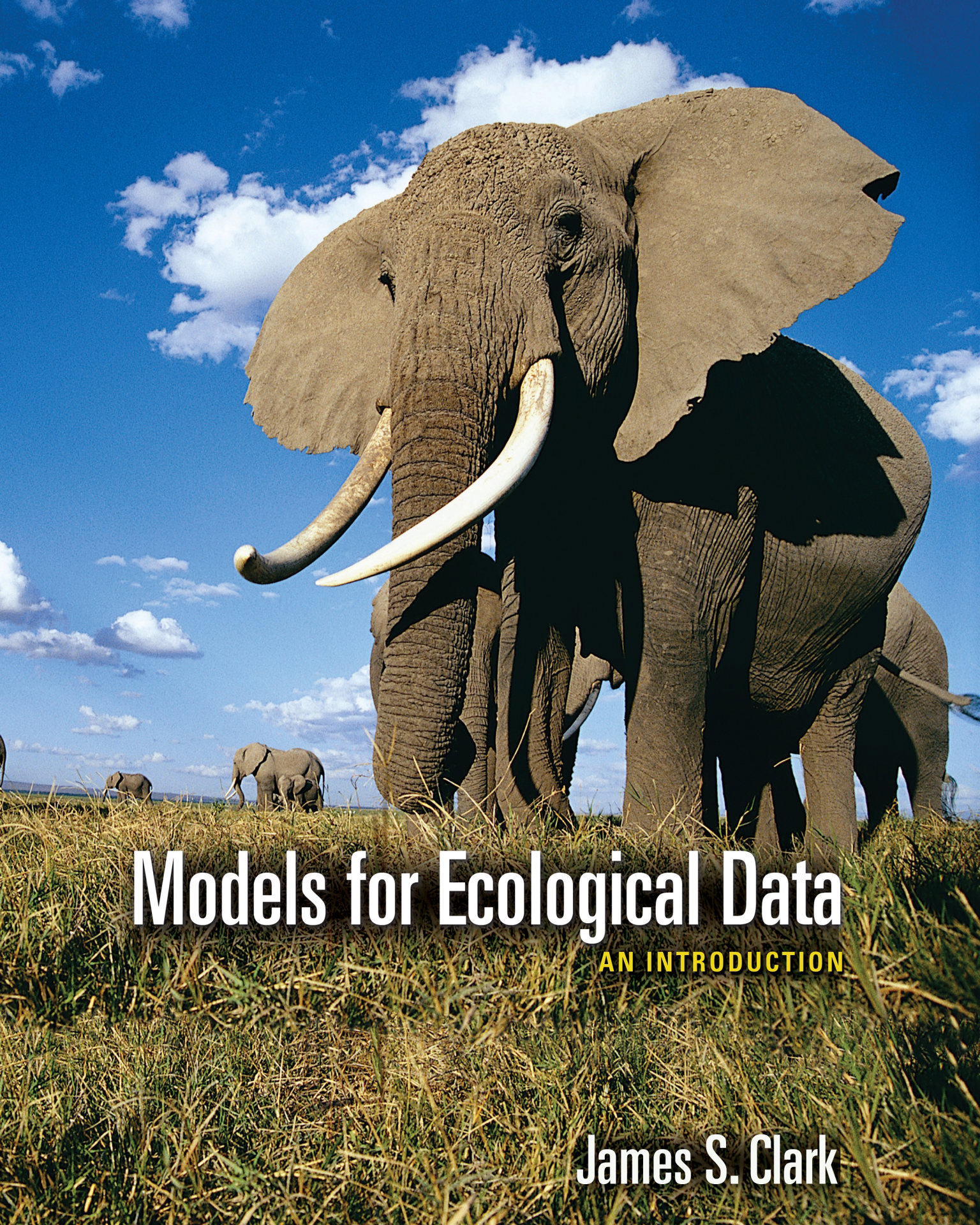
Models for Ecological Data
AN INTRODUCTION
James S. Clark

Models for Ecological Data

Models for Ecological Data

An Introduction

James S. Clark

PRINCETON UNIVERSITY PRESS

Princeton and Oxford

Published by Princeton University Press, 41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press, 3 Market Place, Woodstock,

Oxfordshire OX20 ISY

ISBN-13: 978-0-691-12178-9

ISBN-10: 0-691-12178-8

Library of Congress Control Number: 2006930145

British Library Cataloging-in-Publication Data is available

This book has been composed in Sabon

Printed on acid-free paper. ∞

press.princeton.edu

Printed in the United States of America

1 3 5 7 9 10 8 6 4 2

Contents

Preface

The environmental sciences are witnessing a transformation in how models and data are used to draw inference and make predictions. Not more than a decade ago, ecological data were typically thought of as the products of a controlled experiment designed to test a narrow hypothesis in an abstracted setting. Statistical references used by environmental scientists contained little else. Yet data that meet the assumptions of classical statistical models remain scarce, limited by our capacity to control the environment, the temporal and spatial extent we can afford to examine, and the ability to see relevant variables. Beyond philosophical differences over, say, frequentist versus Bayes, is a more fundamental obstacle: classical methods make demands that environmental data rarely meet.

Meanwhile, environmental questions have expanded to embrace scales of space and time that could draw on suborganism-level (e.g., leaf-level) experiments to remote sensing, provided there was a framework to assimilate such mismatched data sets. Information from uncontrolled and partially observed processes has become a common basis for ad hoc inference and decision. The evolving concept of "data" extends even to such derived products as model output. Many types of information are accumulating faster than they can be assimilated in models; weather records, gene sequences, and remotely sensed imagery are examples. Environmental scientists require the capacity to draw inference based on large amounts of information, much of which cannot be shoehorned into classical models.

With the challenges have come some important new tools. The revolution in how data are used for inference and prediction is driven not only by the scale and complexity of issues confronting environmental scientists, but also by modern techniques for modeling and computation. Environmental scientists are discovering the capacity to build high-dimensional, yet coherent, models at relevant scales, and to accommodate the heterogeneous information that comes from diverse sources and was previously treated in an ad hoc fashion. Together, ambitious goals and emerging machinery have fueled growing demands on how to combine process models with data having context that can vary in space, in time, and among sample units. The approaches are technical and developing rapidly. Critical application of software requires a background that includes distribution theory and basic algorithms. This book is motivated by the need

for an informal treatment of topics that could be accessible to environmental scientists who have some quantitative skills, but lack deep background in statistical theory.

In writing this book, I have tried to reach at least two audiences in the environmental sciences, including the motivated graduate student who has taken introductory courses in statistics and calculus, but has forgotten most of it, and the practicing investigator, who is faced with challenging problems. In my experience, these two audiences often possess some quantitative skills, but are stifled in their efforts to digest methods of modern inference and prediction available from the statistical literature. This introduction to concepts and methods includes material that would typically be contained in graduate-level courses in statistics and applied mathematics. The coverage has been shaped by experiences with students in my Ecological Models and Data course and in the summer institute on Ecological Forecasting at Duke University and by colleagues with whom I have collaborated and shared ideas.

In the years since the writing of this book began I have come to believe that the most difficult expectation to dispel is the one about how it should provide fodder for pithy graduate seminar discussions on the philosophy of statistics. Some readers will be disappointed not to find within these pages emphasis on the controversy common in ecological writings on Bayes. The anticipation for such a book has so dominated the communications I have received from ecologists concerning its preparation, that I will briefly explain why this book may not be what some readers expect.

The grounds for these expectations are several, but I see two factors playing an inordinate role. First is the fact that ecological discussions of Bayes have largely picked up on traditional debates in the statistics literature. These debates have been, at times, so heated (and entertaining[1],) that they have tended to overwhelm the technical developments that have breathed new life into a broad aspect of statistical practice. For the nonpractitioner (most environmental scientists fall in this category), it may be hard not to leave with the impression that "what's new" is changing philosophy.

A second and more subtle factor behind the expectation that new methods can come largely through a seminar format may be an entrenched view that ecologists can do statistics without becoming too involved in technical details. Modern Bayes demands a degree of sophistication with distribution theory that is lacking not only in the statistics courses typically available to ecologists, but also in the math courses taken at the Ph.D. level. The motivated student may gain facility with the deterministic models encountered in linear algebra and differential equations courses, while missing the stochastic models needed for modern Bayes. Even an introductory course in probability and stochastic processes will focus on theory, rather than applications that involve algorithms. In my view, the debates on philosophy can suffer from misunderstanding of some basic technical concepts.

---

[1] For example, Lange et al. (1992) responded to the claim that they were "trying to turn statisticians into Bayesians." Really, they were just trying to turn "frequentists into statisticians."

In contrast to what is often written about statistics in the ecological literature, many statisticians have recently been preoccupied with tools, rather than philosophy. The emergence of modern Bayes in so many disciplines has, to some degree, eclipsed older debates. Ironically, the excitement generated by technical developments has prompted a number of Bayesians to remind colleagues that Bayes is *more* than just machinery. Throughout, I focus on pragmatic concerns, including modeling, computation, and applications, with less emphasis on philosophy than some ecologists may expect from a treatment of environmental statistics. I introduce both classical and Bayesian frameworks, emphasizing how we might want to shift from one to the other as complexity increases. I do not dwell on philosophical divides, largely because I think most students who begin to use new tools will develop an appreciation of these issues along the way. I think that the shift to a Bayesian view of inference is so inevitable that its adoption does not really need any more arguments from people like me. In short, this book is intended for those who want to learn new techniques, including a basic introduction to statistical modeling and computation, emphasizing process models from the environmental sciences. It is a text, rather than a monograph, because the philosophical debates are available elsewhere—in the ecological literature, Hilborn and Mangel (1997) is exceptionally good—and the tools themselves are the source of current focus. A complementary volume edited by Clark and Gelfand (2006b) focuses primarily on applications and assumes a greater level of sophistication.

This effort to introduce state-of-the-art inference and prediction techniques to a nonstatistical audience is behind the nonstandard content and organization. My personal experiences with graduate students and colleagues led me to believe that a text on this material would need to accomplish several things. First, it would have to span a broader range of sophistication than is typical of most ecology and statistics references, starting at a basic level. Many students with the capability and motivation cannot get started with even the best new texts on modern Bayes, and there are now several. On the other hand, if a text starts and remains at a basic level, students fail to appreciate what it can do for them. This is certainly the conclusion many take from ecological literature on Bayes, which seems to say that Bayesians do a lot more work to get about the same results. I have been most impressed with the need for an approach that makes the connections from basic models to environmental application, with plenty of opportunity to see how to get there. Throughout, I have attempted to present material in terms that are as nontechnical and jargon-free as possible. Nonetheless, much here will challenge students entering graduate programs in environmental sciences. This text assumes that the reader is committed to investing seriously in some technical background.

So now for the nonstandard organization. A typical text with this degree of emphasis on statistics would start with probability models and other technical background. I have taken this approach in the past and found that it mostly convinces environmental students that this stuff is no fun and will take more time than they have available to them. More effective for me has been the treatment of the technical tools as progressive goals to be assimilated

incrementally. Rather than starting with the hardest topics, I provide extended appendixes and encourage students to skim them repeatedly. Even the background distribution models can be most accessible when they come in the form of computer applications that are backed up by application.

The first eight chapters can be the basis for a self-contained graduate course, such as my Ecological Models and Data at Duke University. The first two chapters are intended to move rapidly through some motivation, basic concepts, and process models, providing an introduction only. Clearly, there are whole texts devoted to this material. In Chapter 2, my emphasis is on the broad outline of how one might think about process models, with a general introduction to applications and behavior. In my course, I move quickly to inference, which begins in Chapter 3. This is the point where students start to feel more engaged and make connections to data. By the end of my course (Chapters 1 through 8 of this book), many students progress to the point of writing their own Gibbs samplers. Throughout, I attempt to fill in many of the steps that seem obvious to the practitioner, but can baffle the nonstatistician. I emphasize individual goals for the course, with computer labs leading directly to student independent projects. The computer lab manual that accompanies this text, *Statistical Computation for Environmental Scientists in* R, contains substantial cross-referencing with this text. The student projects assure that students apply the tools and move at their own pace.

Chapters 8 through 10 are more advanced. Here I assume additional sophistication and include fewer of the technical details. These chapters provide an overview of the complexity that can be addressed with modern techniques. For some topics (e.g., spatio-temporal models), I have resorted almost exclusively to primary statistics literature. Even with advanced material, I have attempted to explain the basic modeling concepts and how one can begin to develop algorithms for computation. These chapters represent a point of departure for some of the recent excellent texts in this field.

Finally, I have to include a few remarks on literature coverage. This is intended as a text, not a review of either statistics or ecological modeling. I do not attempt to reference all relevant papers. Indeed, the many examples overemphasize work from my own lab for the simple reason that I have worked extensively with these data sets and models and can readily summarize them in this text. Environmental data sets are inevitably complex, far more so than is generally acknowledged in publications. I have tried to emphasize tools to deal with these issues, best done with examples I know well. The coverage also diminishes after 2003. The draft of this book was essentially completed by early 2004. I have included some references through 2006, but these were primarily already known to me before they were published. Thus, the treatment of publications in 2004 and 2005 is necessarily more cursory than that of earlier works.

Acknowledgments

I have benefited from the interactions with many students and ecological colleagues and from the indulgence of several superb statisticians, who have tolerated my meddling in their discipline. Foremost, I have to thank lab members, present and past, who have reviewed drafts of the text, some of them repeatedly. These include Brian Beckage, Kendrick Brown, Phil Camill, Mike Dietze, Michelle Hersh, Janneke HilleRisLambers, Ines Ibanez, Shannon LaDeau, Jason Lynch, Jason McLachlan, Jackie Mohan, Mike Wolosin, and Pete Wyckoff. I owe much to students in my Ecological Models and Data course, who have provided fodder, in the form of interesting modeling problems, while serving as experimental subjects for material in this text and the accompanying lab manual. I thank Peter Morin and Fred Adler for feedback on earlier drafts of the full text. An anonymous reviewer made many valuable suggestions.

Of my statistical colleagues, I must especially thank Alan Gelfand, Montse Fuentes, Michael Lavine, and Chris Wikle. Alan has been a continuing inspiration with his support and friendship. I have abused Chris Wikle with more than one draft of the entire text, and he has been incredibly generous with his time and feedback. I have had the benefit of significant feedback from Michael Lavine over the years and some very constructive collaborations. Montse Fuentes has been supportive through the latter stages of this project. Kate Calder and Chris Pajorek provided very insightful thoughts on presentation of the material in Chapter 10. At various times along the way I have benefited from the additional inputs from Dave Higdon and Peter Mueller. Additional feedback came from Brad Carlin and Tony Ives.

Finally, Chantal Reid has provided valuable criticism and support throughout.

Part I Introduction

The first two chapters lay out goals and challenges and introduce basic elements of models used for ecological data. In Chapter 1 I discuss issues that arise in the development and application of models for environmental inference and prediction. In Chapter 2 I apply some examples from population dynamics as the basis for introducing the different types of models used for ecological processes. These process models are combined with data and parameter models in Parts II, III, and IV.

1 Models in Context

1.1 Complexity and Obscurity in Nature and in Models

This book deals with the use of data and models that can enhance understanding and contribute to prediction. These two goals are complementary. Both involve inference, and model analyses can take the form of predictive distributions. For environmental scientists, the challenge stems from the fact that natural and managed systems are high-dimensional, meaning that many interacting forces are at work (Levin 1998; Clark 2005; Clark and Gelfand 2006a). Much of nature is unmeasurable, unobservable, or both. Much cannot be manipulated. Faced with obscure, complex, and uncontrolled processes, environmental scientists have long recognized the need for abstraction (Schaffer 1981; Caswell 1988). Theoreticians and experimentalists attempt to extract the important relationships from nature so they can be studied in a controlled setting. Ecologists write models with only a few variables and parameters. They design experiments with only a few treatments.

The need to simplify on both the theoretical and experimental sides leaves a gap that can isolate those who analyze ecological models from those who collect and draw inference from data. This gap makes it difficult to test theory with data and to model data in appropriate ways (e.g., Oreskes et al. 1994). The goal of this book is to describe methods that can help to bridge the gap, starting from concepts that underlie traditional process and statistical models, and moving toward modern techniques that allow for deeper integration. This introductory chapter starts with some background and motivation.

1.1.1 Why Ecological Models Are Low-Dimensional

Attempts to abstract key features of a process are an important component of all scientific disciplines. From the conceptual (theory) side, this abstraction is accomplished with process models that contain few variables and parameters. High-dimensional process models are intractable; without simplified models, they cannot be analyzed to yield transparent relationships. Complex process

models are difficult to apply beyond the context for which they were developed. If we can abstract the important elements and develop a simple model thereof, analysis might allow us to understand how the process behaves and why.

The simplification needed to describe systems mathematically often requires assumptions that cannot apply to ecological data. Theorists may speak in terms of processes that apply everywhere, in a general sense, but nowhere in particular. In the light of the complexity mismatch between theory and the real world, it is not surprising that mathematical models are often viewed as irrelevant (e.g., Simberloff 1980). In over a decade of teaching mathematical models I do not need one hand to count the number of times that a basic model described in ecological textbooks has been directly applied to a student data set. When it comes to models, irrelevance can be the price of tractability.

1.1.2 Why Statistical Inference Is Low-Dimensional

Traditionally, the statistical analysis needed for inference and prediction was possible only for data collected under a rigid design. Here again, simplification is achieved through use of model systems. Statistical models simplify with assumptions such as each observation is independent of all others; uncertainty is typically allowed only for a response variable; the variables must be observable, in the sense that we can assume that values assigned to predictor variables represent truth. To meet these assumptions, experiments rely on strict design.

By contrast, ecological data are typically complex and interrelated in space and time (Ellner and Turchin 1995; Scheffer and Carpenter 2003). Broad-scale and long-term interactions with superimposed high frequency and difficult-to-measure fluctuations are pervasive. Only small parts of a process readily submit to experimental control and typically only in terms of highly abstracted experimental designs (Carpenter 1996; Brown 1995; Skelly 2002; Hastings 2004; Clark and LaDeau 2005). The small amount of variance in data that is explained by ecological models reflects the high dimensionality of nature, the many interacting processes that affect observations. The fact that simple models typically account for a small fraction of the total variance, even in experimentally controlled settings, leads to obvious questions: Are the important factors included in the analysis? Do the assumptions of independence and uncertainty confined to the response variable impact inference? Can the experimental results be extrapolated to nature?

Because many important processes cannot be studied in controlled experiments, there is a tendency to overlook model assumptions and apply statistical models where assumptions are violated. Spatio-temporal aspects of data and most sources of uncertainty tend to be ignored. Whether nature is abstracted to the point where simple model assumptions can be satisfied or assumptions are violated in the analysis of natural settings, the barriers between data and theory can be large.

The contrast between the high dimensionality of natural systems versus the simplicity that is manageable in experiments and models explains part of the historic gulf between ecological models and data. Throughout, I suggest that simple process models can be powerful. A goal of this book is to demonstrate

that this gulf can often be bridged, but only if process models can remain simple enough to be tractable, and data/parameter models can be sufficiently sophisticated to allow for the high dimensionality of nature. Rather than design away natural influences, I emphasize inference based on relevant scales and settings; the focus is on bringing models to nature, rather than nature to models.

1.2 Making the Connections: Data, Inference, and Decision

The challenges of inference, prediction, and decision making faced by ecologists are shared by many disciplines. I use two examples to highlight some of the specific issues. To emphasize the generality of both the challenges and the emerging tools needed to address them, I draw the first example from medicine. The second example comes from community ecology.

1.2.1 Example: Soft Data, Hard Decisions

Most information does not come from experiments designed by statisticians, and most decisions are subjective. Consider the example of a standard treatment of kidney stones with extracorporeal shockwave lithostripsy (Madigan and York 1995). The stone is located on a real-time ultrasound image by an operator who will focus shockwaves capable of disintegrating it. If the image is of high quality (I), there is an increased chance that it will be properly identified and disintegrated (D). If disintegrated, there is increased chance of clearance through the urinary tract (C). The decision framework involves interpretation of data with a flow outlined in Figure 1.1a. Note that the medical practitioner is concerned with the probability of successful clearance (C) given image quality (I), or $\Pr\{C|I\}$. Between these two events is the probability of disintegration (D). The decision is subjective, in the sense that, faced with the same evidence and choices, two practitioners could arrive at different decisions.

The basic elements of this problem are familiar to environmental scientists, managers, and policy makers. For instance, information is limited, but it can accumulate with time and experience. To many scientists, these data

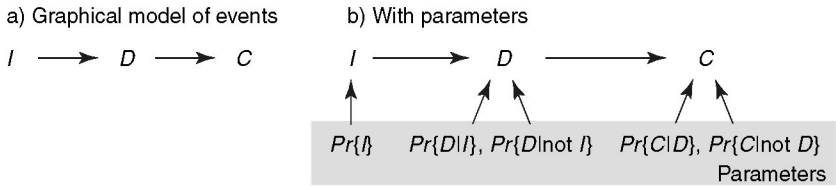

FIGURE 1.1. A graphical model of the kidney stone example (modified from Madigan and York 1995). In (*a*) is the graphical model of events that the image is of high quality (I), the stone is disintegrated (D), and the remnants are cleared through the urinary tract (C). To calculate $\Pr\{C|I\}$ we require five parameters (*b*). By making them stochastic (giving them priors), we can write simple rules for updating knowledge based on experience with new patients.

seem soft. The decision must be based on inadequate knowledge. We may not have the luxury of putting off a decision until hard data can be collected from, say, a series of controlled experiments. However, we would value a means for updating knowledge that can result in better decisions, that is, an adaptive management framework.

How can the practitioner use the model to learn from accumulated experience? The answer is, in large part, technical and a principal motivation of this book. But I provide a partial answer here. The parameters that influence decisions include the probability of C given that disintegration did or did not occur ($\Pr\{C|D\}$ and $\Pr\{C|\text{not } D\}$,[1] respectively), the probability of D given that image quality was good or bad ($\Pr\{D|I\}$ and $\Pr\{D|\text{not } I\}$, respectively), and the probability that the image was of good quality ($\Pr\{I\}$) (Figure 1.1b). The values of these five parameters determine probability of success. Clearly, the more we know about the values of these parameters, the more informed the decision. If we treat these parameters as being fixed values, there is no opportunity for learning. To allow for updating knowledge, the parameters are taken to be random. This random character of parameters allows for regular updating, with current understanding being further refined by each new experience (observation). The posterior knowledge taken from each patient becomes the prior for the next.

Whether the goal is increased understanding (as in, say, inference), prediction, or decision, the model graph provides a road map that facilitates not only modeling, but also computation. It emphasizes the importance of conditional probability, represented by arrows connecting events that are directly linked. In this particular instance, it describes a decision process that involves uncertainty that can be reduced though a prior-update-posterior cycle.

I use graphs like Figure 1.1 to represent models throughout this book. Early in the text, I use the convenient structure that involves data, process, and parameter submodels. This hierachical framework for submodels serves to decompose potentially complex problems into simple elements. By the end of this book I extend this framework to the more general notion of models as networks.

1.2.2 Example: Ecological Model Meets Data

A second example illustrates how the graphical framework of Figure 1.1 extends to ecological processes and brings in some of the challenges that confront ecologists attempting to integrate models and data. Ecological models predict that differences in how species respond to limiting resources can determine whether they can coexist (Levins 1979; Armstrong and McGehee 1980; Tilman 1982; Pacala and Silander 1990; Grover 1991; Pacala and Rees 1998; Murdoch et al. 1998). Is one species better able to exploit a limiting resource than another species? Does the advantage depend on how much resource is present? To evaluate the role of limiting resources, ecologists gather data describing growth responses at different resource levels.

---

[1] The notation | means "given that." Thus, $\Pr\{C|D\}$ means "the probability of event C given that event D has occurred."

Figure 1.2 shows data on seedling response (height growth) to a resource (light). The degree of scatter in Figure 1.2a and 1.2b is not unusual (Kobe 1999; Finzi and Canham 2000). Growth rate data are obtained from measurements of seedling height, together with estimates of light that penetrates the forest canopy. Full sunlight has a value of one, and complete darkness has a value of zero. Although the raw data do not show obvious differences between the two species, the fitted model says that the differences are highly significant, with 95 percent confidence intervals of each assigning near zero probability to the other (Figure 1.2c). It might seem paradoxical that these broadly overlapping data clouds are represented by significantly different models, yet this common situation is rarely mentioned. In fact, it is central to the interpretation of models.

Here is the standard analysis that led to this result. A model might include parameters that describe a minimum light requirement, a half-saturation constant, and a maximum rate, or asymptote. We could write the model as

$$y_{ijt} = \mu(x_{jt};\theta) + \varepsilon_{ijt}$$
$$\varepsilon_{ijt} \sim N(0,\sigma^2)$$

where x_{jt} and y_{ijt} are the predictor and response variables (light and growth) for the i^{th} individual at location j at time t, and μ is the saturating function of light availability x. θ represents the three parameters that determine μ, representing (1)

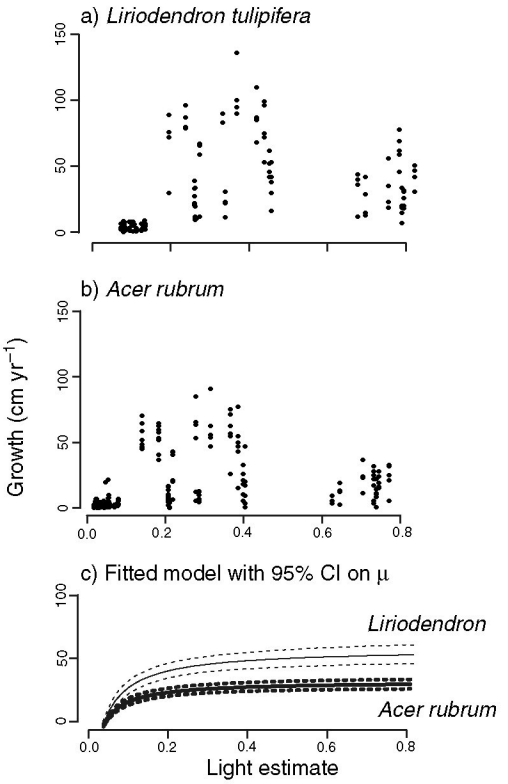

FIGURE 1.2. Growth rate responses to light availability. (*a*) and (*b*) show data obtained as measurements of seedling height growth and light availability. (*c*) shows a traditional model fitted to these data, the solid line being the estimate of μ. The dashed lines are 95 percent confidence intervals for μ. From Clark et al. (2003).

the minimum light requirement, (2) a half-saturation constant, and (3) the asymptotic growth rate. This saturating function does not fully describe the relationship between x and y—there is residual scatter. To accommodate this scatter, this function is embedded within a stochastic shell, which might allow for the fact that the model cannot be correct, and growth rates are not precisely known. Together, deterministic and stochastic elements are termed a *likelihood function*. This model says that y is assumed to have a normal distribution with a mean determined by the function μ and residual variance σ^2. This is the fitted model of Figure 1.2c.

The saturating function μ runs through the data cloud and is described by estimates of three parameters θ, which are represented by error distributions in Figure 1.3. The spread of these error distributions represents the level of confidence in parameter estimates, which depends, in turn, on the number of observations, the scatter, and the model. A 95 percent confidence interval can be viewed as the central 95 percent of the error distribution (although this is not its precise definition, as discussed in Chapter 5). The errors in these estimates are asymptotically zero in the sense that, as sample size increases, the confidence intervals (spread of error distributions) decrease. Hereafter, when I speak of asymptotics, I mean it in this sense of sample size. A predictive interval for the function μ is obtained by propagating the error in parameters θ to error in μ (dashed lines in Figure 1.3).

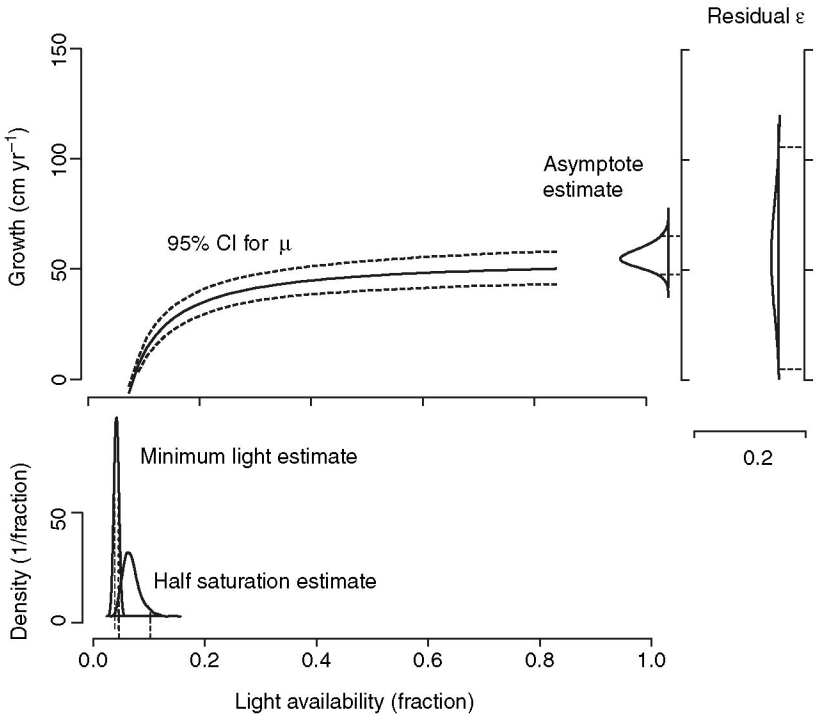

FIGURE 1.3. Elements of the traditional model used to analyze data in Figure 1.1. There are error distributions for three parameters and residual stochasticity in the error term ϵ (From Clark et al. 2003).

Obviously, the 95 percent confidence interval on $\mu(x)$ does not bound 95 percent of the data. The uncertainty in parameter estimates, which is used to construct the envelope for μ, is relatively small in comparison with variability in the data. The bulk of the variation is associated with the error term ε_{ijt}. This scatter is represented by a density shown on the far right-hand side of Figure 1.3. We could calculate a predictive interval for y that incorporates this scatter (it integrates ε) (Feiburg and Ellner 2003; Clark 2003), but that is rarely done. We would do this to predict as-yet unobserved growth response data y. But if this scatter is viewed as noise or error, we might not have much use for this predictive interval—it adds clutter, but not insight. Moreover, if we thought the scatter captured by ε_{ijt} was anything other than error, it might be difficult to justify this model in the first place.

So what is this leftover scatter? I began by saying that the scatter sopped up by ε_{ijt} might be associated with observation errors or model misspecification. But measurements of seedling height can be off by a centimeter. They are not off by a meter. So observation error is not the explanation for the broad scatter in Figure 1.2. If the deterministic part of the model μ is inadequate, we might increase its complexity in terms of a more flexible form or by including additional covariates that could explain the scatter. In either case, we require more parameters. In fact, ecologists have studied seedling growth many times, and measurements of more variables often do not explain much additional scatter. In other words, we often cannot account for this variability by increasing the complexity of the deterministic part of the model $\mu(x)$ (Clark et al. 2003a). If the scatter is not observation error, and we cannot accommodate it by incorporating more deterministic complexity, traditional methods do not leave many options.

In fact, there are many ways in which stochasticity might stand in for unobservable aspects of this relationship. For example, the light data x might be variable or imprecisely known. This brings an additional source of stochasticity and is sometimes termed an error-in-variables problem. I distinguish between the light seen by a plant at location j, x_j, and the observation of it using the notation $x_j^{(obs)}$. To allow that observations depend on the true light level, with uncertainty, I include a density for light observations, $x_j^{(obs)} \sim p(x_j,\phi)$, where ϕ represents any parameters that enter the model for observations. We now have the model

$$
\begin{aligned}
y_{ijt} &= \mu(x_{jt};\theta) + \varepsilon_{ijt} \\
x_{jt}^{obs} &\sim p(x_j,\phi) \qquad \text{uncertainty in } x \\
\varepsilon_{ijt} &\sim N(0,\sigma^2) \qquad \text{error in } y
\end{aligned}
$$

As in Section 1.2.1, we can represent this model with a graph. The basic model that ignores uncertainty in light is represented by Figure 1.4a. If implemented in a Bayesian framework (Chapter 4), I could refer to this as "simple Bayes." The stochastic element in this graph is represented by the connection between σ and y, indicating the error in y. In part b of this figure, there is an additional source of stochasticity associated with observations of x. The graph has increased in complexity to accommodate this additional relationship.

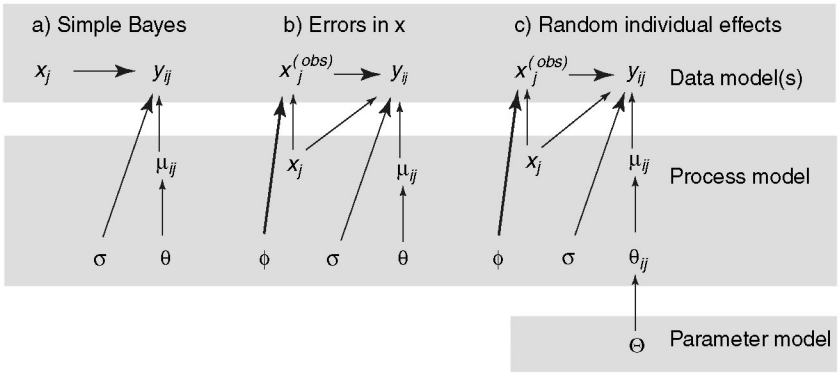

FIGURE 1.4. Four models that can be used to model growth response to light. Complexity increases from a simple model with error in y (*a*), to error in both variables (*b*), to variability among individuals (*c*). Modified from Clark (2005).

But there is still much more going on in Figure 1.2. We expect that individuals will have different responses to light, depending on many factors that cannot be measured. If individual responses result from unmeasured or unmeasurable factors, then we can include random effects in the model. In this case, there is a parameter vector that applies to each individual θ_{ij}. These individual parameters are drawn from a distribution having parameters Θ. This model is:

$$
\begin{aligned}
y_{ijt} &= \mu(x_{jt};\theta_{ij}) + \varepsilon_{ijt} \\
x_{jt}^{(obs)} &\sim p(x_j,\phi) \qquad \text{error in } x \\
\theta_{ij} &\sim p(\Theta) \qquad \text{random individual effects} \\
\varepsilon_{ijt} &\sim N(0,\sigma^2) \qquad \text{error in } y
\end{aligned}
$$

There are now $n+k$ new parameters that must be estimated, one θ_{ij} for each individual, and k parameters in θ, describing population heterogeneity.

The model is getting complex, and the list of potentially important influences is still not exhaustive. For example, we might include random effects associated with location, fixed or random effects for years, and there might be autocorrelation between years or with distance. We might have additional sources of information that are not yet accommodated by the model. Already the model is beginning to look like a network of elements, some of which are known and some unknown. This network perspective is readily depicted with graphs. The modules labeled as "Data," "Process," and "Parameter" on the right-hand side of Figure 1.4 help organize relationships.

With these basic principles in mind, I return to the traditional analysis, which superficially appears to have clarified a relationship involving broad scatter with a tight relationship between resource and growth (Figure 1.2). So where did the scatter go? And did we lose anything important? Using models available in standard software, there is a deterministic function $\mu(x)$ buried

in noise. The function $\mu(x)$ is like the mean response, and it is surrounded by error that is independent of $\mu(x)$. We now have the answer to the first question: the scatter was sopped up by a stochastic shell.

The second question is more difficult to answer. The insights used to extend the simple model came from recognition that observations of height growth are imprecise, light observations or treatments are variable and imprecisely known, individuals have different genotypes, and their responses to light depend on other factors that vary in time and space. Some of these factors can vary on such fine spatial and temporal scales that we could not hope to measure them. Because there are response times associated with many variables, we would have a hard time even deciding on a scale to measure them. If all of these factors contribute to the scatter, then the problem is high-dimensional.

How do the assumptions of the statistical model used to fit Figure 1.2c compare with the factors we identified on the basis of ecological insight? One could argue that at least one of these factors, observation error, is consistent with the statistical model (the deterministic-response-in-a-stochastic-shell approach). This would be the case if there were a deterministic response that applies to all individuals, and it was obscured by errors that result from sampling. The two might be independent if meter stick error does not depend on tree height. But the data suggest that this assumption does not apply here. The measurements could be off by perhaps a centimeter, but not by a meter.

Once we move beyond simple observation error, the deterministic-response-in-a-stochastic-shell model becomes less plausible. The other factors are not well described by this model, because their effects depend on the function $\mu(x)$ itself. Of course, we could make the stochasticity depend on $\mu(x)$ in a specific way by, say, modeling log values. Where such specific transformations can provide a quick fix, use them. For a number of reasons, this quick fix will not suffice here. If the estimates of the light level at which a plant grows are only approximate, then x is stochastic, and, in turn, $\mu(x)$ is stochastic; it is a function of stochastic x. If individuals have different responses to light due to genotype or factors that vary, then there is a function $\mu(x)$ for each individual i.

Is this simply nit-picking? How much of a difference can it make? Would we not fit roughly the same model, regardless of seemingly arcane decisions about what causes the scatter? Can these details affect the inference on the process of interest, example, competition and coexistence? The answer is, "it depends." We will confront this issue repeatedly. For now simply consider a few more points. First, the range of variability in growth described by the fitted function $\mu(x)$ in Figure 1.2c is a small fraction of the total (Figure 1.2a, b). If pressed to identify the dominant term in our model, we would have to say that it is the part we have relegated to the noise or error term, not to the signal $\mu(x)$.

The first point leads to a second. Inference is based on a model that differs substantially from our view of the relationship. If the scatter were modest we might ignore these factors, but how do we justify it here? If we are satisfied with sweeping under the rug the dominant part of the relationship, is there any

point to statistics? Do we have any more confidence in a predictive interval constructed from the traditional model than we would in a line drawn through the scatter by eye? It is worth mentioning that the prevailing confidence in ecological predictions is not great, in part because it is not clear how they relate to underlying processes and to the data used to fit them. The traditional view that the scatter is error underlies the interpretation that one species grows faster at all light levels (Figure 1.5a). In light of the assumptions, can we be sure that tulip poplar outcompetes red maple?

If we trust the data and we trust the theory (e.g., growth is a saturating function of light), then the statistical analysis is the weak link. Off-the-shelf software does not provide flexible options for most ecological data sets. In other words, it may be a bad idea to let default assumptions from canned software determine everything beyond the basic process model.

As preface to techniques covered in this book, consider what happens if we allow for the sources of variability that are consistent with ecological insight (uncertainties in observations, light, and the growth response $\mu(x)$). An analysis that admits these considerations suggests broad overlap (Figure 1.5b). It is true that we could also construct broadly overlapping prediction intervals for the classical approach in Figure 1.5a. This interval could be constructed by putting back the scatter (the stochastic shell) that we threw away to produce Figure 1.5a. I discuss this in later sections of the book. This is generally not done in ecological analyses, because modeling begins from the premise that everything other than $\mu(x)$ constitutes error and is independent of $\mu(x)$. Moreover, the predictive interval obtained by mixing in ε_{ijt} is not the one we obtain under the assumption that many factors contribute stochasticity, and many of them come through $\mu(x)$ itself.

Throughout this book I focus on the weak interface, the critical connection that is needed to evaluate relationships and to predict. Modern approaches use structure to allow for complexity and stochastic elements to stand in for uncertainty. Getting this connection right helps make theory relevant, and it allows us to exploit data fully.

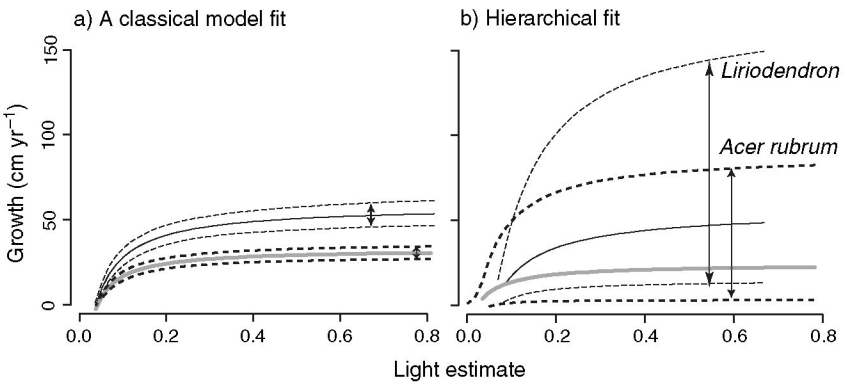

FIGURE 1.5. Comparison of the predictive interval on μ from a classical analysis, as has been standard practice in ecological analyses, with the predictive interval on y from the model in Figure 1.4c.

1.3 Two Elements of Models: Known and Unknown

I emphasize two elements of models, including variables and relationships that are known, either because they can be seen or we can assume them to be true, and those that are unknown, because they are obscure. The former constitutes much of traditional modeling in ecology. Unknowns are treated stochastically. Section 1.3.1 gives a brief overview of process models, followed in Section 1.3.2 by some aspects of unknown model elements, which are taken up in stochastic components.

1.3.1 A Point of Departure: Ecological Process Models

Environmental scientists do not have the luxury of continuing to limit attention to simple models. Low-dimensional models and designed experiments will continue to contribute insight, but the growing demand for relevant inference and prediction calls for the added capacity to address complex interactions. Modeling often begins with a process component to describe how things work. We may seek to determine whether or not relationships exist (e.g., hypothesis tests), to quantify relationships, and to make predictions. Where possible, the modeling strategy will focus on simple process models and, in many cases, allow for complexity on the stochastic side. Still, many processes have multiple interacting elements. Here I mention some of the types of environmental questions that entail integration of data and models, some of which are necessarily high-dimensional. Not all of these examples are included in this book, but they are all amenable to the approaches considered here.

Populations—At a time of intense interest in protecting rare species, the awareness that many nonendangered species undergo extreme fluctuations in density is of great interest. A broad range of questions and processes is addressed using population models and data. What places a population at risk? Will changes in the demography or age structure of a population have consequences in the future? What are the ecological implications of this structure? What kinds of constraints and trade-offs gave rise to observed life history schedules?

Ecologists use models to explore relationships involving individual behavior and dynamics of populations. For example, reproductive episodes can result in oscillations if there are strong feedback effects of density on population growth rate. Although evidence for chaos in population dynamics is weak (Hassell et al. 1976; Ellner and Turchin 1995), feedback involving resources or natural enemies can result in complex dynamics (Elton 1924; Saucy 1994; Ostfeld and Jones 1996; Earn et al. 2000). Oscillations can bring a population seemingly close to extinction, only to be followed by rebound to high abundances. Such fluctuations can occur repeatedly. They can be periodic, and, thus, are potentially predictable. Important strides in the 1970s (May 1974) initiated fertile research that has expanded to include feedbacks and environmental variation (e.g., Bjørnstad et al. 2001). Age represents

one of several kinds of structure in populations. It can have several classes of effects, depending on its contribution to time delays and response to perturbations (Lande et al. 2002; Coulson et al. 2001). The inherent spatial variability in nature (e.g., soils and topography) can explain much pattern in populations, but spatial pattern can also result from processes endogenous to populations themselves. Spatial coupling can result from environmental variation (the Moran effect) (Grenfell et al. 2000), dispersal (Kelly and Sork 2002), or other biotic interactions (Ostfeld and Jones 1996).

The challenges faced by scientists and managers come from the fact that populations are highly variable and subject to many influences, and demographic processes are hard to see. The basic process models used to infer population dynamics are reviewed in a number of recent theoretical ecology texts (Chapter 2). Spatio-temporal variability in demographic rates can complicate models (Wikle 2003a; Clark et al. 2003a) and require many years of data to estimate parameters (Clark et al. 1999b; Coulson et al. 2001). Many of the variables that ecologists would like to infer or predict, such as demographic rates for endangered species and extinction time, occur at low density. Because different processes may dominate at low density, including Allee effects, we cannot simply extrapolate parameter estimates obtained from populations studied at higher densities. Spatio-temporal processes are inherently high-dimensional, in the sense that variables not only interact with one another, but they do so in ways that depend on when and where they occur (Legendre 1993a; Ranta et al. 1995). Natural ecosystems are characterized by both continuous and discrete variation in space (Hanski 1998). Parameter uncertainty may have a large impact on predictions of extinction risk (Ludwig 1999; Ellner and Feiberg 2003). Ecologists have long struggled to characterize land cover in ways that are tractable, yet relevant for population dynamics (Lande 1988; Franklin et al. 2000).

Ecological Communities—Models are used to understand trophic interactions among species (Hutchinson 1959; MacArthur 1972; Tilman 1982, 1988). For example, why are there so many species? Why are some abundant and others rare? Trophic interactions and environmental controls contribute to these patterns. Interactions among species on the same and on different trophic levels must somehow explain much of the pattern (or lack thereof) observed in nature. Why don't the best competitors for the few resources that appear to limit growth drive all others to extinction (Tilman 1988)? How do species interact by way of shared natural enemies (Holt 1977) or other types of indirect effects (Wootton 1993)?

Early ecological models were deterministic, but incorporation of temporal stochasticity has become increasingly popular. Ecologists early suspected that variability in time could have some obvious as well as some mysterious implications for fundamental ecological processes (e.g., Elton 1924; Hutchinson 1959). Fluctuations in the environment coupled with temporal integration by organisms (e.g., long life, dormancy) can contribute to coexistence of species, and it may have evolutionary consequences (Chesson 2000). I consider different approaches for accommodating temporal heterogeneity in Chapter 9.

Spatial relationships complicate modeling, including movement and re-source patchiness. Recent theoretical models have increasingly focused on how predictions of spatial models differ from those of nonspatial ones (e.g., Lewis 1997; Neuhauser 2001; Bolker et al. 2003). A rich literature is developing that emphasizes problems related to spatial covariance and scale (Levin 1992). Theory has facilitated understanding of how aggregation of individuals in space affects interactions such as competition and predation. New statistical approaches allow us to consider processes where spatial relationships are not spatially coherent and change over time (Chapter 10).

Recent models have contributed to our understanding of how disease can affect population dynamics. For example, models have proved invaluable for exploring the spread of AIDS (Anderson and May 1991), foot-and-mouth (Keeling et al. 2001), and the temporal trends in measles (Bjørnstad et al. 2002) and whooping cough (Rohani et al. 1999). Examples of spatio-temporal analyses based on epidemiological data arise in Chapters 9 and 10.

Ecosystem Function—Biogeochemistry involves models and data at a range of spatial and temporal scales. While the 7 percent perturbation of carbon exchange between plants and the atmosphere has been enough to awaken global concerns of climate change, it is one of many interacting cycles that are absorbing the dramatic alterations by humans (Schlesinger 2004). Freshwater shortages, eutrophication from mobilized reactive N and P, acidified precipitation from mobilized S, and pollution of rivers and coastal oceans are just a few of the now-recognized transformations in the chemical environmental upon which life depends. Fertilizer applied in the Upper Midwest impacts fisheries in the Gulf of Mexico (Galloway 2004). Growing awareness that important human perturbations transcend the periodic table (e.g., recent concerns for Pb, Fe, Cl, and B, to name a few) underscores the need to understand linkages. Biology is central to such linkages. Stochiometric relationships demanded by organisms mean that change in the supply of one constituent can cascade through many others (Elser and Sterner 2002). Stochiometric relationships put biology in the driver's seat, as a regulator and a place where nonlinear feedbacks can reside, while simultaneously lending a degree of predictability to element interactions.

With a pending global energy crisis and its broad impact through climate, carbon will keep the attention of biogeochemists for some time (Jackson and Schlesinger 2004). Environmental scientists must anticipate not only the future impact of continuing human perturbations, but they also must weigh in on potential engineering fixes, such as the large-scale N fertilization of soils and Fe fertilization of oceans to stimulate uptake of atmospheric CO_2.

The modeling challenges are great, involving physical processes in soils, waters, and the atmosphere at a range of scales. I mention several continental-scale analyses involving air quality in Chapter 10.

Biodiversity Feedbacks on Ecosystems—Through integration of data and models, the melding of the classically distinct disciplines of community and eco-system ecology has shown that nutrient addition can lead to reduced diversity (Wedin and Tilman 1996), while diversity can play an important role

in nutrient supply (Vitousek et al. 1996) and retention (Tilman et al. 1996). Reorganization of food webs comes with changes in nutrient loading that propagate through primary producers to herbivores (Rosenzweig 1971; Carpenter et al. 2001). Nonindigenous species not only respond to land use changes, native biodiversity, and climate change, but they can also change native ecosystems, altering fire regimes (D'Antonio et al. 2000), promoting spread of infectious disease (some of which are themselves caused by nonindigenous pathogens) (Daszak et al. 2000), and affecting nutrient cycling (Vitousek et al. 1996). How does the loss or addition of species affect ecosystem functioning? Is there a minimal number of species needed to maintain ecosystem function in a given setting? Is redundancy needed as a buffer against change (e.g., Loreau et al. 2001)? Does variability at the level of individual populations propagate broadly? Or are species with similar or complementary function and niche requirements roughly equivalent (Doak et al. 1998; Reich et al. 2003)?

The many scale-dependent issues that arise in the context of ecosystem function can be modeled using the graphical framework outlined in Section 1.2. This framework is applied throughout this book.

Human Dimensions—Models and data are used to provide guidance on potential impacts of climate and land-cover change. The pace of contemporary climate change threatens reorganization of food webs, as species outrun their natural enemies and hosts, forming new networks of interaction, while others are left behind. Anthropogenic climate change has already affected many populations, but to what degree, and how has it propagated through communities? Lags, transient effects, and poorly understood changes in phenology are already combining to produce unanticipated effects on migratory birds, pollinators, and plant reproduction (Parmesan and Yohe 2003). Migration potential of plants became a research priority when general circulation models (GCMs) of the atmosphere suggested that contemporary migration rates would need to exceed apparent dispersal capacities of many species (Pitelka et al. 1997; Clark et al. 1998, 2003b). A growing number of examples from contemporary invasions demonstrate the extensive impacts that can follow from a small change in species richness (Mack et al. 2000; Callaway and Manon 2006). Ecologists are increasingly applying models to help understand which species, communities, ecosystems, and habitats are susceptible to climate change impacts.

Models and data are central to CO_2 fertilization study. Where can we expect CO_2-enhanced growth responses, for which species, and how will changing competitive relationships affect diversity (Hattenschwiller and Korner 1996; Mohan et al. 2005)? How do climate and CO_2-induced shifts in the length and timing of the growing season affect species distributions and interactions, and ultimately ecosystem-level processes (Farnsworth et al. 1996; DeLucia et al. 1999; Parmisan and Yohe 2003; Root et al. 2003; Ibáñez et al. 2006)? How do the combined effects of climate, CO_2, and habitat fragmentation affect migration potential?

These examples, drawn from a range of ecological concerns, have been traditionally modeled deterministically, in the sense that unknowns are ignored

or treated in unrealistic ways. There is a growing number of important exceptions that constitute many of the examples used in this book. Because process models do not include everything that goes on in nature, we need to consider what to do with the things that are left out. As suggested in Section 1.2, these leftovers are treated stochastically.

1.3.2 Stochasticity and a Structure for Complexity

The foregoing processes can be complex and obscure, and involve the types of unknowns described in Section 1.2.2. As discussed for that simple example, statistical inference is used to integrate data with the phenomena that are formulated as process models. That integration involves stochasticity. I began this chapter by saying that a statistical model may be as simple as a process model wrapped in stochasticity, a place to park the scatter. In fact, stochasticity is not confined to places where data enter the model. In Section 1.2.2 it stands in for unknowns, such as differences among individuals that cannot be related to specific variables or processes that are only partially understood. As background and motivation for the methods that follow, I consider the concept of stochasticity a bit further.

In this book, I use the term *stochasticity* to refer to things that are unknown. Although we could speak of a stochastic process, it can be most productive to think of processes themselves as deterministic. *Models* are stochastic, because we cannot measure or specify everything that is relevant. We choose what to treat deterministically and what must be treated stochastically. This pragmatic position is encountered in traditional mathematical treatments (e.g., Taylor and Karlin 1994) and in computer sciences applications such as machine learning (Mitchell 1997; Pearl 2002), and it is unavoidable in modern statistics (e.g., Dawid 2004). Nonetheless, this view of stochasticity may appear to conflict with terminology in the ecological literature, in that it does not suggest that processes are themselves stochastic. When pressed, those who argue that, say, population dynamics are inherently stochastic, may appeal to Heisenberg's uncertainty principle, which says that we cannot simultaneously know a particle's position and momentum. The uncertainty it describes has no demonstrated (or hypothesized) relevance at the level of observable phenomena. There is no obvious answer to the question of whether processes are really stochastic, but there is a practical approach—deterministic relationships are used for relationships that are known; stochasticity stands in for the unknowns. For example, Palmer et al. (2005) discuss the classic Lorenz (1963) attractor, a type of chaotic behavior, produced by a set of three coupled differential equations. Standard model selection arguments (Chapter 6) might tell us to truncate the model to two equations, because the third equation accounts for only 4 percent of the total variance. It is well known that this truncation completely changes the behaviour—it is not even chaotic. So what is the statistician to do with so little information on one-third of the state variables? One answer is a stochastic autoregressive term, which recovers the chaotic behavior, including the Lorenz attractor with the two-equation truncated version. In other words, judicious application of stochasticity stands in for

additional complexity, which may be unknown, unobserved, or both. For an ecological example, births and deaths are modeled as demographic stochasticity in models that treat each individual as a black box; we do not care to model the physiology of individual bodies. But a physiologist might do precisely that and would not view birth and death as stochastic, but rather as explicit processes about which much is known. We could have a philosophical debate about this; thus, I simply adopt this perspective on pragmatic grounds, several of which are outlined here:

1. It helps clarify the role of stochasticity in models—it stands in for things that are unknown.
2. It helps clarify the trade-off between determinism and stochasticity in models. If we could (or desired to) know everything, we would have a high-dimensional, deterministic model. If little is known, we may require a sophisticated treatment of stochasticity. Or we might choose a stochastic representation simply to reduce the complexity of a model.
3. It emphasizes that the terms *noise* and *error* are not always the best way to think about the role of stochasticity. In many cases, variability emanates from many unidentified sources, including from the process itself and from the context of the process. Sources of variability may be unknown, yet still demand careful treatment.

For pragmatic reasons I take the view that there are real processes behind the widely ranging growth rates in Figure 1.2. Stochastic representation is used when we do not know the underlying processes, they cannot be measured, or we do not know how to describe them in a deterministic model. The many components may operate on different scales. Simply collecting them all in one stochastic shell seems to agree with the concept of noise, but the processes they represent should impact the model in diverse ways. Figure 1.4c showed three different ways in which stochasticity stands in for uncertainty in x, in y, and in θ. This example illustrates the pragmatic treatment of stochasticity that is essential to modern data modeling. I began with a deterministic process model that might be motivated by previous understanding (saturating light-growth response). I then identified aspects of the problem that are unknown, including the resource, the differences among individuals, and the process model itself. Stochasticity is introduced for each of these unknowns.

Modern statistical methods provide a framework for complexity that allows for the comprehensive treatment of unknowns that contribute to Figure 1.2d. Assuming the wrong structure for stochastic elements might be just as naïve as an inappropriate function for the deterministic process. The basic approach involves a decomposition of complex problems into those that control the process of interest, those that generate data, and the underlying parameters. This structure admits influences that may impinge in diverse ways, without necessarily requiring that they can be observed or even identified.

The basic structure I follow throughout is outlined in Figure 1.6. This particular decomposition comes from Berliner (1996) and has been used in several recent ecological examples (e.g., Wikle 2003a; Clark et al. 2003b, 2004). Like

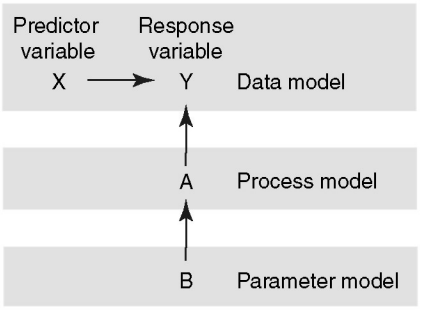

FIGURE 1.6. A graphical structure for an inferential model. Arrows indicate direct relationships that are specified explicitly as conditional distributions.

Figures 1.1 and 1.4, it is a graphical representation that provides a road map for the model. Attention may initially focus on the process level, where we allow for stochasticity in terms of process error or model misspecification. The connection to data accommodates sampling and the stochasticity that results from observation errors, sampling, missing data, and so forth. The parameter model can accommodate structure or relationships among sample units that might result from location, aggregation of individuals or traits, and so on.

By decomposing the problem, we can work directly with a process model that may be simple (and general), while allowing for the complexity that applies to a specific system (parameter models can absorb structure among variables) and data set or experimental design (data model) at other levels. This structure also admits data assimilation, meta-analysis, prediction, and decision analysis. To emphasize this connection between models and data it can be useful to think of inference and forward simulation as two sides of the same coin, one being the inverse of the other.

1.4 Learning with Models: Hypotheses and Quantification

Broadly speaking, the principal products of an analysis consist of (1) quantifying relationships and states (estimates and confidence intervals), (2) testing hypotheses (including model selection), (3) prediction, and (4) decision. Of these four activities, ecological analysis has tended to focus on hypothesis testing. I will not dwell on philosophy, but my inclination to soft-peddle hypothesis testing throughout this book needs some explanation. I discuss hypothesis testing many times, but it will often not be a prime motivation.

Before discussing why I place less emphasis on hypothesis testing than is customary, it is important to first say that hypotheses are critical. Any hope of success requires experiments and observations that are motivated by a clear set of questions with careful thought as to how the range of potential outcomes can lead to progress in understanding. A focus on multiple hypotheses, as opposed to a null hypothesis (e.g., Hilborn and Mangel 1997), can sharpen the approach. The lack of consensus on how to test hypotheses and to interpret them plays into a broader misunderstanding of what models are and how they can be effectively exploited. I summarize some of the issues that arise

with hypothesis testing followed by general considerations concerning the inevitable subjectivity of statistics.

1.4.1 Hypothesis Tests

Consider the (null) hypothesis H_0: $\theta = q$. We collect data and obtain an estimate of θ. Can we make a probability statement about the result? Something about our confidence that reality is different from or the same as q?

Three approaches to hypothesis testing are reviewed by Berger (2003). Fisher significance testing involves a test statistic S that increases in value the farther the estimate is from the hypothesized value q. The larger the value of S, the greater the evidence against the null hypothesis. A P value is associated with S, but it is not the error probability (i.e., the probability of being wrong). It is calculated as the area in the tail of the distribution of S, the tail being everything more extreme than S. If treated as an error probability (as is often done), the evidence against the null hypothesis is overstated. So P values are not error probabilities (they violate the frequentist principle—see below); they are based on values of S more extreme than actually observed.

Neyman-Pearson hypothesis testing has an explicit alternative hypothesis and appeals to the frequentist principle. This principle can be stated in different ways. Berger's (2003) practical version says that if we could repeat the experiment many times, then the calculated Type I and Type II error probabilities (falsely reject the null, falsely accept the null) would agree with the fraction of Type I and Type II errors from the repeated experiments. Whereas the Fisher alternative is vague (H_0: $\theta = q$, H_1: $\theta \neq q$), the Neyman-Pearson alternative is explicit (H_0: $\theta = q_0$, H_1: $\theta = q_1$). As with a Fisher hypothesis test, there is a test statistic S that increases in value with evidence against the null. The method requires a predefined critical value of S for accepting versus rejecting H_0. The outcome does not discriminate between outcomes (values of S) that may be barely significant versus very significant, because the critical value (e.g., $\alpha = 0.05$) only has meaning if it is designated in advance. Unlike the Fisher method, we do not conclude that S had a P value of, say, 0.03, unless this happened to be the preselected critical value. S is either less than the critical value or not, and the conclusion is the same regardless of whether S exceeds the critical value by a little or by a lot. In other words, data sets that lend very different weights to the null hypothesis in the Fisher sense might lead to the same inference in a Neyman-Pearson sense.

Jeffreys hypothesis testing also involves an explicit alternative hypothesis. The null hypothesis can be accepted if it results in a better fit (a greater likelihood) with a probability determined under the assumption that prior probabilities for null and alternative $= \frac{1}{2}$. One criticism of this approach has been the need to specify a prior probability (Fisher 1935).

The point of this summary is to emphasize that thoughtful experts do not agree on how to test a hypothesis. And the challenges mount as we move to high-dimensional hierarchical models, where the concepts of sample size and number of parameters do not have obvious interpretations (Gelfand and Ghosh 1998; Spiegelhalter et al. 2002). The desire to assign a probability statement to

a study is a laudable goal. It is often unrealistic in more than a highly qualified fashion. Ecologists can take sides in the debates and hope to maintain the "illusion of objectivity" (Berger and Berry 1988; Clark 2005). This may be the only option for those wishing to justify heavy reliance on hypothesis tests. At best, hypothesis tests are guidelines.

Even where we can agree on a hypothesis test, designs based on rejecting the null may not yield much. A study deemed unpublishable, because it fails to reject the null, may have been vacuous from the start. If showing an effect is viewed as the only informative outcome, there probably is no informative outcome. Studies founded on questions like "Does X affect Y?" (H_0: $\theta = q$) rarely provide much guidance, regardless of the P value. Such approaches often devolve to sample size (Spiegelhalter et al. 2003). Alternatively, if designed to parameterize a relationship, as opposed to rejecting a null hypothesis, that relationship may be the important product of the study.

1.4.2 The "Illusion of Objectivity"

The foregoing concerns bear on broader issues regarding what models represent, how they relate to theories and hypotheses, how they are used, and what can be inferred when they are combined with data. Models area not truth, just tools.

I will not advocate a distinction between scientific versus statistical models or between empirical versus mechanistic models. Such terms have had a place in the context of classical data modeling, because theory and data can be difficult to combine. For instance, it is common to fit data to a model like that shown in Figure 1.2, and then to use estimates that fit in a more complex model, for example, one that is dynamic with additional assumptions about other variables, perhaps even at different spatial and temporal scales. More desirable would be to fit the data directly to the model that will ultimately be used for prediction (Pacala et al. 1996). This direct connection is more desirable, because predictive intervals depend on the relationships that are lost when models are fitted piecemeal. The full model is harder to link directly with data, because it is complex.

By admitting complexity, modern methods allow for more direct connections between theory and data. The emerging techniques increasingly make distinctions between scientific and statistical models unnecessary. A motivation for setting such terms aside is the sense they can foster that there is a correct model, a best type of data, and an objective design; that is, there should be some formal and objective statistical test that will tell us what is best and correct. Models differ in complexity for many reasons, including not only the complexity of the process, but also how much we can know about it. Models are caricatures of reality, constructed to facilitate learning. In a world where data are accumulating at unprecedented rates, far more rapidly than we can assimilate them in models (e.g., remote sensing, climate variables, molecular data), we need ways to combine the information coming from multiple sources. Much of this book addresses formal structures for integrating multiple data sources within complementary models.

In most cases, there will be more than one way to model the same process. How well the model fits a data set is one of several considerations in deciding among candidate models. In several sections, I discuss model selection as a basis for helping to identify models that can be of most use. I follow the standard practice of parsimony: increasing complexity in a model requires justification. Still, I place less emphasis on model selection than is typically done within a classical approach focused on rejecting a null hypothesis. This diminished emphasis on model selection stems not only from the view that there is rarely a correct model, but also on several characteristics of ecological phenomena. Ecological processes are inherently spatio-temporal, and the best model can vary from place to place and change over time (e.g., West and Harrison 1997). The frequentist concepts based on the idealized notion of resampling an identical population often does not directly translate to environmental data sets. This is inevitable in high-dimensional systems that are subject to processes that operate at a range of scales. This view can shift the emphasis from that of identifying the correct model to one of identifying models that can be useful in different settings and the need to consider model uncertainty as integral to inference. In some cases it can motivate combining models or model averaging.

Finally, formal statistics should not trump all other considerations, especially when it comes to model selection. The fit to data is not the sole basis for model selection. General circulation models of the atmosphere (GCMs) were constructed to embrace certain physical relationships. Although weather and climate models can benefit from better integration with data (e.g., Bengtsson et al. 2003), a hypothesis-testing framework was not the basis for their formulation and does not play a role in continuing model development.

1.4.3 Quantifying Relationships

Unlike hypothesis tests, confidence envelopes are less controversial than many ecologists think. Confidence envelopes estimated for a given model by different methods and stemming from different perspectives often yield similar results. The processes under investigation are often known to exist. There may be no point in rejecting the existence of an effect. There may be value in quantifying it. For example, the effect of interest may contribute a relatively small fraction of the variance to a data set, because there are other large effects. If so, model selection (a hypothesis test) may reject the added complexity of a model containing the variable of interest. This should not necessarily deter us from examining those variables. Lange et al. (1992) found slight improvement in the fit of a change-point model of treatment effects on CD4 lymphocytes, a surrogate for AIDs infection, in patients with advanced HIV infection (models of this type are described in Chapter 9). The more complex model that included the variable of interest was still valuable, despite the fact that it did not explain nearly as much of the total variance as did other factors. Clark et al. (2004) found that temporal effects on tree fecundity overwhelmed individual effects, yet individual differences were

still large. The more complex model that contained individual effects allowed for quantification of individual effects, *in spite of* large stochasticity standing in for other effects. A posterior density centered near zero might suggest use of a simpler model, yet this parameter might be retained simply as insurance against effects on estimates of other parameters.

1.4.4 Learning from Models

Learning is progressive. It entails continual updating or assimilating information as it accumulates (Section 1.2.1). Models have a central role in this process. This process might be represented as

$$\text{updated knowledge} = f(\text{data, previous knowledge})$$

Bayesian methods do this formally as

$$posterior \propto likelihood \times prior$$

Learning can occur sequentially, with the posterior or predictive distribution that derives from information available before serving as a prior to be combined with data that comes available in the future (West and Harrison 1997). Whether or not we do this formally, the process of successively updating understanding as information accumulates is standard practice. The Bayesian framework for learning plays a large role in this book. The likelihood in this expression admits data. For processes that evolve over time, we might term this the *update/forecast cycle*, with the goal being to ingest new information continually.

1.5 Estimation versus Forward Simulation

Model analysis and prediction are ideally based on the same process model as that used for estimation (Figure 1.7). This connection is critical if we are to construct prediction envelopes that accurately reflect uncertainty in data and models. For a number of reasons, few ecological analyses proceed this way. Predictions are typically made from models that are not the same as those used to obtain parameter estimates. And ecological predictions typically come from process models only after discarding the stochasticity (Clark 2003). It is not necessarily wrong to analyze models that are not fitted to data. But arbitrary treatment of uncertainty can make predictions misleading or even meaningless. This book is partly motivated by the fact that there are fewer obstacles in the way of integrating theory and data than there were just a decade ago. New approaches provide a natural framework for data assimilation and for decision (Figure 1.7).

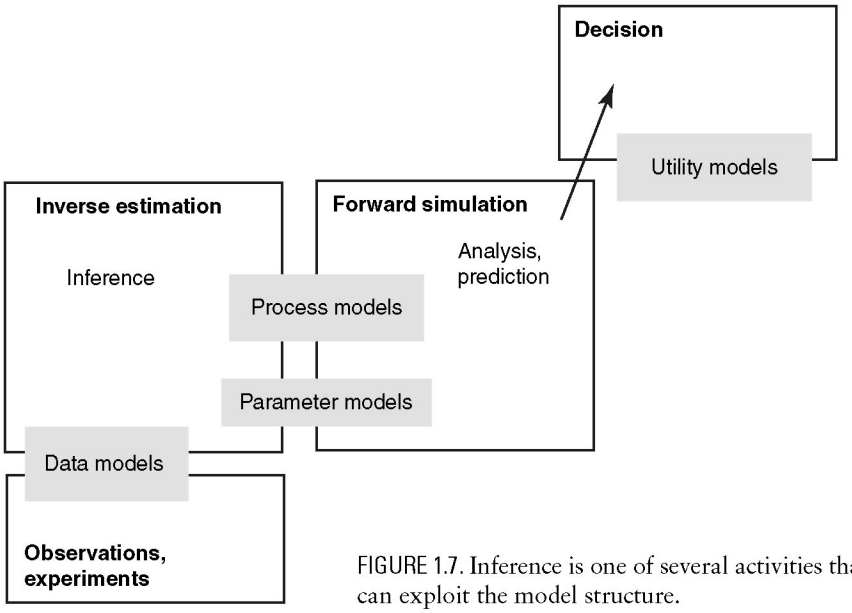

FIGURE 1.7. Inference is one of several activities that can exploit the model structure.

1.6 Statistical Pragmatism

No practicing ecologist is ignorant of the debates over frequentist and Bayesian approaches (e.g., Dixon and Ellison 1996). Already, philosophy has come up in reference to stochasticity (Section 1.3.2) and hypothesis testing (Section 1.4), but it does not play a large role in this book. In light of the philosophical emphasis of ecological writings about Bayes, I should provide some justification for my largely nonphilosophical treatment in this text. Here I summarize several issues that I have written about elsewhere (Clark 2005).

First, the points of controversy have traditionally revolved around *frequency* versus *subjective* concepts of probability (Chapters 4 and 5). The different concepts are especially important when one tries to make sense of classical hypothesis testing, but they also arise within the Bayesian community, in the context of how strictly prior specification must represent current understanding. Ecological writings on Bayes have focused on philosophical differences, advising that an analysis begins with one's philosophical stance, which then prescribes a classical versus a Bayesian approach. This prescription follows past debates in the statistical literature that have become less common and less polarized in recent years. I have expressed my view that the dissipation of such debates and the emergence of modern Bayes in applied fields have less to do with philosophy than pragmatism (Clark 2005). In fact, the philosophy remains important, being central to the challenge of doing objective science using the subjective tools of statistics (e.g., Berger and Berry 1988; Dawid 2004). The expansion of modern Bayes owes much to developing machinery.

Bayesians can now take on high-dimensional problems that were not accessible in the past.

In my view, the persistent focus on philosophy in ecological writings has become counterproductive (Clark 2005). It should not be news to any scientist that strong priors can affect estimates. Weak priors generally do not. The nonpractitioner must wonder at the standard examples provided in many ecological demonstrations of Bayes. Typically a simple Bayesian analysis requires far more work to arrive at essentially the same confidence envelope that could have been obtained with standard software in a classical framework. Given that most ecologists have limited philosophical baggage in this respect, why complicate the analysis? Moreover, the focus on philosophical differences can confuse the issues. For example, although Bayesians refer to parameters as "random," and frequentists do not, in both cases, parameters are fixed. Going further, both approaches view parameters as uncertain. Ecologists have long confused references to random parameters with the idea that the underlying value of a parameter fluctuates. Classical confidence intervals and Bayesian credible intervals both express uncertainty about the underlying true value, which is unknown.

I say more about the underlying assumptions of classical and Bayes models as we begin to implement them. For now, I simply forewarn that both are important in this book, with applications being pragmatic rather than philosophical. By pragmatic, I mean that I avoid unnecessary complexity. The initial models I discuss in Chapter 3 are classical. Bayesian methods require evaluation of integrals, analytically or through simulation. For simple problems this can mean that Bayes requires more effort. I also discuss Bayes for relatively simple problems, but it becomes the dominant approach in later chapters to allow for external sources of information and uncertainty. As models become more complex, the level of difficulty associated with classical and Bayes can reverse. Not only will the Bayesian framework bring flexibility to address complex problems (even where classical methods are still an option), Bayesian approaches can be *easier*, facilitated by new computational techniques.

A pragmatic approach need not be controversial. As suggested above, primary focus on quantification, rather than hypothesis testing, means that the products of a classical and Bayes analysis (e.g., confidence and credible intervals) may often be similar (e.g., Cousins 1995; Clark and Lavine 2001). Controversy can still arise over the role of priors in a Bayesian framework. I will say that informative priors are a good idea, when external information is available. Noninformative priors are useful when information is not available, or when we want to isolate the contribution of a particular data set. Those just testing the Bayesian waters may be motivated by machinery. Those who initially balk at informed priors may come to recognize each new data set as a way of updating what is already known. Priors can be an efficient way to introduce partially known relationships, in contrast to the traditional practice of assuming such relationships are completely known. I return to some of these issues in the final chapter.

Even as models become complex and (in this case) Bayesian, many of the tools will be familiar from classical statistics. Bayesian models will have embedded within them many traditional structures. The power of modern Bayes comes

from the capacity to integrate traditional techniques within high-dimensional models.

Although I embrace both frequentist and Bayesian approaches, there is much that I avoid. In both cases, there is an established framework that underpins the analysis. If I say that I have constructed a confidence envelope based on Fisher Information or a Bayesian posterior, you will have a good idea of what it means (Chapter 5). Although challenging problems can foster excursions into ad hockery (a much bigger temptation for the inflexible classical methods), the formalism of these two approaches is well established.

In this regard, the expanding ecological literature on fitting models with ad hoc approaches that are neither frequentist nor Bayesian is avoided in this book. Many such approaches are highly creative. But there can be numerous pitfalls in terms of inference, prediction, and communication to others. The goal of inference is a probability statement about unobservables—quantities we wish to estimate. A probability statement relies on a framework that is consistent throughout. Ad hockery tends to break down the connection between confidence envelopes and the data used to construct them. Moreover, ad hockery is often unnecessary. When things get complicated, we will go Bayesian. It admits complexity within a consistent, reproducible framework. Although there remains much to understand in the Bayesian world (e.g., selection among hierarchical models in Chapter 8 and complex spatio-temporal models in Chapter 10), there is rapid progress and it builds on a firm foundation. Because we have the Bayesian alternative, there are many recent papers on how to run lines through data points that will not be covered here.

Finally, although I cover basic concepts in classical statistics, I make no effort at comprehensive coverage of classical designs and hypothesis tests. Classical statistics may seem to burden the scientist with a different design and test for each new application. There is substantial jargon that continues to find its way into ecological statistics books, much of which is largely historical and rarely encountered in the modern statistics literature. I focus on general approaches and attempt to minimize jargon.

2 Model Elements: Application to Population Growth

In this chapter I introduce some of the basic types of process models used in ecology. Population growth provides an example, but the principal goal is to summarize basic features of models that are formulated to describe a process. These concepts include model state, linear versus nonlinear models (and how these terms are used in mathematics and in statistics), and structure in models.

Because the focus of this book is data-model connections, my exposition of standard ecological models differs, in some respects, from those contained in most texts. Population models are the basis for some excellent recent ecology texts, including Hastings (1997), Kot (2001), Caswell (2001), Morris and Doak (2003), and Lande et al. (2003). These references are highly recommended for ecologists who study populations, as they contain more analysis of models than is included here. I provide a condensed treatment of this material, but emphasize aspects of models that will be important as we begin to incorporate data in Part II. For ecologists who do not study population-level phenomena, I recommend this section for general background on model development.

2.1 A Model and Data Example

In the early part of this century Raymond Pearl (1925) used a simple sigmoid (logistic) curve, which he sometimes referred to as the *law of population growth*, to predict that the U.S. population would reach an upper limit of 197 million shortly after the year AD 2000 (Figure 2.1a) (Pearl 1925; Kingsland 1985). The basis for this prediction was U.S. Census Bureau data, from which he obtained decadal population statistics from AD 1790 to 1910, and the model in which he placed great faith—so much faith that he reported the inflection point of the curve to lie at April Fools Day, AD 1914. Note from Figure 2.1a that population sizes after 1910 are based on extrapolation.

The future has come to pass, and the modern census data do not look like the predictions of the early twentieth century (Figure 2.1b). History provides two lessons that this and subsequent chapters explore: (1) models can be importantly wrong (they may fit well for the wrong reasons), and (2) model evaluation requires estimates of uncertainty.

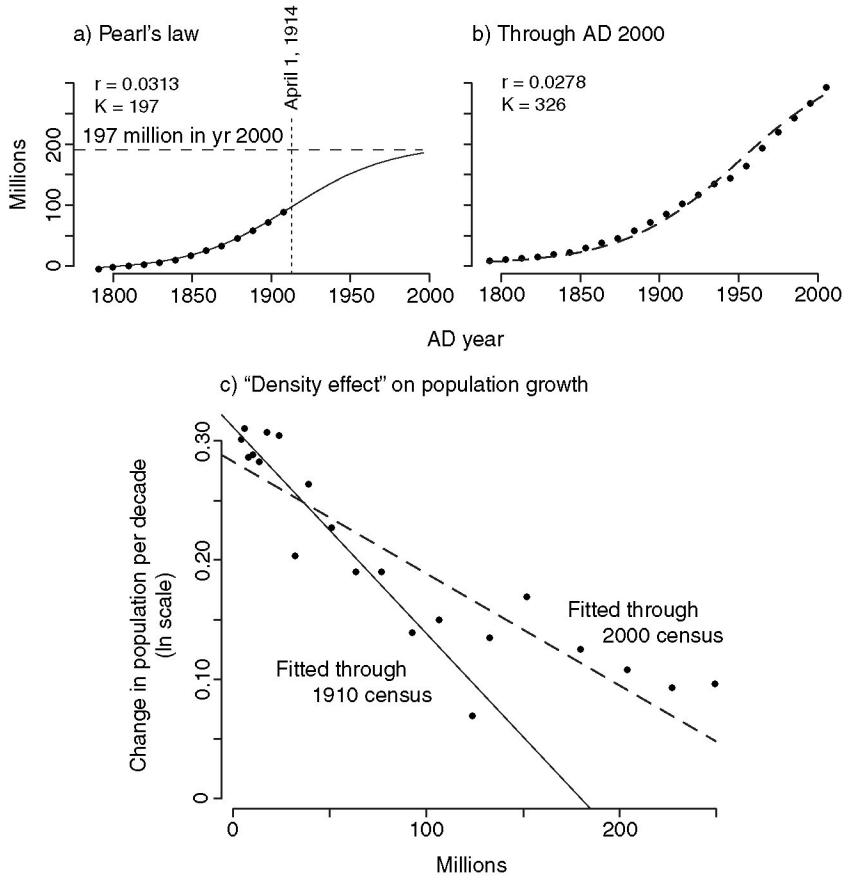

FIGURE 2.1. (a) Pearl's logistic curve fitted to U.S. Census Bureau data from 1790 through 1910. (b) A comparison with the longer series available in AD 2000. (c) The model assumption that population growth declines with density fitted to two subsets of the full census data. The parameters r and K describe per capita growth rate and the carrying capacity, respectively. Census data from www.census.gov/population/censusdata/table-16.pdf.

What did Pearl's model assume? One of the oldest ideas about population growth is that it cannot continue indefinitely (Darwin 1859; Lack 1954). The rate of growth will decline as population size increases, because growth will eventually become limited by resources, natural enemies, and so forth. If density exceeds some carrying capacity, growth rates can be negative, and density will either stabilize or decline.

Model construction begins with an assumption of how density affects growth rate. This relationship is complex and poorly understood. For instance, fertility might decline, mortality might increase, both in ways that vary in time and space. Lacking knowledge of these details, ecologists commonly use an assumption that the per capita rate of growth declines as a constant fraction of population size. The *per capita rate* of increase is the proportionate rate (Appendix B). As a rough approximation, this assumption of constant decline might be reasonable, regardless of whether density influences fertility or mortality,

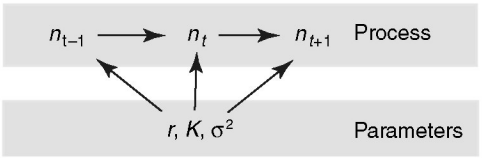

FIGURE 2.2. A graph of the standard logistic model. The process steps forward as governed by two process parameters (r and k) and process error, having variance σ^2.

because both are per capita rates. Let n represent density. This linear relationship is written as

$$\begin{matrix}\text{proportionate} \\ \text{change in } n\end{matrix} = \beta_0 + \beta_1 n$$

There is a potential rate of increase β_0 that applies when density is low (n is close to zero). The slope parameter β_1 is negative and describes the strength of population regulation. This relationship between population size and growth is shown in Figure 2.1c as estimated for examples in 2.1a and 2.1b. This is the logistic model of population growth that was used by Pearl. It is intended to represent feedback of density on population growth. This model could be viewed as having a process that is governed by parameters that must be estimated from data (Figure 2.2).[1] The parameters β_0 and β_1 can be expressed in terms of Pearl's r and K (see below).

The assumptions seem plausible, so why did the model mislead? The logistic model of population growth fits human census data remarkably well (Figure 2.1a). This close fit can inspire confidence that it represents the correct model for population growth. We might quarrel with Pearl's extrapolation from pre-1910 data on uncertainty grounds—had he attempted to determine the uncertainty that could have been quantified as part of the model fit, the subsequent outcome might have been within, say, a standard derivation or so of model predictions. We could then argue about whether the prediction was close enough to be useful (model adequacy). We could have this debate whether or not we agree that this is the most realistic model of population growth (model selection).

So there is a second issue that concerns whether the model captures the process in a reasonable way. Obviously, demographers applying the same model today could reach an equally confident, albeit substantially different, answer (Figure 2.1b). After all, this tight fit is the outcome that fosters confidence and justifies detailed model analysis and projection.

The close fits in Figure 2.1 belie inaccurate description of the process. For example, much of the U.S. population increase resulted from immigration, a process that affects model behavior in a different way than fertility of the resident population. In 1850 (the first year that the U.S. Census attempted to gather nativity data), nearly 10 percent of the censused population was foreign-born. By 1910 (Pearl's data set), foreign-born residents reached nearly 15 percent (www.census.gov/population/www/documentation/twps0029/ tab01.html). In this example, extrapolation was impressively inaccurate, in part, because

[1] This is not precisely the model fitted by pearl.

immigration might respond to extrinsic factors to a much greater degree than does fertility. Immigration came in waves, for reasons that had little to do with U.S. population size. This weak link between population size and immigration means that it is not well described as a per capita rate. Whereas no model would be precisely correct, this one has serious flaws.

Environmental scientists embrace a broad range of questions, consider many types of data, and employ a variety of models. To get started, I devote this chapter to the deterministic elements that are typically included in process models, such as this one for population growth. Some ecologists think of this as the theory or process part of the problem. I include only the most basic tools needed for derivation and analysis, which appear primarily in appendixes. These process models provide fodder for the data modeling that begins in Chapter 3.

2.2 Model State and Time

In this section I cover some background concepts, terms, and notation. First I consider the distinction between *linear* and *nonlinear* models, followed by state variables and structure.

2.2.1 When Is a Model Linear?

There is a simple way to determine whether a model is nonlinear: the second derivative of a linear model is equal to zero. Because this book deals with models that come from both ecological and statistical traditions, there are different terminologies that can create confusion; in this case, we must agree on which derivative. Statisticians apply this distinction with respect to the *parameters* of a model. Consider the equation

$$f(n) = \beta n^2$$

There is a state variable n and a parameter β. (I say more about state variables in a moment.) In a statistical context, this model is termed *linear*, because it is *linear* in β. In other words, $\frac{d^2 f}{d\beta^2} = 0$[2]. For a mathematician, the model $f(n) = \beta n^2$ might be termed *nonlinear*, because

$$\frac{d^2 f}{dn^2} \neq 0$$

In other words, mathematicians tend to think in terms of state variables.

Like mathematicians, ecologists tend to think in terms of state variables and often use the terms *density-independent* and *density-dependent*, respectively, for linear and nonlinear models in the state variable population density

[2] Differentiate once with respect to β to obtain $df/d\beta = n^2$, their differentiate again to obtain zero.

(e.g., May 1981). Some ecologists will find it confusing to see polynomials referred to as "linear models." Recall that statisticians come from a tradition of thinking about parameters. Of course, models can be nonlinear both in parameters and in state variables, such as $f(n) = \frac{n}{n + \beta}$ and $f(n) = 1 - e^{-\beta n}$. I retain this confusion of terms, because it is entrenched. Throughout, I define how the terms are used in specific cases. Except when used in a statistical context, I use the terms *linear* and *nonlinear* with respect to the state variable, such as concentration of a nutrient, population size or density, and so forth.

A second source of confusion comes from the fact that population ecologists often apply the terms *linear* and *nonlinear* to the effect of density on the per capita growth rate, $1/n\ dn/dt$. This terminology is fine, so long as it is qualified by context. However, a model that is linear in per capita growth rate is not a linear differential equation.

In human demography and population biology, linear models (in terms of the state variable density or abundance) are typically used to describe current population growth. Standard human population growth statistics, such as "country X is growing at a rate of Y percent per year," are examples. They are rarely used to predict long-term population trajectories (i.e., over generations), because they cannot be extrapolated with confidence. Many factors that might be ignored in the short term have long-term or cumulative effects. These regulatory factors can directly or indirectly involve density or age structure. Nonlinear models, when used to describe population growth, are termed density-dependent, because the per capita rate of growth depends on the state variable density. Such models describe the phenomenon of population regulation or the tendency for population density to approach some stationary distribution having a mean, a variance, and so forth (May 1973; Dennis and Taper 1994; Turchin 1995). Nonlinear models are also used when the rate of a reaction depends on the concentration of substrate.

2.2.2 Time as a Continuous Variable

Many ecological models are dynamic, with time as an explicit variable. Time can be treated as discrete or continuous. Continuous time models are used to describe processes where change occurs continuously and are written as rate equations. For change in density $n(t)$, we write a differential equation as dn/dt.

Differential equations can be linear or nonlinear. The equation

$$f(n) \equiv \frac{dn}{dt} = rn$$

is a linear differential equation, because the second derivative with respect to the state variable n is

$$\frac{d^2}{dn^2} f(n) = 0$$

This terminology may seem confusing at first, because the solution to this particular linear differential equation is exponential (nonlinear) in the continuous

variable *time*. In population biology, parameter r is sometimes termed the "intrinsic rate of increase." It is the instantaneous contribution of each individual to the population through its birth rate and mortality risk, which is easy to see if we express this rate on a per capita basis,

$$\frac{1}{n}\frac{dn}{dt} = \frac{d(\ln n)}{dt} = r$$

It has the solution

$$n(t + dt) = n(t)\, e^{r \cdot dt} \qquad 2.1$$

where dt is a time increment (Appendix B). I use the notation $n(t)$ to denote time as a continuous variable, to distinguish it from the notation for discrete time n_t in Section 2.2.3. Short-term rates of population change are often reported as annual rates of population growth and presented as percents. These statistics are calculated as 100 r. To determine r based on knowledge of $n(t)$ and $n(t + dt)$, take logs of equation 2.1

$$\ln n(t + dt) = \ln n(t) + r dt$$

and rearrange,

$$r = \frac{\ln n(t + dt) - \ln n(t)}{dt} \qquad 2.2$$

Population increase is described by $r > 0$ and vice versa. Figure 2.3 shows an application of this model, as it was used to describe the twentieth-century increase in the nesting population of black noddies (*Anous minutus*) on Heron Island, Great Barrier Reef (Ogden 1993). The rapid increase, estimated as $r = 0.081$ (8.1 percent per year), cannot continue indefinitely, yet it appears to fit the existing data well.

Nonlinear differential equations are commonly used in ecology. If the environment has a limited capacity to support new individuals, the per capita rate must decline as population density increases. Pearl used the simplest assumption of a straight line (left side of Figure 2.4). This equation could be written as

$$\frac{1}{n}\frac{dn}{dt} = \beta_0 + \beta_1 n \qquad 2.3$$

for two parameters, an intercept β_0 and a slope β_1. In this logistic model, the first parameter is positive so that growth rate is positive at low density. The second parameter is negative, reflecting the negative effect of density on growth rate (Figure 2.4a). Ecologists often write this as

$$\frac{1}{n}\frac{dn}{dt} = r\left(1 - \frac{n}{K}\right)$$

with parameters β_0 being equivalent to r and β_1 equivalent to $-r/K$. This is the "continuous logistic" or (formerly) the "Law of Population Growth" of

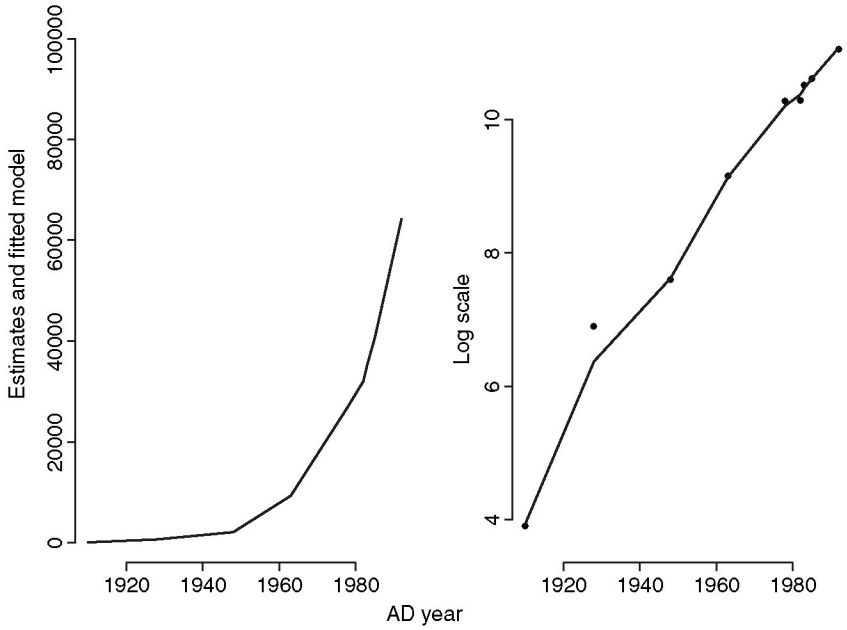

FIGURE 2.3. The black noddy population increase on Heron Island. The exponential model was fitted to this data set, with allowance for process error and observation errors (Chapter 9).

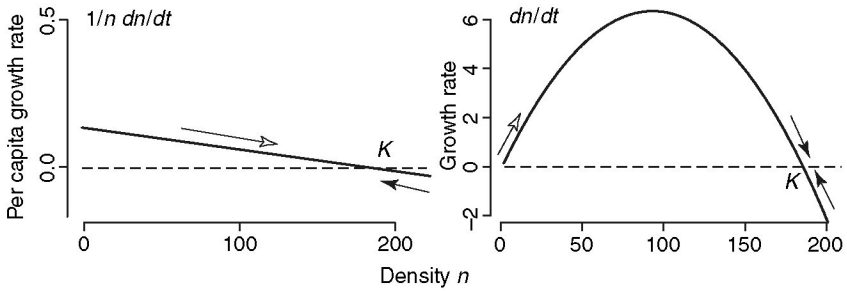

FIGURE 2.4. A linear effect of density on per capita growth rate, $1/n\ dn/dt$, from Equation 2.3, with parameters transformed to r and K. Although the effect of density on per capita growth rate is linear (*left*), this is a nonlinear differential equation, sometimes called the logistic equation for population growth. *Right* is the same model, plotted as simple population growth, dn/dt, rather than per capita growth. Parameters are taken from the fitted model for the BPF moose data (Fig. 2.5). Note that the carrying capacity estimate would be slightly below 200 individuals.

Pearl (Figure 2.1). The logistic model of population growth is nonlinear, because the second derivative of dn/dt with respect to n does not equal zero. This is clear from the "hump" in Figure 2.4b. Nonlinear models usually cannot be solved (the logistic model is an exception—Appendix B).

2.2.3 Discrete Time

Discrete time models describe changes that occur at discrete intervals in time, such as a population of organisms where births and deaths occur at discrete intervals. These may be periodic, example, annual, but they need not be equally spaced in time—Chapter 9 provides examples of unequally spaced intervals. A discrete time model consists of an expression describing change from one time increment to the next. If the time increment is, say, one unit, then there is an expression for $n_0, n_1, n_2, \ldots$. The time increment might be one generation of duration T, in which case there would be an expression for $n_0, n_T, n_{2T}, \ldots$. Discrete time process models are termed *difference equations*.

As an example, the simplest population model expresses density at the next time step as the sum of individuals present at the current time plus new individuals added by births minus those lost to mortality, $n_t = n_{t-1} + B_{t-1} - D_{t-1}$. Let $B_t = n_t b$ and $D_t = n_t \rho$, with the two new parameters being per capita rates. Then

$$\begin{aligned} n_t &= n_{t-1}(1 + b - \rho) \\ &= n_{t-1}\lambda \end{aligned} \tag{2.4}$$

where the composite parameter $\lambda = 1 + b - \rho$ is sometimes termed the *finite rate of increase*. This *first-order linear difference equation* has a solution that we can discover by iteration (Appendix B). This difference equation $n_t = f(n_{t-1}) = n_{t-1}\lambda$ is linear in terms of n, because $d^2f/dn^2 = 0$. In population biology, it is often written in terms of change in log density, $\ln n_t = \ln n_{t-1} + \ln \lambda$, to express change in proportionate terms, that is, on a per capita basis. The per capita rate of change is $\ln \lambda$, and is often represented by the quantity $r \equiv \ln \lambda$.

Nonlinear difference equation models are common in the environmental sciences. An example from population ecology refers to factors that might maintain a population within some range of densities, similar to the continuous logistic model of the previous section. Per capita birth rates, death rates, or both may decline as density increases due to diminishing resources, crowding, and so on. A conventional model that describes a decline in growth rate as density approaches some carrying capacity K is written as

$$n_t = n_{t-1}\lambda^{\left(1 - \frac{n_{t-1}}{K}\right)}$$

As density approaches K, the exponent approaches zero, and density does not change. If density is greater than K, then the exponent is negative, and density declines. Ecologists usually write this model in terms of the parameter r, or $\lambda \equiv e^r$, to obtain an equivalent discrete Ricker (1954) model

$$n_t = n_{t-1}\exp\left[r\left(1 - \frac{n_{t-1}}{K}\right)\right] \tag{2.5}$$

(Appendix B). This is a nonlinear difference equation, because $d^2f/dn^2 \neq 0$. An example is shown in Figure 2.5b, where interannual changes in moose density are modeled with an underlying process described by Equation 2.5.

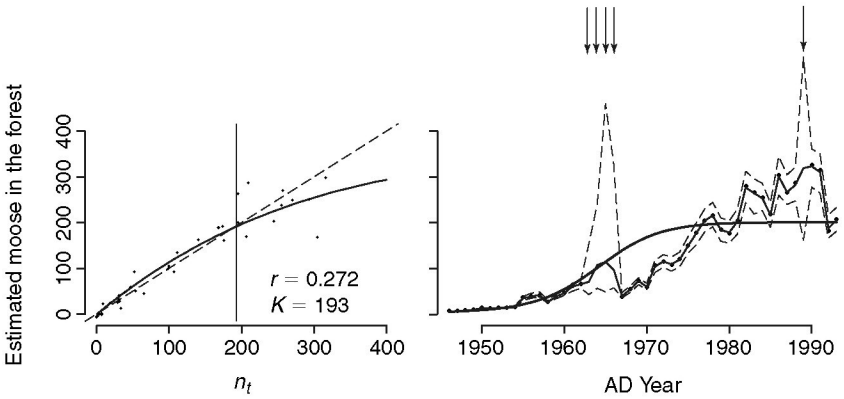

FIGURE 2.5. Models fitted to the BPF moose data (using methods of Chapter 9) compared with the deterministic version of the logistic model (smooth line). Left is the lag-one plot with observations and the fitted model (solid curve). This plot could be used to construct a cobweb diagram (Appendix B) that would show the constant solution to be stable: from any initial density we would converge to $n^* = 193$. This convergence comes from the curved relationship in the figure, *left* (Section B.5.4). *Right*, is the time plot. The vertical arrows indicate years of missing data. Dashed lines bound the 95 percent credible interval, indicating the degree of uncertainty as to the true population size. The uncertainty reflected in both graphs raises doubts about the existence of this stable solution $n^* = K$. From Clark and Bjørnstad (2004).

Discrete time models are often applied to processes that are continuous, either for convenience or for sampling considerations. In the latter case, it is important to recognize that discrete and continuous time models do not always exhibit the same behavior. Most ecological processes are not precisely continuous in time, yet difficult to characterize through a simple discretization of time. There can be some advantages to difference equations, particularly when data are concerned. It is becoming increasingly common to work directly with discretized versions of continuous time models with allowance for the error associated with discretization. Due to the focus on data, this book mostly discusses discrete time models. It is still important to understand continuous models, because environmental processes are often analyzed in continuous time. For example, the continuous formulation

$$\frac{dn}{dt} = rn\left(1 - \frac{n}{K}\right)$$

can be discretized to

$$\begin{aligned} n_{t+dt} &\approx n_t + dn && \text{current density plus increment in } n \\ &= n_t + \left[\left(\frac{dn}{dt}\right)dt + \varepsilon_t\right] && \text{increment in } n \text{ is linear} \\ & && \text{approximation plus error} \\ &= n_t + rn\left(1 - \frac{n}{K}\right)dt + \varepsilon_t && \text{substitute for } dn/dt \end{aligned}$$

To follow this transition from continuous to discrete time, recall from introductory calculus that dn is defined as the small increment in n that occurs with a small time increment, dt. After review of Appendix A, you should recognize this as the Taylor series for n_{t+dt}, with ε_t representing error that comes from ignoring higher-order terms. The discrete and continuous versions will behave similarly, provided that the time increment dt is not too large and nonlinearities are not too large. The process error term ε_t accommodates the error that comes from the fact that I did not integrate the rate equation, but, rather, simply scaled the rate of change, dn/dt, by the time increment dt. This error will often be small relative to other sources of process error, meaning that $rn(1 - n/K)$ is a crude approximation of the growth process. When discretized in this manner, the error associated with ε_t increases with the time elapsed between t and $t + dt$. Difference equations can be unstable due to the coupling of nonlinear dynamics with a time lag dt (May 1974). It is important to note, however, that the instability that can result from using a difference equation to approximate a nonlinear differential equation is less of an issue for data modeling than it is for simulation. With data modeling, we are not solving equations, but rather estimating parameters. We want to allow for process errors, regardless of their source. Scaling of process error is discussed in Chapter 9.

If generations overlap, dynamics depend on some underlying interactions that are not explicitly included in the models considered thus far. The difference equation $n_t = f(n_{t-1})$ says that density at time t is fully determined by density one time step ago. A model for growth that depends on population size in the more distant past could allow for change in age distribution (e.g., Nisbet and Gurney 1982). It is also possible that the population interacts with other species that are not explicitly included in the model (Schaffer 1981). Although unknown, it is possible to allow for such effects deterministically, by including delayed feedback. In Section 9.9, I mention models of the form

$$n_t = f(n_{t-1}, \ldots, n_{t-p})$$

This model says that density at time t directly depends not only on density at $t - 1$, but also at times back to $t - p$. In this case, p is the *order* of the model.

2.2.4 Discrete States and Structure

In addition to continuous versus discrete time, models have continuous versus discrete state variables. In the context of population growth, *state* refers to population size (number of individuals) or density (number per area or volume). The state variable could also refer to the concentration of a nutrient, to the volume or mass of moisture in soil, to the land cover classification of a county, or to the number of infections or deaths in a defined area. Confusion can arise when there are multiple state variables, because the terms *continuous state* versus *discrete state* are sometimes used to refer to the quantity within a given class or category or to the way in which the categories themselves are defined (Table 2.1). The differential equation and difference equation

TABLE 2.1.
Treatment of Time, State, and Structure in Some Common Process Models

	Time		State		Structure		
Process Model	Continuous	Discrete	Continuous	Discrete	Continuous	Discrete	Stochastic
Differential equation	x		x				
Difference equation		x	x				
Partial differential equation	x		x		x		
System of differential equations	x		x			x	
System of difference equations/matrix models		x	x			x	
Discrete state Markov process/cellular automata		x		x		x	
Birth/death process (continuous time Markov process)	x			x			
Integro-difference/ integral projection		x	x		x		
Random effects*		x	x	x	x		x

* Random effects arise in the context of data modeling, discussed in Chapter 9.

models discussed in the previous section contain a single state variable (density) that is continuous, in the sense that density can assume any values greater than or equal to zero (although the zero class is still typically viewed as discrete and requires special consideration—more on this later). Alternatively, a model for the number of births or deaths could be modeled with discrete numbers, the state variable assuming integer values 0, 1, 2, ...

A *structured model* indicates discrete classes or categories. A structured model can be written as (1) multiple attributes for a state variable or (2) multiple state variables. In the latter case, each of the classes is viewed as a different state variable and represented by a separate equation. Together the equations constitute a system of differential (continuous time) or difference (discrete time) equations. In the former case, attributes may be represented as continuously varying (e.g., age) or as discrete classes (e.g., classes for juveniles and adults). Either way, there is structure (e.g., age or stage) associated with the single state variable density.

If the stages are discrete, the full system of equations may be represented with a matrix having elements that describe transitions among states. This is standard practice when models are linear in state variables. Deterministic matrix models have become extremely popular in population ecology, in part because analysis is accomplished with linear algebra (most such models are

linear; see the thorough treatment of Caswell 2001). For example, the demography of a population could be represented as a collection of demographic rates organized in a table or matrix. Matrix models are discrete in time and in structure—all individuals have a specific class or stage to which they belong. However, within a stage, density is a continuous variable. A transition matrix contains demographic rates that are the basis for projecting stage structure forward in time. I take a few moments here to outline some of their key features.

Classes or stages are defined based on age (age-structured model) or other attributes (stage-structured model). Age-structured models can be viewed as a special case of stage-structured models. They are often used when age information is available, having the advantage that transitions among classes (i.e., ages) can occur at the same rate as the passage of time. Individuals who survive advance one age increment in one time increment. There is no uncertainty associated with timing of transition. The U.S. Census Bureau uses age structure, because it is available for a large segment of the U.S. human population. The transition matrix based on age is often called a *Leslie matrix* (Leslie 1945).

Stage structure is typically used when age is uncertain and when stages provide a useful summary of demographic rates. Perennial plants may have a number of stages, including mature individuals, rosettes, and seeds in the soil

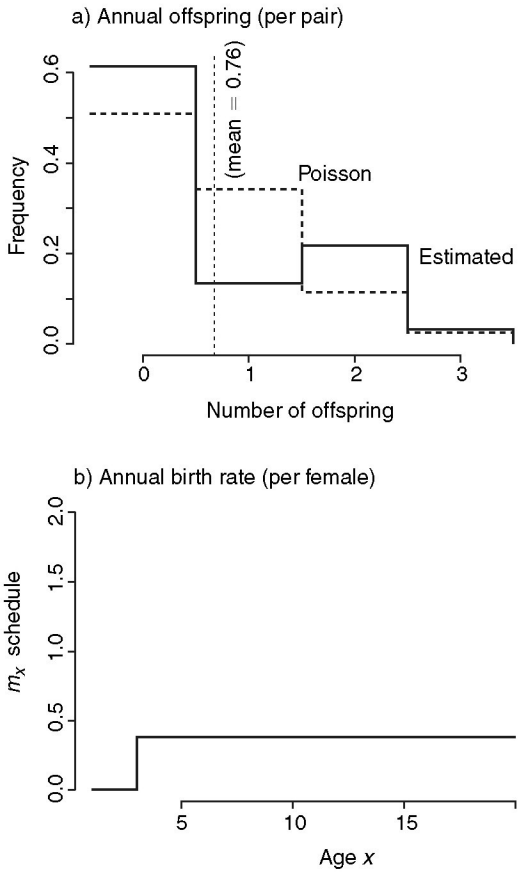

FIGURE 2.6. (*a*) Distribution of offspring from Barrowclough and Coats (1985) (solid line) compared with a Poisson distribution having the same mean (dashed line). (*b*) A fecundity schedule that assumes constant reproductive rate after reaching maturation at age 2 (Equation 2.6).

that cannot be aged accurately. Invertebrates pass through a discrete set of metamorphic stages that often are readily identifiable. Because transitions among stage classes are more variable than are those among age classes, models must accommodate additional considerations. Analysis involves more work, and estimates of transition rates tend to have greater uncertainty.

A classic example for the northern spotted owl (NSO) (*Strix occidentalis*) includes a fecundity schedule m_x, consisting of transitions related to births and deaths. A simple birth schedule includes a maturation age of $x = 2$ years and an average production of $b = 0.382$ offspring per year thereafter. This schedule could be written as

$$m_x = \begin{cases} 0 & x < 2 \\ b & x \geq 2 \end{cases} \qquad 2.6$$

(Lande 1987, 1988; McKelvey et al. 1993) (Figure 2.6b). Deaths are modeled using age-specific survival parameters, often interpreted as the probability that an individual in age class x will survive the interval from, say, t to $t + 1$. If age classes and time are measured in the same units, then survival is sometimes estimated as the fraction of individuals in age class x that survive during the age interval $(x, x + 1)$ relative to the fraction present at the beginning of the age x,

$$s_x = 1 - \rho_x = \frac{n_{x+1}}{n_x}$$

where n_x is the density of individuals in age class x and ρ_x is interpreted as the probability of mortality. The *survivor function* can be obtained from the age-specific survival parameters as

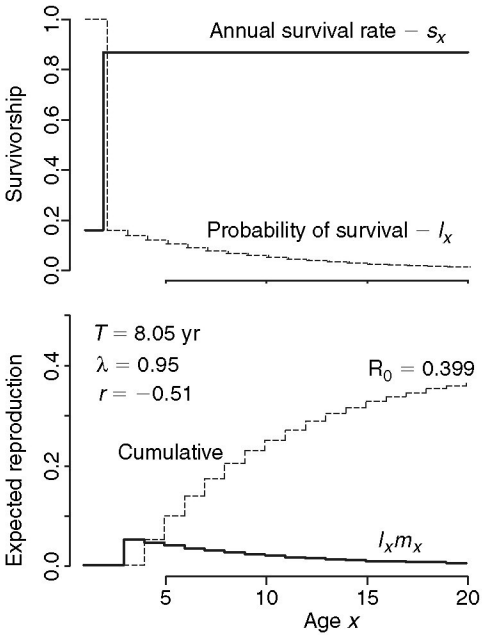

FIGURE 2.7 Life history schedules in this example for NSO. (*a*) Survival the first year is low, resulting in a large initial decline in the survivorship curve l_x, followed by geometric decline thereafter. (*b*) At the time of birth, the expected reproduction from an age x individual, m_x, must be scaled by its probability of surviving to age x, that is $l_x m_x$. Net reproductive rate R_0 is this expected rate summed over all age classes (Equation E.1).

$$l_x = \prod_{i=1}^{x-1} s_i = s_1 \times s_2 \times \cdots \times s_{x-1} \qquad 2.7$$

and interpreted as the probability of surviving to age x, obtained as the joint probability of surviving all previous age intervals (Figure 2.7a).[3] NSOs have high juvenile (first year) mortality. McKelvey et al. (1993) use a low juvenile survival rate of $s_1 = 0.159$ and a higher rate of $s = 0.868$ for subadults and adults. The plot of survivorship shows the large decline during the first year and slower decline thereafter (Figure 2.7a).

Analysis of population growth rate, population structure, and contributions of demographic rates to population growth rate can be accomplished directly from these equations (Appendix E) or using matrix algebra (Appendixes C, E). For the latter, demographic rates are assembled in a matrix **A**, having elements a_{ij}, describing the contribution of stage j to stage i. Here is a matrix for the NSO model just discussed,

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & b \\ s_1 & 0 & 0 \\ 0 & s & s \end{bmatrix} \qquad 2.8$$

The contribution of stage j to the population at the next time increment is determined by the elements in j^{th} column of **A**. Individuals in stage 1 can only survive to stage 2 ($a_{21} = s_1$) or die, with probability $1 - s_1$. Individuals in stage 3 not only can survive in stage 3 ($a_{33} = s$), but they can also produce offspring ($a_{13} = b$). Several examples in this book involve matrix population models (Sections 9.16, E.3). If a population grows linearly, and it possesses a stable stage structure, then the growth rate is obtained as the dominant eigenvalue of **A** (Appendix E).

Instead of discrete classes, there may be a continuous range of states, with the rate taken to be expectations for a random variable. The states may be represented by a continuous function describing how this expectation varies with size or age. For example, fecundity, natality, or age-specific fertility schedules can be represented by smooth functions of age or size (Figure 2.8). For the continuous age variable a, define expected fecundity $m(a)$ = expected offspring per unit time per age-a female per time. Because a is continuous, per capita births during the age interval $(a, a + da)$ are approximately $m(a)da$ for some age interval da. Time and age are often measured on the same units, with $da = dt$, although this need not be the case. This expectation is a continuous function of time and of state. As in Figure 2.8, such functions are often expressed in terms of size, say, $m(x)$ for diameter x, rather than age a. To obtain the age schedule, we need an additional piece of information, the growth rate dx/da,

---

[3] The indexing of classes needs careful thought (e.g., Cawell 2001). I adopt the following convention. Define the first age class to be $x = 1$. The individual enters this age class at birth. At this point, the survivor function is $l_1 = 1$. At the end of age class 1, the survivor function is $l_2 = l_1$ X $s_2 = s_1$. This is the probability of surviving to the beginning of age class 2. If the survival rate is the constant rate x_1 then $l_x = s^{x-1}$.

$$\frac{dm}{da} = \frac{dm}{dx}\frac{dx}{da}$$

(recall the chain rule). Continuous survival functions are summarized in Section E.6, with inference discussed in Sections 3.8 and 8.2.

In continuous time, structure can be included as systems of *ordinary differential equations* (ODEs) and *partial differential equations* (PDEs). A system of coupled ODEs describes rates of change among discrete stages. There is a rate equation for each state, and equations are coupled by terms that express the continuous transitions among these discrete states. Ecological models are often posed as coupled ODEs.

A PDE represents how the condition of the state variable changes continuously with time, in this case represented by the index of age and time on the state variable n,

$$\frac{\partial n(a, t)}{\partial t}$$

Models involve an additional equation for boundaries (the *boundary condition*). For an age-structured model, there is a boundary at age zero, which can be a function describing fecundity. Finally, there will a function for the initial condition. Unlike a simple differential equation, where the initial condition is a constant, with a PDE, the initial condition is a function of (in this case) the variable a, describing the age structure at time $t = 0$. PDEs apply to many types of structure, including size, physiological state, and location. Metz and Diekmann (1986) provide an overview. Parameterization can be challenging (Wood 2001), and stochasticity is typically ignored or limited to simple Gaussian variability. Some basic concepts are included in Section E.7.4. They appear in Chapter 10, where we consider diffusion.

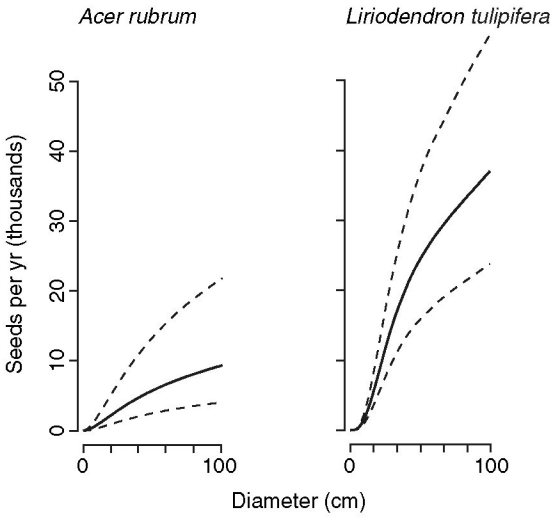

FIGURE 2.8 Fecundity schedules for two tree species treated as a continuous function of tree size. These are schedules for expected fecundity. Actual fecundity shows large fluctuations (chapter 10). From Clark et. al. (2004).

Integral projection models are discrete in time, but continuous in state (Easterling et al. 2000). The term *integro-difference* refers to the integrated effects of all stages at the current time on the density of stages at the next time. This combination provides an alternative to matrix projection models for cases where stages are not clearly differentiated, but shift more-or-less continuously. Over a discrete time increment, the population structure, described by a continuous variable x, changes as

$$n_{t+1}(x) = \int_0^\infty k(x|y)n_t(y)dy$$

where the kernel $k(x|y)$ represents transitions from state y to state x. The kernel summarizes transitions that result from fecundity, survival, and growth. Although not yet widely used, integral projection models could be more broadly applied (Section E.8), because they do not assume a single transition parameter. I do not explicitly consider this approach here, because Bayesian approaches in Chapter 9 can be used to address a wide range of deterministic and stochastic components of growth rates (Section 9.17.4). A brief introduction to stochasticity in models follows.

2.3 Stochasticity for the Unknown

Continuous time models can accommodate stochasticity (Bartlett 1960; Dennis et al. 1991; Lande et al. 2003), but they are typically treated in a deterministic fashion. In ecology, deterministic models are used to model expected values of a variable, for example, change in the mean density of a population. In such cases, the population *size* is assumed to be large. *Density*, being the number of individuals per unit area or volume, is a real number that results from division. Thus, a large population does not imply that density is large, because density could be defined over an arbitrary area or volume. Rather, both the population *density* and *size* must be sufficiently large that we can ignore the stochasticity represented by individual births and deaths. Large population size does not assure that such stochasticity can be ignored, because individual births and deaths can be important if density is effectively low. Moreover, variability among individuals can and, perhaps typically, does have impact even when populations are large. I devote substantial attention to this in later chapters.

Models where the state variable is represented by a discrete number of items or events involve probabilities of transition that result from, say, births and deaths. Unlike continuous state population models, which can be stochastic, but are typically treated deterministically, models for discrete numbers or events are invariably stochastic. The fluctuations that result from the fact that births and deaths are necessarily discrete events are termed *demographic stochasticity* (May 1973; Engen et al. 1998), although the same term has been applied to variability among individuals in terms of fitness (Lande et al. 2003). Because this book focuses on data-model connections, I deal with discrete

counts more often than is typical in traditional ecological models. A connection between discrete and continuous models is obtained with a deterministic, continuous model for an underlying process, such as natality or risk, superimposed with a stochastic, discrete model for the events themselves, such as births and deaths, for observations thereof, such as counts, or both. An example is shown in Figure 2.6, where the fecundity schedule m_x is a deterministic relationship (Figure 2.6b) for the mean number of births, from which actual births are taken to be stochastic (Figure 2.6a).

Several types of stochastic models are used for ecological processes. A *Markov chain* is a process where there is a probability associated with transitions among states, and each transition probability depends solely on the current state. In population models, transitions result from birth, growth, and death. A model for discrete numbers of individuals in each state could involve the same matrix that was described for the deterministic model in Equation 2.5, but now transitions occur stochastically, with each column of the matrix being one element of a multinomial parameter vector. Such models are taken up in Chapter 9.

There are many combinations of discrete versus continuous time, state, and structure (Table 2.1), and stochasticity can enter in a number of ways. For example, population models that are continuous in time, but have a discrete number of individuals in a state are termed *birth-death processes*. Let $p_n(t)$ be the distribution of densities n at time t, where $n = 0, 1, 2, \ldots$. A rate equation dp_n/dt describes the instantaneous rate of change in the probability of state n. Because time is continuous in this formulation, development of the model entails writing down instantaneous probabilities of transition from one state to a neighboring state. A birth-death process is a *continuous time Markov chain*. Birth-death processes are sometimes implemented in discrete time. If several births and deaths might occur from one time increment to the next, there are a potentially large number of transition possibilities, each with an associated probability. This model is sometimes termed a *branching process*. A matrix can be used to define probabilities of transition from a given state to any other state. Rather than defining a large number of transition probabilities among all pairs of possible states (population sizes), we might focus on a random variable representing the number of new individuals that enter or leave a population during a given time step (Kot 2001; Caswell 2001). I do not dwell on continuous time birth-death models, because they have not been well developed for data modeling, which is our focus, and potential for analysis is limited. Solutions are available for some simple linear models, and there are useful results for some nonlinear models (Nisbet and Gurney 1982; Ricciardi 1986). Overviews of such models include Bartlett (1960) and Kot (2001).

A *random walk* is a Markov process where there are nonzero probabilities of moving to either of the adjacent states, there is zero probability of all other transitions in a single time increment, and the transition probability does not depend on the state variable itself. Models of random walks arise in a number of contexts, often where the state variable is location rather than population size. A biased random walk has unequal probabilities of moving to adjacent states, with the bias sometimes being termed *advection*.

In addition to some of these standard models that regularly appear in the ecological literature, there is an additional class of models that is increasingly applied to data, termed *random effects* models (Table 2.1), which come from the statistical literature (Clark 2003; Clark et al. 2003a). In this case, the structure is defined stochastically to allow for the differences among sample units that will be estimated from data, as opposed to being defined in advance. The variability among seedlings in terms of growth response to light is an example (Section 1.2.2).

2.4 Additional Background on Process Models

The general classes of models mentioned in this chapter are thoroughly discussed in a number of recent texts on ecological theory applied at the population level (e.g., May 1973; Edelstein-Keshet 1988; Murray 1989; Hastings 1997; Kot 2001; Caswell 2001; Case 2000) and ecosystem level (DeAngelis 1992; Agren and Bosatta 1996). Some basic elements of analysis are provided in the form of appendixes for basic process models (Appendix B), matrix manipulations (Appendix C), and life history calculation (Appendix E). This material is placed in appendixes, because it is referenced in several sections of the book and is intended for use on an as-needed basis.

Part II Elements of Inference

In this section I move directly to inference, the practice of estimating unknowns based on models and data. I begin with the likelihood, introduce inference by maximum likelihood, and then mention the method of moments. The remaining chapters of Part II take up confidence envelopes and the basics of model selection.

3 Point Estimation: Maximum Likelihood and the Method of Moments

3.1 Introduction

The basic models from Chapter 2 are used to describe a process. The process might be viewed as the core of a stochastic model that brings in sources of uncertainty and additional variability associated with the collection of data. I now consider connections between models and data.

Statistical inference involves constructing and evaluating models in the context of data. Inference can be motivated by a desire to understand the process that generated the data, to make predictions, and/or to inform a decision-making process. The process can involve comparing models, evaluating plausible parameter sets, and assigning goodness of fit. Probability statements are made in light of data, which might come from several sources and accumulate over time. Decisions involve combining the uncertainties estimated from the analysis with perceived benefits and consequences (gains, losses, risks) of alternative decisions.

There are two common frameworks for formal statistical inference that have broad application. I refer to these as classical and Bayes. It is important to understand both. Likelihood is fundamental to both frameworks, so I begin here. I then take up maximum likelihood for point estimation. A *point estimate* is the value of a parameter that finds most support in the data set. After introducing the concepts using examples based on survival, I consider applications to population growth and some standard approaches to survival analysis. At the end of the chapter, I discuss the method of moments and some general considerations about sampling distributions. This is a good time to begin familiarizing yourself with the material in Appendix D.

3.2 Likelihood

Likelihood is the probability of a data set given that the model for those data is deemed to be correct (Fisher 1934). The *likelihood principle* says that a model (with parameter values) is more likely than another if it is the one for

which the data are more probable. This principle provides a basis for saying which models explain a data set better than others, that is, model selection. The easiest way to grasp this concept is with an example.

3.2.1 An Exponential Model

To estimate mortality rate from an experiment, begin by specifying a model. Initially I assume that there is one observation, the life span of a plant. If mortality rate is continuous and constant, say, ρ per unit time, then the probability that death occurs at an age between a and $a + da$ can be represented by the notation $\Pr\{a < a_1 < (a + da)\}$ (Section E.6). This notation represents the probability that the observed time of death, a_i, occurs after age a and before age $a + da$. The subscript i could indicate that this is the time of death for the i^{th} individual. This is a joint probability of two events, (1) the plant is still alive at a, and (2) the plant dies before $a + da$. If these two events are independent, then the joint probability is their product:

$$\Pr\{a < a_i < (a + da)\} = \Pr\left\{\begin{matrix}\text{die now given that}\\ \text{plant is still alive}\end{matrix}\right\} \times \Pr\left\{\begin{matrix}\text{plant is}\\ \text{still alive}\end{matrix}\right\}$$

$$\approx h(a)da \times l(a)$$

The factor $h(a)$ is the age-specific risk per unit time, and it is scaled by the duration of the increment da to obtain the (approximate) probability of death between a and $a + da$ (Figure 3.1). The last factor $l(a)$ is the survival function, or the probability of surviving until age a (Equation E.18). If the age-specific risk is a constant $h(a) = \rho$, then $l(a) = e^{-\rho a}$, and we have

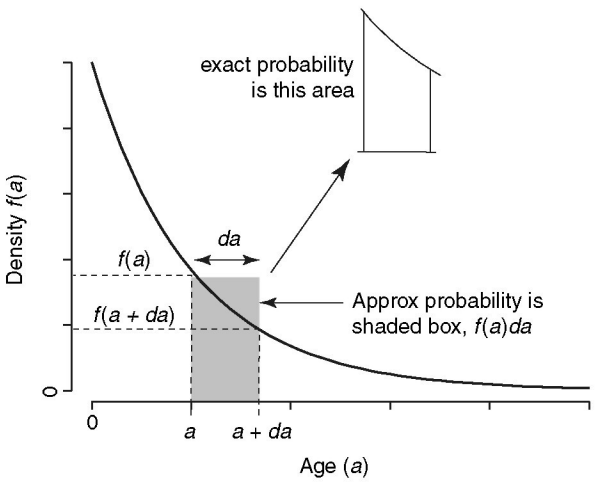

FIGURE 3.1. The relationship between probability of death in an age interval from a to $a + da$ (extracted at *upper right*) and the approximation $f(a)da$ (shaded). As the interval da becomes small, the two areas converge.