International Statistical Review

International Statistical Review (2026)

Check for updates

Book Review

Foundations of Bayesian Statistics for Data Scientists: With R and Python
Alan Agresti, Maria Kateri, Ranjini Grove and Antonietta Mira
Chapman & Hall/CRC, 2026, 452 pages, £62.99, paperback
ISBN: 978-1-0412-0292-9

Readership: Advanced undergraduate and master's students in data science, statistics and mathematics, as well as applied researchers and practitioners seeking an introduction to Bayesian methods.

Bayesian statistics is now a standard component of modern statistical science, yet introductory textbooks still face a difficult pedagogical challenge. Some emphasise conjugate models and analytical derivations, leaving readers poorly prepared for contemporary modelling practice. Others move quickly to software and computation, sometimes at the expense of the statistical principles needed to understand what is being fitted and why. *Foundations of Bayesian Statistics for Data Scientists* occupies a useful position between these extremes. Its contribution is primarily pedagogical rather than methodological: It offers a broad, model-oriented route into Bayesian statistics for readers who already have a foundation in classical inference and wish to move towards Bayesian reasoning, modelling and computation.

The book is organised into seven substantive chapters. An introduction to Bayesian statistics is followed by chapters on inference for proportions and means. The discussion then turns to linear and generalised linear models, followed by Markov chain Monte Carlo (MCMC) computation and diagnostics. A final chapter addresses model choice and a wide range of extensions. Substantial appendices provide computational material in R and Python, together with solutions to odd-numbered exercises. The intended audience is clearly defined. Although introductory in Bayesian statistics, this is not an introduction to statistics in general: Readers are expected to know basic probability, likelihood, point and interval estimation, significance testing and linear regression.

The book's most distinctive pedagogical feature is its comparative structure. Familiar frequentist methods are repeatedly reviewed before their Bayesian counterparts are introduced. Confidence intervals, P-values, least squares, maximum likelihood and classical regression inference therefore serve as bridges to Bayesian ideas. This approach should work particularly well for students emerging from a conventional statistics curriculum. It helps readers understand not only how the two inferential traditions differ but also why they can sometimes yield similar numerical conclusions, for example, under weak prior information or with large samples. The price of this accessibility is that readers with strong prior training may find the route to specifically Bayesian material longer than necessary.

The opening chapter illustrates both the strengths and the limits of this approach. Rather than beginning immediately with Bayes' theorem, the authors first ask what should precede any

© 2026 International Statistical Institute.
All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

analysis: what question the study was designed to answer, how the data were collected, what sources of bias may be present, which variables are relevant, what assumptions are reasonable and to what population the resulting inference applies. This is a welcome choice, as it places Bayesian analysis within a broader process of statistical reasoning rather than presenting it as a collection of formulae. The subsequent contrast between frequentist and Bayesian inference is concrete and accessible, particularly in the discussion of confidence intervals and p-values.

The opening treatment is deliberately traditional. Bayesian probability is introduced primarily through a subjective interpretation, in which probability reflects a person's assessment of uncertainty given available information and experience. This is an accessible starting point, but it is narrower than the range of prior constructions encountered in contemporary practice. Priors may also arise from external evidence, hierarchical structures, formal reference criteria or regularisation objectives. The book later broadens the practical discussion of priors, but the initial philosophical framing is intentionally simple. The breadth of the intended audience also produces unevenness in level: After a useful conceptual comparison of inferential paradigms, the text returns to elementary material on sample spaces, events and probability axioms. This makes the book self-contained, although mathematically mature readers may find some early sections slower than necessary.

The chapters on proportions and means provide a bridge from foundational concepts to statistical modelling. Their scope is broader than that of an introduction confined to conjugate prior-posterior calculations. The contents include shrinkage, decision-theoretic interpretations of estimators, posterior intervals, prediction, improper priors, empirical Bayes methods, Bayes factors and sensitivity to prior specification. These ideas establish useful conceptual continuity, as shrinkage, prediction and prior sensitivity reappear later in more complex models.

The chapter on linear models offers perhaps the clearest illustration of the book's central trade-off. Before introducing Bayesian regression, the authors reconstruct the classical framework in considerable detail, reviewing least squares, correlation, regression towards the mean, multiple regression, centring, standardisation, sums-of-squares decompositions, R-squared, adjusted R-squared and classical t- and F-based inference. For students with uneven preparation, this is a considerable advantage. The discussion is careful in ways that are sometimes absent from more compressed data-science texts. In particular, the authors distinguish the conditional mean structure from the distributional assumptions of the normal linear model and make clear that least-squares estimation itself does not require normality, whereas familiar classical inferential procedures require additional assumptions.

Particularly valuable is the discussion of marginal and conditional effects. The book emphasises that a predictor's coefficient in a multivariable model need not agree with its unadjusted association and may even differ in sign. The treatment of centring and standardisation is similarly useful, as parameterisation is directly linked to interpretation. These are important points for students, who are often taught to interpret coefficients mechanically. For readers already comfortable with regression modelling, however, the extensive classical review may feel disproportionate in a book whose primary purpose is Bayesian statistics.

The chapter on generalised linear models (GLMs) is, in my view, one of the book's strongest parts. Logistic and Poisson regression are introduced within a unified framework comprising a response distribution, a linear predictor and a link function. The need for nonlinear links is motivated by the relationship between the permissible range of the conditional mean and the

© 2026 International Statistical Institute.

17515823, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/insr.70060 by Ciara Wade - NICE, National Institute for Health and Care Excellence , Wiley Online Library on [27/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

unrestricted range of the linear predictor. Thus, the logit and log links emerge as modelling devices with a clear purpose rather than as formulae to memorise. The discussion contrasting transformation of the response with direct distributional modelling is balanced: The authors explain why suitable probability models often provide a more coherent analysis of non-normal responses, while avoiding the simplistic conclusion that transformations are always inappropriate.

The Bayesian treatment of GLMs is also commendably candid about prior specification. Priors on regression coefficients can be difficult to interpret under nonlinear links, and the book discusses standardisation, elicitation on the response scale, diffuse and heavy-tailed priors and shrinkage. The Bayesian case is especially persuasive when grounded in practical modelling problems. The discussion of binary-data configurations in which maximum-likelihood estimates may diverge, while proper priors yield finite posterior inference, gives students a concrete reason to view Bayesian regularisation as more than a philosophical alternative to classical inference. The chapter is highly ambitious, extending from basic GLM structure to overdispersion, negative-binomial models, hierarchical priors, multilevel models and generalised linear mixed models. The breadth is useful for orientation, although readers will need specialised sources for deeper treatment of several of these topics.

The chapter on computation follows the same broad philosophy. Monte Carlo methods and Markov chains lead to effective sample size, R-hat, trace plots and autocorrelation diagnostics, followed by Metropolis, Metropolis-Hastings and Gibbs sampling. Hamiltonian Monte Carlo and approximate Bayesian computation are also introduced. The ordering is sensible: Models and inferential ideas are established before computational algorithms become the focus, reducing the risk that students equate Bayesian statistics with running MCMC software. The emphasis on diagnostics is particularly welcome. Modern software makes complex Bayesian models easy to fit, but not necessarily to fit reliably. More advanced algorithms necessarily receive briefer treatment, so their role is more orientational than instructional.

The final chapter is unusually wide-ranging for a foundational text. It covers model choice, Bayes factors, posterior model probabilities, model averaging, sensitivity analysis, heavy-tailed priors, classification, network models and missing-data methods. This breadth connects the foundations of Bayesian inference to a wider data-science landscape, but it also makes the chapter somewhat eclectic. The book's originality lies less in any individual topic than in bringing these topics together in a single transition text that links a conventional statistics curriculum to a broad Bayesian repertoire.

The computational organisation deserves separate comment. R is used throughout the main text, whereas Python is developed primarily in an extensive parallel appendix. The Python material is substantial and largely mirrors the progression of the substantive chapters, covering estimation, prediction, linear and generalised linear models, hierarchical modelling, MCMC diagnostics, classification and missing data. This makes the book useful to both R and Python users, although the two languages are not symmetrically integrated into the main narrative. The separation between statistical development and software is both a strength and a limitation: it prevents the main exposition from being dominated by code, but readers working computationally may need to move repeatedly between chapters and appendices.

As a course text, the book has several practical strengths. Technical sections can be omitted in less advanced courses, exercises range from routine practice to methodological extensions and selected solutions are provided. The use of real datasets across social, biomedical and other

© 2026 International Statistical Institute.

17515823, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/insr.70060 by Ciara Wade - NICE, National Institute for Health and Care Excellence , Wiley Online Library on [27/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

applied domains gives the book an empirical flavour that should appeal to data-science students. The worked examples generally link output to substantive interpretation rather than using data merely to demonstrate syntax. The book is therefore especially suitable for advanced undergraduate or master's courses in statistics, data science or quantitative disciplines, where students have already encountered classical inference but have little or no Bayesian background.

Overall, *Foundations of Bayesian Statistics for Data Scientists* is a substantial and carefully structured introduction. Its principal strength and limitation arise from the same design choice. By repeatedly reconstructing familiar frequentist ideas before introducing Bayesian alternatives, the authors make Bayesian methods accessible to conventionally trained students and reveal important connections between inferential traditions. The cost is that experienced readers may find the route to specifically Bayesian material longer than necessary, while the breadth of coverage means that several modern topics can only be introduced briefly. The book is best understood neither as a coding-first data-science manual nor as a mathematically advanced treatise on Bayesian theory, but as a transition text: a broad statistics textbook that guides readers from familiar inferential methods towards Bayesian modelling and computation. For its intended audience, that is a valuable and well-executed contribution.

Diego Andres Perez Ruiz: *diego.perezruiz@manchester.ac.uk*
Department of Social Statistics, School of Social Sciences, The University of Manchester, Manchester, UK

© 2026 International Statistical Institute.

17515823, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/insr.70060 by Ciara Wade - NICE, National Institute for Health and Care Excellence , Wiley Online Library on [27/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License