

Preface

"Science is thought to be a process of pure reductionism, taking the meaning out of mystery, explaining everything away, concentrating all our attention on measuring things and counting them up. It is not like this at all. The scientific method is guesswork, the making-up of stories. The difference between this and other imaginative works of the human mind is that science is then obliged to find out whether the guesses are correct, the stories true. Curiosity drives the enterprise, and the open acknowledgement of ignorance."

(Lewis Thomas, *Sierra Club Bulletin*, March/April 1982, p. 52)

The above quote from Lewis Thomas is a wonderfully succinct summary of the scientific process. It includes two main components that directly relate to the objectives of this text: the "making up of stories" and how we "find out whether the guesses are correct, the stories true." The "stories" we tell in science are called theories and models. Theories may be thought of as a general summary of how the world works, developed from many observations (in biology these would be in natural systems and in laboratory settings) and allowing us to generate hypotheses about what we expect to occur under different circumstances. Models are elaborations of particular aspects of theories, developed to elucidate specific issues that the theory can address.

A major theory in biology is that an organism's characteristics (morphology, size, physiology, behavior, etc.) are shaped by the processes of natural selection. Organisms with characteristics that enable them to better survive and reproduce are selected for. This means that, over time, all else being held fixed, organisms with those characteristics will increase in frequency, *if* such characteristics are heritable. Here, heritability means that the tendency to display such characteristics is possible to pass on to the organism's offspring. This is the essence of evolution by natural selection, and this general theory has produced an enormous number of hypotheses that have been found to be accurate from observations and experiments on a vast array of organisms and biological systems.

A large number of models rely on the theory of evolution to make predictions about the genetics, behavior, morphology, and relationships between organisms. These models allow us to predict the genetic response of bacteria populations in a human or livestock host to antibiotic use, the changes in strains of the HIV virus in an infected individual, the circumstances under which we expect to find organisms that reproduce once during their lifetime as compared to

those that reproduce several times, and what shapes and sizes of leaves we expect to find on plants in desert versus those in temperate systems.

The second component of the above quote deals with determining whether our scientific stories are "correct." This is done by using data of all types to determine whether the model hypotheses and predictions are consistent with the data. We call this process "testing" or "evaluating" models, and the first part of this text deals with some of the basic methods for using data to test models. This is part of the field of statistics, and an objective in this text is to help you appreciate how we can use the methods of descriptive statistics to summarize and display data. It is then possible to use these data summaries to evaluate model hypotheses. Throughout this text, we will use data to point out the utility of mathematics in investigating biological questions. We have chosen examples with data from across the biological sciences, but we make no attempt to cover the huge breadth of examples that could be used to illustrate each mathematical topic.

Models in biology can be of several different types. They can be expressed verbally, mathematically, or in computer language. Models can also be "real" in the sense that biomedical science often uses animal models as substitutes for humans or uses cultures of cells to investigate responses that would be more difficult to determine from experiments on whole organisms. Similarly, rather than trying to carry out experiments on whole ecosystems, ecologists use microcosms and mesocosms (think of these as aquaria and terraria with different collections of species within them) to determine how the ecosystem might respond to changes, such as harvesting.

In using any of these models, we should ask the same basic questions: what is the objective for the model (e.g., what do you want to use the model to learn about), and how you would assess whether the model is useful for the purposes for which you intend to use it? In biomedical science, for example, there is a large literature on the use of animal models that provide some guidance as to whether a particular animal (or cell) is an appropriate model to use to address a question for which the ultimate goal is to relate the results to humans. Unfortunately, there is little agreement on the best ways to determine whether a particular model is "correct," and so mice may be used as a surrogate for humans, for example, in evaluating the efficacy of a new drug, but the response of the mice may be quite different from what a human response would be to the drug.

Our objective in this text is to help you learn about the variety of mathematical methods used to create and evaluate models in biology. The general aim of the text is to show how mathematics and associated computational tools may be used to explore and explain a wide variety of biological phenomena that are not easily understood with verbal reasoning alone or from simply analyzing the data that come from experiments. We provide an introduction to mathematical topics that have been found to be of use in analyzing problems arising across the biological sciences. So we expect that it will be useful if you are a student in biology, agriculture, forestry, wildlife, veterinary science, pre-medicine, or another pre-health profession. Our goal is to provide an overview rather than a detailed introduction to any particular topic. We assume that you have had prior exposure to high school algebra, geometry, and trigonometry.

Another objective of the text is to point out the utility of computational methods and to encourage you to develop a basic understanding of the underlying ideas of computer programming. While you have likely been using computers throughout your education and in many facets of your everyday life, we suspect that a large fraction of the readers of this text will have had little if any prior experience with the underlying programming (often called coding) that determines what a computer does. Computational methods have become more and more prevalent throughout biology, and we believe it is important for modern biologists to be aware of the basic ideas of computer algorithms. For that reason, we have chosen to illustrate examples throughout the text using the computer tool Matlab. This provides a means to address more complicated biological questions than we could otherwise do with only hand calculations or

using a standard numerical calculator. We also use Matlab to help you learn about some of the basic coding constructs that underlie all digital computers.

Modeling is the creative process by which a model is developed. This includes determining the objectives for which the model is being developed, the choice and construction of the model, the evaluation criteria to be used to determine if the model is useful for the purposes for which it is being constructed, and an iterative procedure for model modification (or elimination if the approach is deemed unsuitable) to meet the objectives. Modeling is often inherent in the scientific process of observation, identifying patterns, hypothesis formulation, setting criteria to evaluate the hypotheses, abstracting the key features of the system under consideration, carrying out further observation or experiment, and evaluating the hypotheses based on the chosen criteria. The abstraction carried out in the scientific process typically involves a model or set of models that are applied to suggest appropriate experiments or observations and infer the implications of the assumptions inherent in the abstraction. The modeling process can point out the need for more data in order to create a useful model, and, as is true of science, models are modified regularly to incorporate new features and account for new data.

While this text includes many models, it is not our objective to teach you how to be effective in doing the modeling necessary to develop new models. Rather, we expect that the concepts and skills we incorporate in the text will help you appreciate the variety of components of the modeling process, including assessing hypotheses based on data, formulating a mathematical description of a system based on assumptions, and determining the implication of the assumptions by analyzing the model. The models we use throughout the text are designed to help you develop intuition for the concepts and techniques considered. So we use models mainly to motivate the utility of the mathematics being discussed and to illustrate how the mathematics can provide insight in the life sciences and be related back to observations. Given the proliferation of mathematical models in every area of the life sciences, our intent is also to help you read critically the literature in your chosen area. One of the benefits of learning mathematics is that the same concepts and methods often apply in diverse areas so that you will see the same mathematical tools used in very different areas of biology. Many fine texts that present models in particular subdisciplines of the life sciences are available. Our expectation is that after you complete this text, you will be well prepared to expand your expertise by reading through other, more specialized texts on modeling in biology in areas as diverse as pharmacokinetics, population ecology, cell signaling, and genomics.

Summary of Goals for This Text

- Develop a reader's ability to quantitatively analyze problems arising in the biological areas of interest to them.
- Illustrate the great utility of mathematical models to provide answers to key biological problems.
- Develop an appreciation of the diversity of mathematical approaches that are useful in the life sciences.
- Provide experience using computer software to analyze data, investigate mathematical models, and provide some exposure to programming.

The methods utilized in this text to meet these goals include the following:

1. Encouraging hypothesis formulation and testing for both the biological and the mathematical topics covered.
2. Encouraging investigation of real-world biological problems through the use of data.

3. Reducing rote memorization of mathematical formulae and rules through the use of Matlab.
4. Providing biological motivation for each main mathematical component introduced through a set of examples that are returned to regularly.

Note to Instructors

Although long considered the language of science, mathematics has slowly become more integral to biological fields as it has been historically in chemistry and physics. Although there is a long history of mathematical approaches in certain areas of biology, it has been only within the past two decades that an appreciation for the ubiquitous utility of mathematics has arisen across the life sciences. Several national reports (e.g., *BIO2010: Transforming Undergraduate Education for Future Research Biologists* from the National Research Council in 2003, the 2009 HHMI/AAMC report on *Scientific Foundations for Future Physicians*, and the NSF/AAAS 2011 *Vision and Change in Undergraduate Biology Education* report) have pointed out the need for quantitative and computational components to be closely linked to all areas of biology education.

Happily, a large number of new texts are devoted to the interface of mathematics and biology, with several of these focused on entry-level courses for life scientists. In most cases, these texts focus first on calculus topics. The pedagogy of the courses for which this text has been developed over the past two decades is that an entry-level course should provide an introduction to the variety of mathematical topics of importance in many areas of biology and that this goes well beyond the calculus that is typically part of an entry-level course sequence for undergraduates in science. The course was fostered by discussions at two workshops at the University of Tennessee, Knoxville, in the early 1990s at which gatherings of leading quantitative biologists encouraged a broad-based collection of material be included, with ample motivation by biological data. Additional information gathered by LJG noted that one of the most critical skills that biology faculty felt their undergraduates needed that was not being adequately addressed was interpreting and making graphs. Thus, this text starts with biological data and its analysis, developing basic concepts in descriptive statistics that are rarely included in entry-level math courses for life science students. Additional discussions with numerous faculty led LJG to compose a set of quantitative concepts that ideally all undergraduate life science students would be exposed to during their undergraduate careers. This set of concepts was included in the BIO 2010 report and includes rates of change, scale, equilibria, stability, structure, interactions, stochasticity, visualizing, and algorithms. This text provides introductory material for each of these concepts in a relatively simple mathematical framework using questions motivated by biological data.

One of the major themes of the undergraduate mathematics education reform movement over the past several decades has been the "Rule of Four." This acknowledges that students have diverse learning styles and that it is appropriate to incorporate alternative methods to enhance quantitative concept and skill comprehension through four complementary approaches: symbolic, numerical, graphical, and verbal. From our experience guiding life science students through quantitative courses over many decades, we have found that an additional approach can be beneficial: providing motivation for the concept or skill through observations and data on biological topics about which the students have some direct experience (e.g., heart rate changes during exercise, drug dosing, and growth of pets or gardens) or intuition from our activities as living beings (e.g., population changes and seasonal and diurnal changes of organism behavior). We thus posit that for interdisciplinary courses and readers, for which this text is designed, it is appropriate to follow a "Rule of Five" for conceptual development that includes biological data and examples that enhance intuition regarding the concepts.

In the units of this text, we try to incorporate each of these five components of learning development in the expectation that this will not only cover all potential learning styles of readers but additionally reinforce appreciation for the utility of mathematics by making explicit connections to biology. In this process, we realize that many readers may well have limited formal exposure to the science of biology (as opposed to our own direct experiences as organisms), so we have tried to limit inclusion of biological areas that typical students would not be exposed to until beyond the first year of their curriculum and about which they likely would have little intuition. This necessarily has focused our examples at levels of the biological hierarchy above the molecule and cell, with little emphasis on biochemistry, systems biology and biophysics of the cell, developmental biology, and neurobiology. We encourage instructors to enhance the biological topic coverage with examples from their own experience that may assist in student comprehension and would definitely appreciate suggestions based on your experiences in introducing these.

Using This Text

The topic coverage and approach we follow is concordant with suggestions in the reports on U.S. undergraduate life science education mentioned above. The text can be utilized in several ways to help institutions meet learning goals for life science students. This text has been used in draft form over the past 5 years for a two-semester, 3-credit-hour sequence taken by many life science and agriculture students at the University of Tennessee, Knoxville. In this form, it has been taught by a broad range of mathematics and life science faculty, instructors, and graduate students.

The format has been both one with small sections of approximately 30 students meeting three times weekly and one with a large-lecture format for up to 200 students with two large-class 50-minute meetings per week and a single 85-minute lab/practice session. The coverage has been essentially the entire text, with the first course covering Units 1 through 3 and the second covering Units 4 through 7. A summary of the course structure and discussion of the challenges in teaching a course covering the diverse topics included is in Gross [31].

While many national reports on life science education encourage incorporation of a diversity of quantitative concepts and skills as undergraduate learning objectives, a large fraction of formal course requirements for life science majors include only calculus. However, the MCAT exam structure has changed to include data-based and statistical reasoning while noting that an understanding of calculus is not required. We recognize that as a foundational quantitative approach in much of modern science and engineering, many institutions wish to maintain a calculus emphasis in their life science curricula. Thus, we have tried to make the text portion that focuses on continuous methods, Units 4 through 7, self-contained so that it can be utilized for a single-semester course if desired. Drafts of the text have been used in this manner for several years at Unity College, for which the course has begun with Sections 5.1 and 5.2 on sequences and their limits and then moved to Unit 4. In this case, if Matlab is being used, the instructor may wish to incorporate a initial lab period that covers the basic Matlab procedures and constructs included in earlier units.

If more class time is available in a single introductory course for which a calculus emphasis is desired, we suggest starting with Unit 1, continuing with Sections 5.1 and 5.2, and then moving to Units 4 through 7. This will provide background on many of the data analysis skills required for the MCAT (e.g., using, analyzing, and interpreting data in figures; using measures of central tendency and dispersion to describe data; and recognizing and interpreting linear, semilog, and log-log scales and calculating slopes from data found in figures). Linking this with a biostatistics course that includes hypothesis testing would then provide sufficient background for all quantitative components of the MCAT while also providing an overview of calculus. An alternative, for a single course that emphasizes noncalculus topics, is to include Units 1 through 3.

This may then be followed with a biostatistics course for those students who wish to prepare only for the MCAT, a course based on Units 4 through 7 of this text, or a calculus course or sequence developed for a broader audience of students.

We have incorporated Matlab throughout this text both to introduce basic constructs used in computer coding and to introduce the concepts of algorithms that are applied throughout computational methods used in every area of science. However, the text components related to Matlab are included at the end of each chapter and may be skipped. Because of the ongoing growth of the use of the R language in many areas of biology, an online adjunct to this text is available that uses R rather than Matlab for all the sections at the end of chapters.