Alan Agresti, Maria Kateri, Ranjini Grove, and Antonietta Mira

Foundations of Bayesian Statistics for Data Scientists, with R and Python

Dedication

- Alan is grateful to Jacki for her support and encouragement, not only for this book but for the long run of book projects in our 40 years together.
- Maria is grateful to her parents, Athena and Dimitris, for making everything possible through all stages of her life, for their encouragement, and for always being there when needed.
- Ranjini is grateful to her students, past and present, for continually inspiring her to grow, to question, and to become a better teacher.
- Antonietta is grateful to her father, Eugenio, for his enduring inspiration and intellectual companionship, and to her mother, Lavinia, whose generous encouragement has been a constant source of motivation and care.

Contents

Preface

This book presents an overview of the Bayesian approach to applying the most important inferential methods of statistical science. The primary intended audience is undergraduate and masters-degree students who plan to become data scientists. Statistical science is by now a large subject, having many distinct specialties. This book highlights the aspects of Bayesian statistics with which we believe that a data scientist should be familiar:

- Bayes rule and its applications
- Conjugate and non-conjugate prior and posterior distributions
- Inference for a proportion and for comparing proportions
- Inference for a mean and for comparing means
- Inference for linear models for quantitative responses
- Inference for generalized linear models for binary and count responses, including hierarchical models
- Bayesian approaches for model checking, model comparison, and model averaging
- Markov chain Monte Carlo computational methods

For a prerequisite, the student should be familiar with calculus and have taken a Statistics course that presented basic rules of probability, probability distributions and expectations, and the tenets of the traditional, so-called *frequentist* approach to statistics, including sampling distributions, likelihood functions, the basic inferential methods of point estimation, confidence intervals, and significance tests, and linear regression models. Each chapter briefly reviews the primary frequentist methods for the topic of that chapter before introducing corresponding Bayesian methods. For this background, we also occasionally refer to a companion textbook, *Foundations of Statistics for Data Scientists, with R and Python*, written by two of us (Agresti and Kateri). Like that book, this book presents some substantive theory as well as the methods, so the book is designed for a reader who wants to *understand* Bayesian methods rather than merely know how to use them. For each Bayesian method, we show how with certain prior distributions or a large sample size, nearly identical results occur as with corresponding frequentist methods. We also explain how the Bayesian analogs of frequentist confidence intervals and P-values have simpler and more natural interpretations, but require the additional structure of prior distributions and sometimes increased computational complexity.

As in the Agresti/Kateri companion textbook, the focus is not on mathematics but on showing how to implement the statistical methods with modern software, emphasizing appropriate simulation methods. To use and properly interpret Bayesian or frequentist inferential tools of modern statistical science, computational skills are as important as mathematical skills. Throughout the book, examples with real data show how to use the free statistical software `R` to implement Bayesian statistics. The book also contains software appendices that present greater detail about `R` as well as introducing `Python` to conduct

Bayesian statistical analyses. The `Python` appendix shows analyses for the same examples introduced in the chapters with `R`. The software appendices also introduce additional analyses that supplement the examples presented in the chapters.

Use of this book as a course textbook

This book is designed as a textbook for a course on the Bayesian approach to statistical science for undergraduate and masters-level students majoring in Data Science, Statistics, or Mathematics. It can supplement a book such as the Agresti/Kateri textbook in a course on the foundations of statistical science (such as a "mathematical statistics" course) or it can be used in a follow-up course of about 30 lectures that focuses on an introduction to Bayesian statistics. It also serves graduate-level programs that have a heavy focus on statistical science, such as econometrics, psychometrics, and operations research. Furthermore, it should be useful to undergraduate and graduate students in the social, biological, and environmental sciences who choose Statistics as their minor area of concentration, so they can learn about the Bayesian versions of frequentist statistical methods with which they are already familiar.

Instructors may prefer to skip some of the less central or more technical material, such as the sections that have an asterisk (*) next to their titles. Because the `Python` appendix shows analyses for the examples introduced in the chapters with `R`, an instructor can use either as the main software for a course on Bayesian statistical methods. Each chapter contains many exercises for students to practice and extend the theory and methods. The exercises are grouped into two parts: Exercises in *Data Analysis and Applications* request that students perform data analyses similar to the ones presented in that chapter. Exercises in *Methods and Concepts* ask questions about properties of Bayesian statistical methods, conceptual questions about their bases, and also provide extensions of that chapter's results. Appendix C contains brief solutions for most odd-numbered exercises.

In preparing this book as a companion to the Agresti/Kateri textbook, Agresti has taken main responsibility for Chapters 1-6 and Sections 7.3, 7.5, A.3.4, and A.3.5, and Kateri has taken main responsibility for Appendix A about `R` and Appendix B about `Python`, the expanded appendices about `R` and `Python` at the book's website, and most of the figures and some software examples in the chapters. Agresti and Kateri have added two co-authors to this book project who have contributed additional material. Ranjini Grove has contributed exercise solutions and created a structured package of teaching resources —- including a syllabus, labs, and lecture slides – intended to guide instructors in integrating the book seamlessly into their courses. Antonietta Mira has contributed Sections 7.1, 7.2, and 7.4 and their exercises and solutions, as well as slides and .Rmd files to further help instructors prepare their own material in teaching using this textbook. We authors welcome you to email us about any errors that you notice or with comments that we can take into account in any future editions of this book.

Additional resources

The book's website `http://bayes4ds.rwth-aachen.de` contains all the data files, longer and regularly updated versions of the `R` and `Python` appendices, corrections, and resources for instructors.

Acknowledgments

Many thanks to those who provided comments or suggestions for us in preparing this book. We give very special thanks to Prof. Linda J. Davis, who gave us a great number of helpful suggestions and pointed out many errors that needed correction. She is an incredibly helpful reviewer, the best that any of us have ever had. Thanks also to Chris Ickler, David Ríos Insua, and Cristiano Varin, who also gave us excellent feedback and numerous helpful suggestions. Others who gave us feedback include Sid Chib, Francesco Denti, Anna Gottard, Alessandra Guglielmi, David Hitchcock, Beth Johnson, Gabriele Marchi, Kerrie Mengersen, Anish Mukherjee, Antonio Di Noia, Hubert Pawlusinski, Fabrizio Ruggeri, and Erica Trofimov. Thanks to Jonas Meyer for help in setting up the book's website. Grateful thanks also to the reviewers who provided comments about our manuscript to Taylor & Francis. Finally, special thanks to Lara Spieker, Editor of Statistics for CRC Press and Taylor & Francis, for her encouragement and support in this book project.

ALAN AGRESTI, University of Florida (emeritus)
Gainesville Florida and Brookline Massachusetts, USA (agresti@ufl.edu)

MARIA KATERI, RWTH Aachen University
Aachen Germany (maria.kateri@rwth-aachen.de)

RANJINI GROVE, University of Washington
Seattle Washington, USA (grover4@uw.edu)

ANTONIETTA MIRA, Università della Svizzera italiana and Insubria University
Lugano Switzerland and Como Italy (antonietta.mira@usi.ch)

November 2025