Preface

Why This Book?

This book is about the process of gaining new knowledge about ecology using models and data. We wrote it for several reasons. An overarching motivation is that our satisfaction with scientific work has been palpably enhanced by understanding this process from start to finish, with no gaps where faith must fill in for understanding. We write this book because what we write about is easily the most intellectually satisfying material we have learned in our careers. It has made the work we do every day more enjoyable. We are confident that your research in ecology will be accomplished with greater satisfaction and reward if you master the concepts we describe here.

There are more specific motivations as well. We teach graduate students ecological modeling using Bayesian methods. Colleagues whose students have taken our classes often ask us what they should read to get a big picture understanding of the Bayesian approach. They seek a book explaining statistical principles of Bayesian modeling written in language accessible to nonstatisticians. They may never write a line of computer code implementing Bayesian methods, but they realize the importance of understanding them. Our colleagues want to be able to appreciate contemporary scientific literature, to review papers and proposals, and to mentor students who *do* know how to write code. We decided to write the book they asked for.

Of course, there are now many excellent texts on Bayesian modeling in ecology (e.g., Clark, 2007; McCarthy, 2007; Royle and Dorazio, 2008; Link and Barker, 2010; Kéry, 2010; Kéry and Schaub, 2012), texts that we use all the time. For the most part, these books emphasize computational methods over concise explanation of basic principles. Many of these are difficult to appreciate without a background in mathematical statistics—training that most ecologists lack. Our book will complement the existing crop by providing the basic understanding of mathematical and statistical principles needed to use more advanced texts effectively.

A third motivation also comes from our personal experience. Students and colleagues often come to us with a modeling problem they are working on, a problem that uses Bayesian methods. They can write code in one of the popular implementations of Gibbs samplers, WinBUGS (Lunn et al., 2000) or JAGS (Plummer, 2003), and often they bring a thick stack of code with them. Although they can create computer programs that give them answers, they cannot write the mathematical expression for the model that underpins their work. They are unsure about their starting point and hence are not entirely confident about where they have ended up. They have difficulty writing a mathematical expression that clearly communicates their analysis in manuscripts and proposals. As you will see as this book unfolds, we believe that reliable analysis must begin with a model written in mathematical symbols and operators, not in a computer language. Writing models is the foundation of good science.

The ability to write models leads to our next motivation. There is a diminishing set of important questions in ecology that can be answered by a single investigator working alone, even a very good one. Instead, what remains are problems solvable only by the application of intersecting sets of talents, skills, and knowledge. This book offers hold-in-your-hand evidence of the value of collaboration between a statistician and an ecologist, but there are many other examples (Gross et al., 2002, 2005; Clark, 2003b, 2005; Latimer et al., 2006; Farnsworth et al., 2006; Cressie et al., 2009; Rotella et al., 2009; Webb et al., 2010; Eaton and Link, 2011; Wilson et al., 2011; Fiechter et al., 2013; Peterson et al., 2013). These collaborations require a mutual understanding of basic principles and a shared vocabulary.

Our final reason for writing this book relates to the first one. Most ecologists working today, save perhaps some of the youngest ones, received training in statistics emphasizing procedures over principles. We emerged from this training understanding how to do "data analysis" using a suite of recipes—t-tests, analysis of variance, regression, general linear models, and so on. The diligent and ambitious among us might have added work in sampling and multivariate statistics. Mathematical statistics was reserved for statistics majors.

The outcome of this approach to training is seen in a revealing way in the frontispiece of a once widely used text (Sokal and Rohlf, 1995), which displays a table of analyses resembling a dichotomous key in taxonomy. If our data are like x, then we should use analysis y; otherwise, we should use analysis z. Those of us trained this way had precious little understanding of *why* we should use analysis y over z beyond the authority provided by that table. Moreover, there was a limited range of kinds of data for which this taxonomic approach would serve, and we were stymied if the observations we worked hard to obtain were not found somewhere in the table. Sometimes we would heroically bend those observations to make them fit in one of the table's narrowly defined cells. A narrow range of approaches to analysis constrains the questions that ecologists are willing to ask—we are uncomfortable posing questions for which we see no analytical route to insight. If the only tool in our locker is analysis of variance, then the world we study must be composed of randomized plots. Some of the current Bayesian books are organized in a similar way—around procedures rather than principles.

All research problems that ecologists seek to solve have aspects in common and aspects that are unique. The unique features of these problems argue for a principled approach to insight. A lean set of modeling principles can substitute for volumes of statistical facts. Understanding these principles enables us to design routes to insight uniquely suited to each of the diverse problems we confront over the course of a research career. Providing this understanding is the main reason for writing this book.

Goals

The overarching goal of this book is to train ecologists in the basic statistical principles needed to use and interpret Bayesian models. An essential part of meeting that goal is to teach how to write out accurate mathematical expressions for Bayesian models linking observations to ideas about how ecological systems work. These expressions form the foundation for inference. It is our ultimate aim to increase the intellectual satisfaction of ecologists with their teaching, research, and peer review by providing a solid, intuitive understanding of how we learn from data and models in the Bayesian approach. Finally, we aim to enhance the quality of collaboration between ecologists and statisticians by training ecologists in the mathematical concepts and language of statistics.

Approach

Organization

The book is organized to provide the understanding needed to support a general process for model building in ecology, a process that applies to virtually all research problems. We outline a flow of tasks in building models that we have found helpful explaining the process (figure 0.0.1). The sequence here is not immutable, but it offers a useful schematic of the steps needed to build a revealing Bayesian model. We will return to this diagram throughout the book to show how specific topics fit into the larger task of model building.

We have organized the book in three parts. The aim of part I is to provide a basic understanding of the principles underpinning Bayesian analysis (fig. 0.0.1). We begin part I with a preview of the entire book to motivate what will follow. We then spend some time encouraging thinking about deterministic models and how they have been traditionally used in ecology (fig. 0.0.1 A). Next, we cover basic principles of probability and probability distributions. Think of this as a crash course in mathematical statistics, a prospect that might not be thrilling on the face of it, but we urge you to read this material carefully if it is not familiar to you. Gaining familiarity will allow you to understand the powerful ideas that follow and, ultimately, to understand what you are doing when applying Bayesian methods to your own research.

Knowledge of statistical distributions provides the foundation for understanding maximum likelihood approaches to parameter estimation. Likelihood is a central element of Bayesian models. We explain the link between likelihood and Bayesian approaches to inference by explaining the theory underpinning the Bayesian approach. We develop the concept of Bayes as "likelihood reweighted." We then dissect Bayes' theorem piece by piece to explain its components: the posterior distribution, the likelihood, the prior, and the marginal distribution of the data. We introduce some uniquely Bayesian concepts likely to be unfamiliar to most ecologists, for example conjugate relationships between likelihoods and priors, which turn out to be critical in the second part of the book. We finish part I by applying the basic statistical concepts we have developed to specific problems in ecological research, illustrating how mathematical expressions are built to link models and data (fig. 0.0.1 B). We show how models of complex phenomena, models with many parameters and latent quantities, can be broken into manageable chunks using hierarchical modeling.

Part II lays out the nuts and bolts of how we use the principles developed in part I to learn about parameters, unobservable states, and derived quantities. We clearly explain Markov chain Monte Carlo and Gibbs sampling, the numerical algorithms that have revolutionized the ability to gain insight from hierarchical models (fig. 0.0.1 C). Part II closes by describing how we check models to assure their fidelity to statistical assumptions and how we make inference from a single model or from multiple models (fig. 0.0.1 D).

Part III includes a series of problems and worked solutions drawn from several subdisciplines of ecology. These problems require application of the concepts and principles we developed in part I and II, emphasizing model specification (fig. 0.0.1 B), a skill that we believe is not emphasized in other texts. Our intention in part III is to encourage active model building and to show how the same approach to model specification can be fruitfully applied to a broad range of problems in ecology.

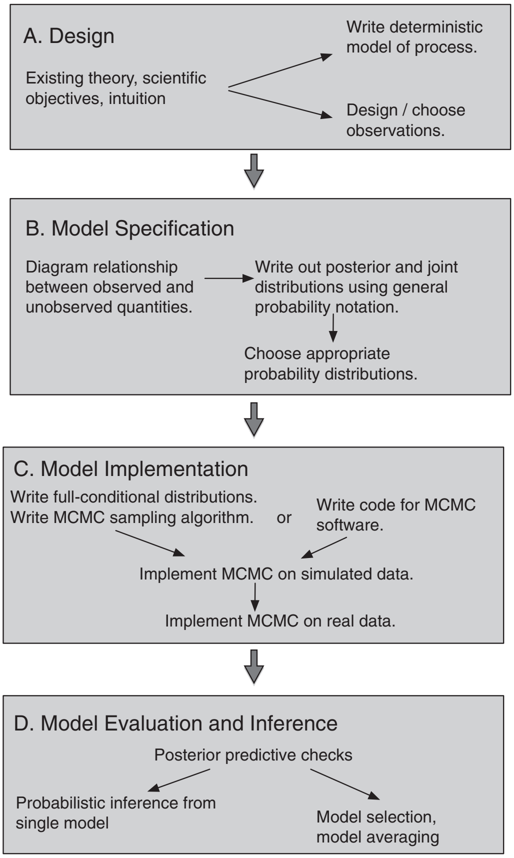

Figure 0.0.1. Gaining insight from Bayesian models involves the same sequence of steps for virtually all research problems, steps that fall into four broad groups. The sequence of steps is indicated by the long and short arrows. This book is organized to explain these steps in a logical way. (**A**) Design is not uniquely Bayesian, but we include it here because we want to encourage the thoughtful development of mathematical models of ecological processes as a starting point for analysis (chapter 2). (**B**) The premise of this book is that mathematical models must be combined with data to allow us to learn about how ecological systems operate. Chapters 3, 4, 5, and 6 show how we specify models to include data. (**C**) A key idea is that a properly specified model provides all we need to know to implement the enormously powerful algorithm Markov chain Monte Carlo (MCMC). We provide a principled understanding of how and why MCMC works in chapter 7. (**D**) We then cover how to use output from MCMC as a basis for inference from single models (chapter 8) and from multiple ones (chapter 9). Finally, we return to the key process of model specification **(B)** in chapter 11 by providing a series of problems challenging you to formulate Bayesian models.

Crosscutting Themes

We err on the side of excessive explanation of notation and equations. We believe that a formidable impediment to understanding statistics by ecologists is that ecologists, for the most part, don't get up in the morning every day and write statistical models. The authors of most statistical textbooks *do*. Consequently, notation that is compact and efficient in the eyes of the practiced is murky for the rest of us. We promise to use a consistent notation in expressions that are fully explained, believing that clarity trumps elegance. We cannot avoid equations, but we can strive to make them understandable.

We use model diagrams[1] throughout the book. These sketches portray stochastic relationships among data, unobserved states, and parameters. We show how these diagrams, properly composed, provide a blueprint for writing out Bayesian models, and ultimately, for designing the samplers that allow us to obtain probability distributions of the quantities we seek to understand.

Unlike many existing books on Bayesian methods, ours will not emphasize computer code written in any specific language. We will teach algorithms but not coding. We choose this approach because there are an increasing number of software packages that implement algorithms for Bayesian analysis (e.g., Lunn et al., 2000; Plummer, 2003; INLA Development Team, 2014; Stan Development Team, 2014). The diversity of this software is likely to expand in the future. Including today's favorite flavor of code in our book assures that it will become obsolete as a new favorite emerges. Moreover, writing your own algorithms in the programming language of your choice is not all that difficult and offers the added benefit of forcing you to think through what you are doing. We make a single exception to this "no-code" rule in part II by including a script to illustrate the relationship between mathematical expressions and their implementation.

[1]Formally known as directed acyclic graphs or, more succinctly, Bayesian networks.