponential Distributions

I put the Poisson and exponential distributions together in one section partly because of their common connection to events occurring over time and partly because neither provided enough material for a whole section. The focus is on the modeling issue of deciding when these distributions are appropriate. As with many other topics, the calculations are routine and largely done by computers, but the results are meaningless without a proper interpretation.

Several of the problems involve characterization of a data set and qualitative examination of the fit to a distribution. Some of these sets fit a distribution beautifully, while others do not fit at all well. Problem 3.7.14 develops the idea of using the Poisson distribution to approximate the binomial distribution. There are a number of rules-of-thumb that one can find that indicate when the approximation is good. Of course this is somewhat subjective, and it is far better to offer a formula for error approximation, which allows the user to decide how much error is tolerable. The formula I present in this problem is based on an asymptotic approximation for relative error along with an observation that appears to connect maximum absolute error with maximum relative error. The error approximation is

remarkably accurate for the values used in the problem, and it seems to be accurate over a broad range of parameter values. Problems 3.7.15 and 3.7.16 introduce the negative binomial distribution, which I felt should be included for the benefit of wildlife biologists although not in the text itself.

## 4 Working With Probability

This chapter is the least coherent one in the book. It could not have been otherwise. Many reworkings of Chapter 2 eventually resulted in its current form, which seems very coherent to me now. A year from now I may think the order of sections could be improved, however. There are tradeoffs in imposing a linear structure on a multidimensional graph.

In probability, there was a need for sections on descriptive statistics, basic probability, conditional probability, probability distributions, samples and inferential statistics, and parameterization of probability models. In early versions, I had the topics in this order, with the chapter break in the same place as it is now. That made for a nine-section Chapter 3 followed by what was then a very short Chapter 4 of just three sections. In the classroom, I found the long interlude of three sections on basic and conditional probability to be an intrusive break between descriptive statistics and probability distributions, which I felt belonged together. In the end, I opted to remove conditional probability from Chapter 3, making the chapters more equal in length and making Chapter 3 highly coherent. Experience with students has confirmed for me that this is the correct pedagogical decision. No harm comes from delaying conditional probability, and I believe that instructors who are able to get past the initial reactions of “You can’t do that” and “That’s not how I learned it” will come to agree with me. I would even go so far as to say that I would rather scrap conditional probability for lack of time than to scrap material on probability distributions or sampling. Instructors anxious to have enough time for conditional probability should skip Section 4.5 and could also opt not to do the Poisson and exponential distributions of Section 3.7. Of course they can also insert Sections 4.6 and 4.7 into their original location between Sections 3.2 and 3.3.

### 4.1 An Introduction to Statistical Inference

This section was added in the final revision of the book, in response to a reviewer’s request for much more material to help students understand sampling. This was certainly an oversight on my part. All of the other chapters began with at least one introductory section that explores a topic without getting into the technical details. There was clearly a need for the same thing in this chapter.

The question to be addressed in choosing a topic for the opening section was “What is the important intuition that students should get about applications of probability?” Just as the answer to the question for the modeling chapter was an understanding of the importance of demographic stochasticity, the answer for this chapter was an understanding that random does not mean representative. This point has already been made once, in the discussion of Mendel’s pea experiments that appears in the discussion of Example 3.4.4, where I observed that approximately 1 in 20 biologists repeating Mendel’s experiments would obtain

evidence that was not sufficient to suggest the dominant-recessive combination that figures prominently in Mendelian genetics.

Where I usually prefer a biological setting in which to introduce a key idea, I chose a setting from everyday life for this section. The point to be made runs counter to our intuition, and I felt that it could be made more strongly in a context in which our intuition is strong. We all have the experience of seeing something that seems unusually rare and wondering whether there is any significance to its appearance.

Three main points are made in this section and supported by the problem set. First, the test for a discrete outcome to be considered unusual must be based on the probability of obtaining that outcome or smaller (if the result is below the mean) rather than the probability of obtaining that specific outcome. This is a hard point to understand and a hard one to explain. I have tried to do so with specific examples from the calculations. It should be fairly easy to explain why the specific probability is inadequate. Where the number of discrete outcomes is quite large, each of them has a low probability even though some are close to the mean.

The second main point is brought out by the specific numerical values used in the story. The initial set of data probably seems convincing to most people, but turns out to be merely suggestive. The second set of data falls in that gray area where as ordinary humans we are utterly convinced but as scientists we remain skeptical. At the risk of being overly didactic, I put the key conclusion in italics: *The standard for a result to be considered unusual is much stricter than the standard needed to arouse suspicion that a result is unusual.* In work on education research projects, I have learned that even professional researchers are often overly interested in whether a statistical test returns a “significant” result, missing the additional questions of whether the difference found to be significant is meaningful and whether the unusual data can be supported by a plausible theoretical explanation. I would rather accept something that is significant at the 94% level and backed by theory than something that is significant at the 96% level and seems to be inexplicable.

The third main point is the one that drove the choice of example, that random samples do not have to be representative. A probability of 0.05 does not merely mean that something is unlikely—it also means that we will see it in one case out of twenty. When we are using data to inform scientific judgment, we have to keep in this lesson firmly in mind; our results could be anomalous simply because we were unlucky enough to have chosen a random sample that is not representative. Put another way, a research goal is to obtain a representative sample, but we have no way to accomplish this goal directly. The best we can do is to obtain a large random sample. The larger the sample, the more likely it is to be representative, but the only way to be sure is to repeat that large sample numerous times. Of course this cannot be done in science, but it can be done using computer simulations. The problem set provides opportunities for students to do just that. The size of sample needed to reproduce probability results to four decimal digits is staggering, as the student sees in Problem 4.1.1, where we draw one million numbers from a binomial distribution and compare the results to the theoretical probability. In Problem 4.1.3, we test the results of a sample of some given size by collecting a lot of comparable samples and examining the aggregate results. Of course they never approach the accuracy achieved in the example with one million samples. In Problem 4.1.4, we determine how large a sample is needed to reach a given level of confidence that an observation of 2.5% successes is unusual when 5% is expected. The sample size required

for what appears to be an obvious difference is much larger than most people would guess it to be.

I believe that learning to program a simulation on a computer is a worthwhile skill on its own. I use software packages to sample a probability distribution or to compute theoretical probabilities, but not to run simulations. Problem 4.1.5 asks the student to write a very simple simulation that is a prototype for agent-based modeling. I do not discuss this topic in the book, but it could be mentioned by the instructor in conjunction with this problem.

Problem 4.1.6 will be inaccessible to most students, but I have included it as one of a small number of problems to provide mathematical challenges for those students who would enjoy them or benefit from them.

## 4.2 Tests on Probability Distributions

Connecting models with data is a recurring theme of mathematical modeling. Chapter 2 features three sections that deal with parameterization and selection for deterministic models. This chapter features two sections that deal with similar topics for probabilistic models: this section and Section 4.5. Thematically, these two sections complement each other; I have placed this section second because it serves the dual purpose of providing a tool for understanding the central limit theorem, which appears in Section 4.3. Students have a very difficult time understanding distributions of sample means as distinct from the underlying distributions. With a quantitative test for normality, we can see how larger samples produce means that become normally distributed. In Section 4.4 on inferences about populations, we can use a normality test to justify the normality assumption required to relate standard deviation to probability.

The question of whether a data set appears to come from a normal distribution can not be resolved in the affirmative, but only in the negative. On the other hand, data that is not inconsistent with a normal distribution can safely be considered as normally distributed; hence, there is no great need for an affirmative result.

As in my treatment of AIC, I do not shy away from using results presented without proof, provided the conceptual basis is clear. The core of AIC is the idea of combining the residual sum of squares on a graph of  $y$  vs  $x$  with a measure of model complexity. The specific manipulation used to combine these quantities and the interpretation of AIC differences is presented without proof. Similarly, the core of the Cramer-von Mises test statistic is a residual sum of squares measured on a graph. The specific correction for small sample size and the corresponding confidence levels are presented without proof.

This section also includes a brief optional discussion of outliers. The most commonly used test for outliers, Grubbs' test, is somewhat complicated, so I present a more conceptual method in the text and defer the more sophisticated test to the problem set.

Most of the problems on normality testing are routine examples that appear here to justify the assumption of normality that will be made when the same data sets are investigated in Section 4.4. Issues of what makes a distribution not be normal are explored in several problems in different ways. Problem 4.2.5 is a data set that fails the normality test only because the data is too highly rounded. Problem 4.2.8 is a data set that is exponentially distributed rather than normally distributed, and problem 4.2.9 looks at uniformly distributed data and data drawn randomly from a uniform distribution. Problems 4.2.4 and 4.2.10 ex-

plore what happens when two normal distributions are combined into a single data set, the first using experimental data and the second using theoretical data. Problem 4.2.11 looks at normally distributed data that is contaminated with data from some other distribution.

### 4.3 Probability Distributions of Samples

The only way to obtain useful data in the face of demographic stochasticity is to use large samples rather than individual values. It is therefore essential for biologists to understand the basic facts of sampling distributions. The subject is difficult because of the need to clearly distinguish the underlying distribution from the distribution of sample means. In my experience, the best way to get students to understand the relationship between individual values and sample means is to have them write computer simulations to compute sample means. This is the approach taken in this section. We have the simple theory to see how the standard deviation changes with sample size, but the change from an underlying distribution shape to a normal distribution is mysterious. Computer simulations show the gradual progression of the distribution as it becomes less skewed. Analytical evidence using the Cramer-von Mises test complements the visual evidence from histograms. I spend a day in class guiding the students through the process of writing the computer simulation, and then I assign problems using the simulation to be done at home. Most, but not all, of the problems in this section are of this type. Problem 4.3.1 asks students to prove the theorem about means and standard deviations for sums of distributions. Problem 4.3.2 requires only means and standard deviations. Problem 4.3.8 analyzes a wildly inaccurate statement about probability that I heard on a news report from an otherwise trustworthy source, NPR. I posted the comment, “Your story on climate change contained a wildly inaccurate statement equating one standard deviation to rolling snake eyes three consecutive times. In actual fact, a sum of 6 on six dice represents 3.6 standard deviations from the mean. It is unfortunate for NPR to inadvertently make a negative contribution to mathematics education. Not to worry—I am writing a mathematics book and will a positive contribution out of your error by using it for an exercise.” Parts d and e of that problem explore the relationship between the granularity of a mean of discrete sums and the closeness to normality.

### 4.4 Inferences About Populations

This section presents the key probability results used in inferential statistics, but without emphasizing the 5% and 1% significance levels. Given the focus on modeling, I pay particular attention to the question of whether an approximation valid in the limit  $n \rightarrow \infty$  is useful for finite  $n$ . This is particularly relevant in the probability result for inferences about proportions. I chose not to present a lot of detail in the discussion here, but the individual instructor might want to add to the text presentation. The issue is that discrete distributions are granular. The probabilities  $P[X \leq 5]$  and  $P[X < 6]$  are the same in a discrete distribution because there is no possibility of an outcome less than 6 and greater than 5. In a normal distribution, these probabilities differ by  $N(z_R) - N(z_L)$ , where  $z_L$  and  $z_R$  are the  $z$  values obtained from  $x = 5$  and  $x = 6$ , respectively. The magnitude of this difference is a measure of the granularity of the discrete distribution.

The problem with granularity is that we could approximate  $B(5)$  by either  $N(z_L)$  or  $N(z_R)$ , depending on whether we interpret  $B(5)$  as  $P[X \leq 5]$  or  $P[X < 6]$ . Theorem 4.4.2 is based on the assumption that the best strategy is to use  $N(z_M)$  instead, where  $z_M$  is the  $z$  value corresponding to the intermediate point  $x = 5.5$ .

Most of the problems combine basic calculations from the two theorems with more sophisticated questions designed to give students experience in interpreting probability results and a general sense of how more data allows for stronger conclusions.

## 4.5 Estimating Parameters for Distributions

The connection between experimental and theoretical science was introduced in Section 2.2 and further explored in the three sections on empirical modeling. We return to that theme in this section with a brief look at the problem of using experimental data to obtain parameter values for probability distributions. After the standard material on confidence intervals for normal distributions, I present a standard example of finding the success probability for a binomial distribution and then turn to a more complicated situation of an ecologist who wants to estimate the size of a population that cannot be counted directly. The mark-and-recapture method is of great interest to ecologists, but probably not to anyone else, and it could easily be omitted.

Problems 4.5.7, 4.5.8, and 4.5.12 have the students analyzing data they collect for themselves. I recommend doing at least one of these for classes with an emphasis on modeling. People do not have good natural intuition for probability and statistics, and the best way to develop intuition is to collect and analyze data. Problem 4.5.6 is of particular importance for understanding likelihood functions. The problem deals with a very large sample of Bernoulli trials; the problem guides students to the discovery that the likelihood for a normal distribution is the probability density function, suitably scaled. This is what allows us to make a connection between confidence intervals and values of the likelihood function.

## 4.6 Conditional Probability

This section presents a standard approach to the subject of conditional probability, albeit rather late in the book for many people's taste. In my experience, students find conditional probability to be conceptually more difficult than probability distributions. The idea that new information changes the probability of something that has already happened is unnatural, as seen in discussions of the Monty Hall problem that occur every few years. This is the problem made famous by the old television game show, "Let's Make a Deal," which I am old enough to remember watching. In one segment, a contestant guesses which of three doors shields the one real prize. The host, Monty Hall, then opened one of the other doors to reveal a joke prize. The contestant was then given an opportunity to change doors. It is not obvious that doing so confers an advantage, and not all contestants did. I recommend doing this problem in class with students because it offers a nuanced example of conditional probability.

A conceptual way to reason through the Monty Hall problem is to appreciate that revealing the joke prize behind one door cannot change the location of the real prize. The

probability that the initial choice was correct was initially  $1/3$  and remains so. This probability is independent rather than conditional. Opening a door does change the probability that the real prize is behind it from  $1/3$  to  $0$ , not because the prize has moved but because we can see it was never there. We have new information about that door, but we don't have new information about our original choice. So far, so good. The hard part is seeing that the total probability must be  $1$ , so the probability of the real prize being behind the third door must have gone up from  $1/3$  to  $2/3$ . This probability is conditional on the opening of the second door. This is where our intuition fails. How can seeing what is behind door 2 change the probability of the prize being behind door 3 and not the probability for door 1? Why aren't the new probabilities both  $0.5$ ? If students can appreciate why the probability for one door is independent and the other conditional, then they will have made progress in understanding the concept of conditional probability.

The mechanism that makes the probability of the selected door stay at  $1/3$  and the nonselected door go up to  $2/3$  is subtle. Opening door 2 provides information about door 3 and not door 1 because Monty Hall could have opened door 3 if it shielded a joke prize, but he didn't. He would not have opened the door chosen by the contestant at this time anyway, so his choosing not to open it provides no information. His choosing not to open door 3 provides evidence that increases the probability of it being the correct one.

I find the conceptual argument very convincing, as will most mathematicians, but most students remain skeptical. I suggest having them detail the three cases. We can assume without loss of generality that the contestant has chosen door 1; if not, just put new labels on the doors to make it so. The three cases correspond to the three possible prize locations. In each case, we can see whether switching is successful. For example, if the prize is behind door 1, the host opens either door 2 or door 3 and the player wins by standing firm. If the prize is behind door 2, the host opens door 3 and the player wins by switching to door 2. Similarly, if the prize is behind door 3 the player wins by switching. Two of the three possible scenarios reward the switching strategy.

## 4.7 Conditional Probability Applied to Diagnostic Tests

Many of my students are in premedicine programs, and for them it seemed particularly useful to discuss the mathematics of diagnostic tests, which is an application of conditional probability. The standard treatment of this material is to present Bayes' formula as a theorem. Mathematicians may feel that this is satisfactory because the proof is in the book, although I expect few actually discuss the proof with their classes. As a modeler, I always prefer a conceptual approach to a formula-driven approach. Here, I present a sample tabular method for solving problems in which we know the conditional probability in one direction and want the conditional probability in the other direction. The key idea is that the joint probabilities provide all the necessary information. A two-step approach is to use the given information to obtain all four joint probabilities and then use the latter to obtain any other probabilities of interest.



## 5 Dynamics of Single Populations

This chapter develops the mathematics of discrete and continuous models for a single dynamic quantity. We begin with separate presentations of numerical methods for simulations and graphical methods of analysis before combining both types of model into a single section on linearized stability analysis. We progress from simulation to graphical methods to analytical methods because simulations serve as an informal introduction and graphical methods provide a starting point for the development of analytical methods. Combining discrete and continuous models into a single chapter allows us to draw a sharp contrast between the graphical methods for the two types while also emphasizing the very close connection between the analytical methods. Three lessons follow from these juxtapositions. First, graphical methods for continuous models are very simple, while graphical methods for discrete models are complicated. In later chapters, readers will see that there are no graphical extensions of cobweb analysis to multivariable systems, while phase line analysis extends naturally to nullcline analysis in the phase plane. Second, graphical methods are superior to analytical methods for single dynamic models. Cobweb analysis can suggest stability of cycles of length  $2^n$  with much less effort than analytical methods, and both cobweb analysis and phase line analysis can demonstrate global stability. Third, instability is driven by movement away from a potential long-term solution in both cases, but only in discrete models can instability be driven by rapid movement toward such a solution. Rapid movement toward an equilibrium in a continuous model causes stiffness, which complicates simulation but does not affect stability. These broad themes should be emphasized throughout the chapter.

Analytical solutions are generally not in the toolbox for mathematical modeling with dynamical systems. Of course they can be used for any single autonomous differential equation, but the solution formulas provide no results of value that could not be obtained more easily without them. While methods for solving differential equations are interesting from a mathematical point of view, they serve biologists merely as a barrier to dynamical systems analysis and a reason for employing a discrete model where a continuous model would be better.

### 5.1 Discrete Population Models

This section introduces discrete population models using the solution formula for the discrete exponential model and simulations for other models. Simulations also provide a context for the definition and an informal discussion of stability. Most of the problems in this section use simulations to introduce models that will be analyzed in the problem sets for Sections 5.2 and 5.5. Problem 5.1.6 develops a model to explore the behavior of 2-cycle solutions of the discrete logistic equation. The algebra is complicated, but there are also simulations that use several growth rates to illustrate possible solution behaviors. These simulations could be run without the accompanying theoretical analysis. This problem is of more mathematical interest than biological interest. Discrete models contain the implicit assumption that life events are synchronized, and it is this synchronicity that makes the instability of the discrete logistic map possible. Continuous models should be used for populations whose life history does not impose synchronicity, both to avoid instability caused by model choice and because the analysis of continuous models is more powerful.

## 5.2 Cobweb Analysis

Cobweb analysis is a useful graphical technique for discrete dynamic equations. While most of the information it provides can be obtained by analytical methods, it has the advantage of being intuitive rather than strictly formal.

The key to understanding cobweb analysis is separation of the processes of construction and analysis. I begin with a detailed description of the construction process along with an explanation for why the method does what it does. Once the reader understands the construction, it is an easy matter to understand the analysis.

Mathematicians might be tempted to skip this section, but I recommend against doing so. I use cobweb analysis to motivate the later development of linearized stability analysis. I also find that cobweb analysis is a useful conceptual tool for explaining why rapid change can lead to instability in discrete models but not continuous models.

## 5.3 Continuous Population Models

Mathematicians can help biologists overcome their natural reticence toward continuous models by making them seem less mysterious. This requires us to focus on interpreting differential equations as prescriptions for calculating rates of change, which in turn means that we need to emphasize graphical methods of analysis and simulations while deemphasizing analytical techniques for finding solution formulas. The only solution formula used in this section is that for exponential growth, which can be obtained from a table of derivatives without need for a solution method. Other equations are studied using Euler's method for simple simulations. This method is discussed in the text because it has some conceptual value, and it can be used for simulation on today's fast computers by selecting a very small step size. For any extensive amount of simulation, it is far better to use the classic Runge-Kutta method of order 4. This method is summarized in Appendix C without the mathematical derivation that is far beyond the scope of this book. The method has the advantage of being almost as good as methods used for commercial software while also being simple enough to code manually. I do not recommend using commercial software to run differential equation simulations. It is much easier in any computer programming environment to write one's own code for this method than to track down and implement a professionally written subroutine, and it also removes the mystique from using tools that are beyond any understanding.

The renewable resource model presented in Subsection 5.3.4 is somewhat complicated, but I believe its richness makes it well worth the effort. It serves as a context for several examples and problems and it can be used to explain a number of real biological phenomena. It also illustrates the continuing importance of technical algebra skills, as contrasted with many other analytical skills whose importance has waned with technology. Computer algebra systems do algebra by brute force and are incapable of obtaining results that require any subtlety. The careful rearrangement of the equilibrium equation in Equation (5.3.11) results in the illustration of equilibrium solution cases in Figure 5.3.3. A similar technique is put to use in some of the problems of the section. These problems require the reader to perform algebraic manipulations, but they do not require the reader to determine the appropriate manipulations to perform. Some readers may see enough of a pattern to be able to do similar work on their own, but this is not required in this book.

Many of the problems in this section explore the properties and uses of the renewable resource model (5.3.8). Others introduce the Schaefer fishery model, the Holling type 2 version of the renewable resource model, and a model of lake eutrophication. The eutrophication problem is slightly more complicated than the Holling type 3 renewable resource model and occupies almost two full pages of the book. I would not use it with most classes, but it is an excellent case study for students who have strong mathematical skills and an interest in environmental issues. If I use it with a class, I either make it (along with follow-up problems in Sections 5.4 and 5.5) an extended project or else I dedicate at least one class period to working on it in groups. Of course this takes time away from presentation of new material, but this may be an acceptable trade-off between depth and breadth.

## 5.4 Phase Line Analysis

Phase line analysis is a simple technique that relies entirely on the interpretation of an autonomous differential equation as a prescription for a rate of change in terms of the state of the system. It is a simple matter of identifying intervals, marked out by equilibrium values, that distinguish increasing states from decreasing states, with results for global stability as well as local stability. This comparison of this section with Section 5.2 illustrates the general fact that continuous models have simpler properties than discrete models, in spite of their equations being less intuitive. Because motion on the phase line is continuous, it is impossible for the state of the system to jump across an equilibrium value. In contrast, motion on the state curve of a cobweb plot is done in large jumps, making possible a sequence of jumps that bypass a fixed point in alternate directions without ever being trapped in domain of attraction. This behavior requires two dimensions for illustration, making the cobweb technique much more complicated than the phase line and ruling out the possibility of an extension to discrete systems of two variables. In contrast, students can appreciate while studying the phase line that it should be possible to use two phase lines at right angles to display systems of two differential equations. Motion in the plane generated in this way can occur in a variety of directions, so it is not obvious how such a phase plane could be used.

The problem set for this section is quite limited because of my preference not to include technical exercises for modeling topics. The phase line is too simple to require much practice for its mastery. There are a few problems in Chapter 7 that ultimately come down to analysis of single differential equations, and these will of course make use of phase line analysis.

## 5.5 Linearized Stability Analysis

The methods for linearized stability analysis are derived from the graphical methods of cobweb analysis and phase line analysis by connecting the visual idea of zooming in on a fixed point or equilibrium point with the analytical idea of applying a linear approximation when arbitrarily close to a given point. As generally presented, the stability requirements for discrete and continuous systems seem to have nothing in common, but there actually is an important connection that is revealed by writing discrete models as equations for rates of change rather than new amounts. With similar forms for the types of equations, we see that one of the stability requirements for discrete equations is identical to the requirement for continuous equations that the solution must be attracted to the equilibrium point (fixed

point in the discrete case), while the other requirement is necessary to rule out the possibility of instability resulting from an attraction that is too strong. As we saw in graphical analysis, too strong an attraction matters in the discrete case and not the continuous case, so naturally there is an additional requirement for stability in discrete equations.

The problem set for this section consists primarily of problems that complete investigations started in earlier sections. In each case, the student is asked to determine stability results and confirm that they are consistent with results obtained using simulations and graphical methods. Problem 5.5 determines the requirements for stability of a 2-cycle for the discrete logistic equation; this serves as an excellent example of what can be accomplished by stability analysis, but the calculations are more complicated than what I would expect from most biology students.

## 6 Discrete Dynamical Systems

This short chapter presents discrete linear dynamical systems and the matrix algebra necessary for their analysis. It was originally written before the rest of the book as a stand-alone module for the Research Skills in Theoretical Ecology that my biologist colleague Brigitte Tenhumberg and I created for an NSF-funded undergraduate research program. The students in that course were primarily pre-freshman biology majors. They were generally strong students, but their background in mathematics was limited to what they had seen in high school. The purpose of the course was to provide students with an authentic theory-and-experiment research experience. The overall research program is described in Example 2.2.3, and the full course is described elsewhere.<sup>4</sup> The research focused on the population dynamics of aphids, and we used a stage-structured linear dynamical system model.

The reader will no doubt wonder why discrete nonlinear systems do not appear in this chapter. This topic appears in Section 1 of Appendix A instead, having been placed there for two reasons. Discrete nonlinear systems are more complicated than continuous linear systems, so it makes good pedagogical sense to do the continuous ones first. I also wanted to put this topic in the appendix to deemphasize it. Discrete models are often used for settings more suitable for continuous models. I sometimes see the argument that discrete models should be used when the data is discrete. This is incorrect. The purpose of models is to capture the relevant dynamics of real-world settings, which includes dynamic properties that are not measured as well as ones that are. Discrete models should only be used for populations where life history events are synchronized, such as annual plants with seed banks or fish that spawn at the same time each year. My advice is that the material on discrete nonlinear systems should only be studied after mastering the material in Chapter 7.

There are two additional discrete topics that could have been put in Chapter 6, but were placed in Appendix A instead: Markov chain models and Boolean models. I put these in the appendix for pedagogical reasons; a reader who does not have enough time for these topics and Chapter 7 should choose the latter so as to have a thorough introduction to dynamical systems.

---

<sup>4</sup>See Ledder, Tenhumberg, and Adams, An interdisciplinary research course in theoretical ecology for young undergraduates, in *Undergraduate Mathematics for the Life Sciences: Models, Processes, and Directions*, ed. Ledder, Carpenter, and Comar, Mathematics Association of America, 2013, Chapter 10.

## 6.1 Discrete Linear Systems

The standard pedagogical order in mathematics is definition-theorem-proof-application. This has the advantage of mathematical clarity, but it is of dubious pedagogical value. There are infinitely many mathematical definitions that could be made; in practice, we only define things that we already know will be useful. On pedagogical grounds, we should start by discovering something useful through simulations or thought experiments, and only then should we develop the formal mathematics. This is the philosophy that drives the material in this opening section of the chapter. We start with simulations of a simple discrete linear system. By choosing just the right quantities to graph, we discover the biological fact that a discrete linear system has an asymptotic growth rate and stage structure. Mathematically, these quantities are the dominant eigenvalue and the associated eigenvector. Graphs of ratios of successive populations converge to the dominant eigenvalue, and graphs of ratios of populations of various stages converge to fixed ratios that represent the eigenvector.

The theory behind the methods for finding eigenvalues and eigenvectors is hard to understand because there is no simple concept represented by the determinant. The calculations for eigenvectors are additionally complicated by having an underdetermined system. These difficulties can be overcome for small systems by a careful conceptualization of the problem. We get the standard eigenvalue problem by looking for a vector whose proportions are not changed by the matrix multiplication. An equally valid method is to hypothesize a stage structure that is unchanged by matrix multiplication. We assume initial conditions having this property; this assumption leads to expressions for the populations at time 1, which must be equal to those obtained by matrix multiplication. The problem is simplified by the arbitrary assumption that one of the stages has initial population 1. We cannot do this in a general setting because there is no way of knowing whether an eigenvector has a component of 0. But in the context of population models, the matrices are nonnegative and the Perron-Frobenius theorem guarantees that the eigenvector for the dominant eigenvalue is strictly positive; hence there is no loss in generality in arbitrarily taking one of the components to be 1. This simplifies the mathematical problem by reducing the number of unknowns so that the system is no longer underdetermined. Mathematicians may object to the use of a method that is not fully general; my justification is that I am not advocating the method for practical use but merely using the method to build intuition. In a similar vein, we can obtain the dominant eigenvalue approximately by graphing the characteristic polynomial rather than by a formal numerical method.

Two of the problems are case studies that are worth spending extra time on for a course that emphasizes modeling. Problem 6.1.9 is based on a paper produced by a Research Experience for Undergraduates team at the University of Nebraska. Problem 6.1.10 is an exercise that was part of the Research Skills for Theoretical Ecology course mentioned in the introduction to Chapter 6 of this guide. Any mathematical modeling investigation that begins with a verbal narrative is necessarily limited. Real science begins with observations and experimental data, not a verbal narrative that summarizes what has been learned from this beginning. BUGBOX-population was written as a way of giving students a chance to generate their own observations and data, so that they would have to create the narrative as well as the model. I have used this problem many times with great success. I spend a class period guiding the students through part (a), and then I have them do parts (b) and

(c) outside of class. I spend a second class period helping the students write the computer program to do parts (d) through (f), and they finish the problem at home.

## 6.2 A Matrix Algebra Primer

Section 6.1 presents an informal approach to analysis of discrete linear systems, and Section 6.3 repeats the analysis using the mathematics of matrix algebra. This section develops the matrix algebra background necessary for Section 6.3, focusing on the properties of homogeneous linear algebraic equations.

## 6.3 Long-term Behavior of Linear Models

Having previously discovered the existence of an asymptotic growth rate and stage structure for a discrete linear system and learned about the solution structure for homogeneous linear systems, the reader is now ready for the formal mathematics of the eigenvalue problem. This section is similar to that in any linear algebra book, except that it is limited to nonnegative matrices, for which the dominant eigenvalue is always positive and the associated eigenvector is positive as well. This allows us to avoid the complications of complex eigenvalues and multiple eigenvalues. The point of this section is not to make the reader an expert on the eigenvalue problem of matrix algebra, but to give the reader enough knowledge to be able to manage any matrices that arise in population models.

For those interested in mathematical modeling, there are four extended case studies in the problem set for this section. Problem 6.3.9 looks at a model for a plant with a complicated life history, which appears in Hal Caswell's book on matrix population models. This problem is a good example of how much modeling can accomplish if done carefully. The model presented by Caswell is quite complicated, but the work in this problem results in a much simpler model that produces almost identical results; this is of biological significance because it helps us distinguish between model features that are critical and ones that make very little difference. Problem 6.3.10 is the modeling done in the Research Skills for Theoretical Ecology course mentioned earlier. Problem 6.3.11 is another investigation that grew out of a Research Experience for Undergraduates project. I have simplified the model to make the analysis easier for students, but the overall result of the analysis is unchanged. The conclusion is worth seeing for students interested in wildlife conservation—not all endangered species can be saved. There is little doubt that the Serengeti cheetah will become extinct in the wild. Problem 6.3.12 examines the population dynamics of loggerhead sea turtles that appears in one of the best-known papers on a wildlife conservation issue. Here we see how it is sometimes possible to make a big difference in survival chances for an endangered species, provided the data and modeling are able to identify a critical intervention.

## 7 Continuous Dynamical Systems

The typical reaction from mathematicians looking at this chapter is that it is too difficult for students with only a limited calculus background. In my view, this is because the topics covered in the chapter are generally deferred to a differential equations course taught

to students with a background that includes the full calculus sequence and a course in matrix algebra. This background is needed for a full development that includes all analytical methods, but a lot of useful work in dynamical systems can be done without them. It is not necessary to obtain analytical solutions even for linear systems. It is not even necessary to calculate eigenvalues, as we can obtain the needed qualitative results merely by examining the entries in the Jacobian matrix. Nor is any background in multivariable calculus necessary, as students with a solid understanding of parameters have no difficulty computing partial derivatives.

The material does make some demands on students, but those demands are primarily on algebra and graphing skills rather than calculus or matrix algebra skills. The only calculus skills used in the chapter are ability to interpret a differential equation as a statement about a rate of change and ability to differentiate simple functions of more than one variable. Similarly, only the minimal matrix algebra skills covered in Chapter 6 are necessary as background. In practice, my students have done very well on this material.

While the other chapters in the book have one section that serves as a relatively informal introduction to the theme of the chapter, this chapter has two such sections. Both sections 1 and 2 showcase important modeling techniques; additionally, the graphical and analytical techniques developed in the main part of the chapter (Sections 7.3 through 7.5) require introductory examples with different features.

Most of the problems in the first two sections introduce models that will be analyzed in the problem sets of later sections. Instructors should look at the list of extended case studies that appears at the beginning of Chapter 7 before choosing the problems to use in the first two sections. There should be enough variety to set up problems in later sections, but it is important not to do too many, as each one requires much more time and effort than typical textbook problems.

## 7.1 Pharmacokinetics and Compartment Models

Compartment models made a brief appearance in Chapter 2 and are dealt with in much more detail here. The primary example of the section considers a generic pharmacokinetics model using a narrative of lead poisoning rather than drug interaction. One reason for using this narrative is that the parameter values are such that we can identify a two-component model that produces results almost identical to the three-component model. Mathematical treatments of dynamical systems deal with the problems as original defined, but mathematical modelers should always be on the lookout for approximations that simplify without introducing much quantitative error, as in the teasel model introduced by Caswell and studied in detail in Problem 6.3.9. Both graphical and qualitative methods of analysis are complicated by extra dimensions, so reducing the dimensionality of a problem makes it much easier to study. In the lead poisoning model, I use simulation results to identify the correct simplification and then nondimensionalize the simplified model. The preferred method is to do the nondimensionalization first and use the parameter values to motivate the simplification. This is deferred to a later section, but the calculations needed to derive the simplified model are the subject of Problem 7.1.1. The remaining problems in the section introduce several epidemiology models. In this section, students are asked only to interpret and nondimensionalize models that will be analyzed in later sections.

The choices given for scales to use in nondimensionalization in both this section and the next are not always obvious. In some cases, I have explained them in a footnote, while I have asked students to suggest explanations in other cases. The choices for dimensionless combinations of parameters are more subtle. I have yet to see a treatment of scaling in any book that does a good job explaining these choices; usually the explanation is strictly based on dimensional arguments, which offers no explanation at all since that only identifies an infinite set of possible scales. Scales should be based on a combination of representative sizes and algebraic convenience, while parameter combinations should be chosen first to make as many of them factor out of a differential equation as possible, then to make as few of them large or small as possible, and finally for algebraic convenience. The instructor who is not an expert in asymptotic analysis should realize from this discussion that the correct choice of scales and dimensionless parameters is far beyond the scope of students at this level. I could not attempt to teach students to be expert modelers, but it was reasonable to expect them to be able to do the mechanics of nondimensionalization and to appreciate the advantages of the correct choices made by an expert.

## 7.2 Enzyme Kinetics

Enzyme kinetics is a natural choice for an example of a dynamical system. Many of the students in my class are biochemistry majors, for whom enzyme kinetics is an important subject. At the time they take the course, these students have probably heard of the subject from an introductory biology course, but will not have seen it discussed from either a chemical or mathematical point of view. Even advanced biochemistry students will be familiar with the Briggs-Haldane approximation (usually incorrectly attributed to Michaelis and Menten), but they will have no idea where it comes from. Deriving something that students have had to take for granted gives them a great lesson in the power of mathematics. The enzyme kinetics model is another illustration of the benefits of a proper nondimensionalization using appropriate scales for the variables. Because of differences in initial concentrations and reaction rates, the two differential equations have intrinsic time scales that are considerably different. With a correct choice of scales and dimensionless parameters, the dimensionless system has a small parameter that factors out of one of the derivatives and can therefore be written on the left side of the equation, as in Equation (7.2.11). The presence of this small parameter provides a mathematical justification for the assumption that the corresponding equation is quasi-steady on the long time scale. In fact, the corresponding equation is quasi-steady on the short time scale because changes encoded in the faster equation happen too quickly for the changes encoded in the slower equation to accumulate. The effect of having two time scales is demonstrated numerically in Figure 7.2.1 before being treated analytically in Section 7.2.3.

The problems serve several different functions. The first four supplement the text material on enzyme kinetics. Four of the remaining problems introduce models that will be used as case studies for the methods sections that follow. Of these, Problem 7.2.6 stands out as different from the other three. It turns out that standard analysis methods for dynamical systems are neither possible nor necessary for this model. The only clue for this that is readily apparent is that the equilibria for the system are not isolated. As in Section 7.1, the choices of scales and dimensionless parameters is not always obvious, and students should



not be expected to make these choices themselves. In some cases, it is reasonable to ask them to explain why one set of choices is better than another set.

### 7.3 Phase Plane Analysis

The phase plane is a natural extension of the powerful phase line technique to two dimensions. The method for graphical analysis in the phase plane uses nullclines, which are curves on which one of the two dynamic variables is constant. Phase line plots always provide conclusive evidence of domains of attraction because movement in a one-dimensional world is highly restricted. Nullcline plots can be more or less revealing of system properties, depending on the extent to which the nullclines restrict the more extensive possibilities of movement in two dimensions. Students must learn some subtleties to master nullcline analysis, but this is not overly difficult. Nullcline arguments can often be made with parameter values restricted to intervals rather than specified as numerical values. This is necessary if the results are to apply to large regions of the parameter space rather than being limited to specific examples.

The Michaelis-Menten system from Section 7.2 is an ideal first example for nullcline analysis because the system of nullclines is much simpler than those for the predator-prey systems commonly used as first examples. I prepare students for nullcline analysis by first using the phase plane merely as a device for turning simulation results into an animated graph of the system state. It is a natural extension to ask what we can say about solution curves in the phase plane without having simulation results to work with.

Most written descriptions of the process for constructing nullcline plots initially focus on the nullclines rather than the regions bounded by them. In my experience, students find it easier to determine what I call a compass quadrant in a region than to determine the direction for the heads on the arrows located on a nullcline. They do better when we teach them to start with one region, usually near the origin. Once we know the correct compass quadrant for one region, we can work our way around the plot with a combination of nullcline arrows and compass quadrants. This does not work for the Michaelis-Menten system because nullclines pass through the open first quadrant into the origin. However, it works nicely in most nullcline plots, such as that of Figure 7.3.5. In that example, we can look at linear approximations near the origin to determine that the compass quadrant is southeast. This means that the arrows on the nullcline boundaries for the region must point either down or to the right. It is easy to distinguish these, as the nullclines have already been identified as to type. In my own practice, I draw vertical marks across the nullclines when I plot them in the first place. I then start with compass quadrants and add arrowheads to the nullcline ticks by working my way to those ticks through the adjoining regions.

The conclusions that can be drawn from nullclines depend on the topological properties of the flows across the nullclines. A no-egress region traps solution curves in a two-dimensional analog of the way solution curves are always trapped in the one-dimensional phase line. Absent any no-egress regions, strong conclusions about stability of equilibria cannot be drawn. Nevertheless, nullclines are in a sense more powerful than linearized stability analysis; when stability is proven by a nullcline plot, the stability is clearly global for a portion of the phase plane.

Problem 7.3.1 illustrates how we cannot draw stability calculations from nullclines in the absence of no-egress regions. It does this by exhibiting a system in which the only parameter

represents a ratio of time scales, leading to a common nullcline plot but parameter-dependent stability. Without analytical methods, we can only determine the stability by numerical simulation; the problem is completed in the next section with analysis of the Jacobian matrix.

## 7.4 Stability in Linear Systems

Two sections are devoted to the analytical technique for local stability analysis. In this first section, we consider only linear systems, such as the lead poisoning model from Section 7.1. This puts the use of eigenvalues to determine stability in this section, while postponing the linearization component of the analysis to the next. Eigenvalues are sometimes the easiest path to stability results, but often it is easier to work directly with the system matrix by means of the Routh-Hurwitz conditions. For some reason I do not know, many treatments of stability analysis in texts only presents the eigenvalue method. This may be a case of mathematical bias against certain methods that I have noted in discussing earlier chapters. Here there does not seem to be any reason for such bias, as the theorem that justifies the Routh-Hurwitz conditions is easy to prove in the  $2 \times 2$  case and not overly difficult in the  $3 \times 3$  case. Perhaps the method is overlooked by some mathematicians because it does not scale up to higher dimensions the way the eigenvalue method does. This is an argument for including the eigenvalue method rather than an argument for discounting the Routh-Hurwitz method. Good modeling practice can often reduce larger systems to smaller ones, and any system of three components is easy to analyze with the Routh-Hurwitz conditions. Higher-dimensional systems require more conditions and more work to check them, but it is still more difficult to find the eigenvalues without specifying all parameter values.

The problem set for this section is quite limited because very few real biological models are linear. It is important not to spend too much time on this section. It should be seen primarily as background for the following one.

## 7.5 Stability in Nonlinear Systems

This section provides the principal tool for determining local stability of systems, by means of the eigenvalue analyses of the Jacobian matrix evaluated at the various equilibria. Again, it is usually more efficient to use the Routh-Hurwitz conditions rather than to compute the eigenvalues, the exception being the case where there are enough zeros in the Jacobian to decouple the eigenvalue problem into smaller dimensional components. There are usually one or more equilibria in which the eigenvalues can be determined by inspection.

The problem set consists primarily of case studies that were begun in Sections 7.1 and 7.2. Some are quite detailed, and instructors may want to assign only parts of problems.

## 7.6 Primary HIV Infection

I changed my mind several times about whether my treatment of the classic Perelson HIV model should be written in the text as a complete case study or broken into components and placed in the problem sets. I finally decided to distinguish this particular case study in the text for two reasons. First, there are plenty of other case studies broken into components, so

there was no great need for another. Second, this case study brings together all of the themes of the chapter into one grand summary. I often find that I can improve on standard published models by applying modeling techniques not done in the original investigation. These minor improvements are not publishable as papers because they produce no new results, but it seems to me that they are worthy of publication in some venue because they are simpler than prior treatments. I consider it to be analogous to papers that consist of new proofs of previously-proven theorems, which I consider to be generally worthy of publication on pedagogical grounds even if the improvements are relatively minor.

Because the treatment of the HIV model is intended as a complete case study, there is limited scope for problems based on the text material. Instead, I have crafted a collection of problems based on another physiological model dealing with the immune system. I discovered this general area of modeling at a talk by Angela Reynolds, one of the co-authors of the paper from which these problems were drawn. Where almost all other examples in the book are concerned with systems that are completely known, these problems are concerned with a system that remains a topic for active research. The models presented in the problems are reduced versions of models presented in the Reynolds et al paper, which are all reductions of a full immune system. There is obviously a place for comprehensive models, but there is also a place for simplified models that examine subsystems of larger systems.

## A Additional Topics in Discrete Dynamical Systems

The three topics that appear in this appendix could have been placed in Chapter 6 instead. I did not do so for various reasons. Section A.1 is analogous to Section 7.5 on continuous nonlinear systems. However, the mathematics of discrete systems is more complicated than that for continuous systems; in particular, graphical methods are weaker or absent and stability conditions are more complicated. Consequently, discrete models should be used only when rendered necessary by synchronization of life history events in a population. The fact that data is frequently collected at discrete times is not an argument in favor of discrete modeling. Indeed, my colleague Bo Deng argues that it is better to choose random times for data collection so as to avoid effects that are mere artifacts of the data collection. Because of these difficulties, it makes good pedagogical sense to study continuous nonlinear systems before discrete nonlinear systems. The same argument does not hold for linear systems, since the matrix algebra methods that arise naturally with discrete linear systems are needed for continuous linear systems.

The remaining sections are on topics that are peripheral to the central theme of Part III. I included these topics for the benefit of those readers who want an introduction to them, but it seemed a poor choice for their placement in Chapter 6 to interrupt the logical flow of the dynamical system narrative.

### A.1 Discrete Nonlinear Systems

Chapter 7 features three sections on analysis of continuous systems, while the appendix contains but one section on analysis of discrete systems. The first analysis section in Chapter 7 contains graphical analysis for which there is no counterpart in discrete mathematics,

except in the sense that Boolean models can be thought of in such terms. Two sections were needed for linearized analysis because of the need to present material on both the analysis via eigenvalues and the construction of the Jacobian matrix for a nonlinear system. Only one section is now necessary for linearized analysis of discrete systems because of the presumption that the reader has learned about linearization and the Jacobian in Section 7.5. Readers who prefer to disregard my advice to study continuous systems first will need supplement Section A.1 with a study of linearization and the Jacobian matrix.

The material in the section is generally similar to that in Section 7.5 aside from the inclusion of the Jacobian as a new topic in the latter. It is somewhat harder to obtain general results for discrete systems with arbitrary parameter values, but it is not impossible if the systems are small. Most of the examples in the section and problem set are systems of just two components. We can find the eigenvalues for these systems to determine stability; alternatively, we can use the Jury conditions, which are analogous to the Routh-Hurwitz conditions, albeit a little more complicated.

## A.2 Markov Chains

I wanted to include one topic in the area of bioinformatics since this is a an important area where mathematics is used for biology. I chose the problem of determining phylogenetic distance because it uses matrix methods developed earlier in the book and because the biological question of how closely species are related is of general interest. The presentation of Markov chain mathematics is focused on the specific problem of determining phylogenetic distance using the Jukes-Cantor model, which is the simplest of the models. The Markov chain concept is simple enough, as is the general idea that phylogenetic distance is greater than total genome change because some mutations reverse other changes, but the calculations necessary to produce the final result are complicated. I recommend omitting that portion of the section unless the students have very strong algebra skills.

## A.3 Boolean Algebra Models

I also wanted to include a brief introduction to Boolean algebra models, and I eventually decided to place it in the Appendix so as not to interrupt the flow of the dynamical systems chapters. As with the rest of the book, my focus is on concepts and modeling. For this reason, I restricted consideration to Boolean networks with only three components. With only eight possible states, it is not difficult to work out the graph by hand, and there is some conceptual value in doing so. Adding another component doubles the number of states, which greatly increases the work needed to produce the graph and poses additional problems in finding a layout that makes the graph readable. In practice, people use software to produce the graph. I would have done this if the goal were to analyze real gene regulatory systems rather than to develop the concepts of Boolean network analysis.

I find the notation for Boolean networks to be difficult to read. I have to keep reminding myself which symbol refers to which logical operation and which components are activbe for a given state. I have no solution for the first difficulty, but I believe the second difficulty is solved by the alternative notation I introduce for the labeling of the states. It takes me less time to interpret  $XZ$  as the state in which  $X$  and  $Z$  are present and  $Y$  absent than

the corresponding notation of  $(1, 0, 1)$ . I'm sure that people who work in the field quickly internalize the notation, but I think that students will appreciate the greater readability of the alternate notation.

The problems focus on a specific application of Boolean network analysis. Suppose we know the effect of each component on the other. A particular component can be promoted or inhibited by any of the other components. In cases where production of a component is affected by presence or absence of more than one component, the logical connection between the relationships may not be clear. For example, we may know that both  $X$  and  $Y$  promote  $Z$  but not know whether both must be present, either can be present, or only one must be present. These three cases lead to different sets of recurrent states, so each makes a different prediction that could be tested by experiment.