

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

| | |
|---|----------|
| Dedication | 3 |
| Preface | 1 |
| 1 Introduction to Bayesian Statistics | 5 |
| 1.1 Classical (Frequentist) versus Bayesian Statistics | 6 |
| 1.1.1 Frequentist Statistics: Probability Distributions for Observations, not Parameters | 6 |
| 1.1.2 Bayesian Statistics: Probability Distributions for Observations <i>and</i> Parameters | 7 |
| 1.1.3 Concepts of Probability: Frequentist and Subjective | 8 |
| 1.2 Probability Rules and Bayes' Theorem | 9 |
| 1.2.1 Probabilities and Conditional Probabilities of Events | 9 |
| 1.2.2 Example: Diagnostics for Disease Screening | 10 |
| 1.2.3 Bayes' Theorem and its Generalizations | 11 |
| 1.3 The Bayesian Approach to Statistical Inference | 13 |
| 1.3.1 Bayesian Prior and Posterior Distributions | 14 |
| 1.3.2 Prior Distribution and Likelihood Function Determine Posterior Dis- tribution | 15 |
| 1.3.3 Binomial Distribution and Its Likelihood Function | 15 |
| 1.3.4 Binomial Likelihood with Uniform Prior Induces Beta Posterior Dis- tribution | 16 |
| 1.3.5 Example: Proportion Supporting a Woman's Choice about Abortion | 18 |
| 1.3.6 A Criticism: Bayesian Inferences Are Subjective | 19 |
| 1.4 Bayesian Point Estimates | 19 |
| 1.4.1 Posterior Mean, Median, and Mode | 20 |
| 1.4.2 Shrinkage with a Bayesian Posterior Mean Estimate | 20 |
| 1.4.3 Decision-Theoretic Evaluation of Estimators: Loss and Risk Functions* 21 | |
| 1.4.4 Bayesian Risk: Decision-Theoretic Justification for Posterior Mean and Median* | 22 |
| 1.5 Bayesian Posterior Intervals | 23 |
| 1.5.1 Percentile Intervals and Highest Posterior Density (HPD) Intervals . | 23 |
| 1.5.2 Example: Proportion Supporting a Woman's Choice Revisited | 24 |
| 1.5.3 Interpretation: Bayesian Posterior Intervals versus Frequentist Confi- dence Intervals | 25 |
| 1.6 Bayesian Significance Testing and Prediction | 25 |
| 1.6.1 Significance Testing based on Posterior Probabilities for Parameter Regions | 26 |
| 1.6.2 Example: Proportion Supporting a Woman's Choice Revisited | 26 |
| 1.6.3 Predicting Future Observations: Bayesian Posterior Predictive Distri- bution | 27 |

| | | |
|----------|--|-----------|
| 1.7 | More on the Bayesian Approach | 28 |
| 1.7.1 | Non-informative Prior Distributions: Uniform and Improper | 28 |
| 1.7.2 | Likelihood Principle and Frequentist Violations of It* | 28 |
| 1.7.3 | Exchangeability and Conditional Independence of Observations* | 29 |
| 1.8 | Bayesian and Frequentist Statistics as Components of Data Science | 31 |
| 1.8.1 | Key Aspects of Frequentist and Bayesian Statistical Analyses | 32 |
| 1.8.2 | Some History: Bayes, Laplace, and Bayesian and Frequentist Statistics | 34 |
| 1.9 | Chapter Summary | 36 |
| 1.10 | Exercises | 37 |
| 2 | Bayesian Inference for Proportions | 43 |
| 2.1 | Prior Distributions for a Proportion | 43 |
| 2.1.1 | Beta Family of Prior Distributions | 43 |
| 2.1.2 | Jeffreys Prior Distribution | 45 |
| 2.1.3 | Conjugate Beta Prior and Posterior Distributions | 46 |
| 2.2 | Bayesian Point Estimates of a Proportion | 46 |
| 2.2.1 | Bayesian Posterior Mean Estimate of a Proportion | 46 |
| 2.2.2 | Example: Bayes Estimates with Three Prior Distributions and the Same Data | 47 |
| 2.2.3 | Bayesian Updating: Posterior Distribution Becomes Prior Distribution for Future Data | 50 |
| 2.2.4 | Bayesian Estimation and the Bias/Variance Tradeoff | 50 |
| 2.3 | Bayesian Posterior Intervals and Significance Tests for a Proportion | 51 |
| 2.3.1 | Review of Frequentist Inference for a Proportion | 52 |
| 2.3.2 | Posterior Intervals for a Proportion: Percentile and Highest Posterior Density | 53 |
| 2.3.3 | Influence of Sample Size and Prior Distribution on Posterior Intervals | 54 |
| 2.3.4 | Posterior Probability Analogs of P -values | 55 |
| 2.3.5 | Bayes Factors for Hypothesis Testing | 56 |
| 2.3.6 | Bayesian Prediction of Future Observations | 58 |
| 2.4 | Bayesian Inference for Comparing Proportions | 59 |
| 2.4.1 | Review of Frequentist Inference for Comparing Two Proportions | 59 |
| 2.4.2 | Posterior Estimation of Difference Using Beta Prior Distributions | 60 |
| 2.4.3 | Analogs of Significance Tests about Difference of Proportions | 61 |
| 2.4.4 | Multiple Comparisons of Several Binomial Parameters | 63 |
| 2.5 | Bayesian Inference for Categorical Variables with Several Categories* | 64 |
| 2.5.1 | The Multinomial Distribution | 64 |
| 2.5.2 | Conjugate Dirichlet Distributions for Multinomial Parameters | 65 |
| 2.5.3 | Example: What Percentage of People are Very Happy? | 65 |
| 2.5.4 | Frequentist Test of Independence of Two Categorical Variables | 66 |
| 2.5.5 | Bayes Factor for Evaluating Plausibility of Independence | 67 |
| 2.6 | Empirical Bayes Approach Estimates Hyperparameters* | 68 |
| 2.6.1 | Empirical Bayes Approach for Proportions | 69 |
| 2.6.2 | Example: Estimating Biden/Trump Presidential Election Results | 69 |
| 2.7 | Logit-Normal Prior Distributions for Proportions* | 72 |
| 2.7.1 | Normal Distribution for Logit Transformation of a Probability | 72 |
| 2.7.2 | Example: Estimating Biden/Trump Presidential Election Results | 74 |
| 2.7.3 | How to Select Prior Distributions for Proportions? | 76 |
| 2.8 | Chapter Summary | 77 |
| 2.9 | Exercises | 78 |

| | | |
|----------|--|------------|
| 3 | Bayesian Inference for Means | 87 |
| 3.1 | Bayesian Inference for a Normal Mean, Conditional on the Variance | 88 |
| 3.1.1 | Conjugate Normal Prior and Posterior Distributions | 88 |
| 3.1.2 | Shrinkage and Precision for the Posterior Distribution of the Mean | 89 |
| 3.1.3 | Parameterization of Variance of Prior Distribution with Imaginary Observations | 90 |
| 3.2 | Inference for a Normal Mean with Variance Unknown | 91 |
| 3.2.1 | Review of Frequentist Inference for Mean of a Normal Distribution | 91 |
| 3.2.2 | Bayesians Inferences Can Handle Nuisance Parameters | 92 |
| 3.2.3 | The Gamma and Inverse Gamma Distributions | 92 |
| 3.2.4 | Finding the Joint Posterior Distribution of the Mean and Variance | 94 |
| 3.2.5 | Marginal Posterior Distribution of Mean and the t Distribution | 95 |
| 3.2.6 | Example: Bayesian Analysis for Anorexia Therapy | 96 |
| 3.2.7 | Significance Testing Analogs for Means | 100 |
| 3.3 | The Monte Carlo Method for Approximating a Distribution | 100 |
| 3.4 | Inference for Means Using Improper Prior Distributions | 103 |
| 3.4.1 | Deriving Posterior Distributions with Improper Prior Distributions | 103 |
| 3.4.2 | Example: Anorexia Study Revisited | 104 |
| 3.4.3 | Future Predictions and Prediction Intervals | 106 |
| 3.5 | Bayesian Inference for Comparing Two Means | 107 |
| 3.5.1 | Review of Frequentist Inference for Comparing Two Means | 107 |
| 3.5.2 | Posterior Distribution of $\mu_1 - \mu_2$ for Conjugate and Improper Prior Distributions | 108 |
| 3.5.3 | Example: Comparing Mean Housework Hours for Men and Women | 110 |
| 3.5.4 | Significance Testing Analogs for Comparing Two Means | 114 |
| 3.5.5 | Bayesian Inference When Only Summary Statistics Are Available | 115 |
| 3.6 | Bayesian Inference for Multiple Means | 115 |
| 3.6.1 | Review of Frequentist Analysis of Variance for Comparing Means | 115 |
| 3.6.2 | Bayesian Modeling of Multiple Means | 116 |
| 3.6.3 | Example: Comparing Mean Holding Times on Phone for Service | 117 |
| 3.6.4 | Checking Assumptions | 119 |
| 3.7 | More Properties of Bayesian Inference* | 120 |
| 3.7.1 | The Exponential Family and Conjugate Prior Distributions | 120 |
| 3.7.2 | Poisson Model Has Conjugate Prior Gamma Distribution | 121 |
| 3.7.3 | Example: How Many Close Friends Do You Have? | 122 |
| 3.7.4 | Bayesian and Maximum Likelihood Large-Sample Performance | 124 |
| 3.7.5 | Robustness to Specification of Prior Distribution: Cauchy Distribution | 125 |
| 3.7.6 | Example: Anorexia Data with Cauchy Prior Distribution | 126 |
| 3.8 | Chapter Summary | 127 |
| 3.9 | Exercises | 128 |
| 4 | Bayesian Inference for Linear Models | 135 |
| 4.1 | The Linear Model and the Normal Linear Model | 135 |
| 4.1.1 | Least Squares Model Fitting | 136 |
| 4.1.2 | Correlation and Regression Toward the Mean | 137 |
| 4.1.3 | Multiple Regression: Linear Model with Multiple Explanatory Variables | 138 |
| 4.1.4 | Statistical Inference for Normal Linear Models | 139 |
| 4.1.5 | Example: Normal Linear Model for Mental Impairment | 140 |
| 4.2 | Bayesian Approach to Normal Linear Model | 142 |

| | | |
|----------|---|------------|
| 4.2.1 | Prior and Posterior Distributions for Normal Linear Models | 142 |
| 4.2.2 | Example: Bayesian Linear Model for Mental Impairment | 144 |
| 4.2.3 | Establishing Causality from an Association* | 146 |
| 4.3 | Extending the Bayesian Normal Linear Model | 147 |
| 4.3.1 | Categorical Explanatory Variables: Indicators for Categories | 148 |
| 4.3.2 | Example: Comparing Mean Fertility for Natives and Migrants | 149 |
| 4.3.3 | Permitting Interaction between Explanatory Variables | 151 |
| 4.3.4 | Example: Linear Model for Scottish Hill Races | 151 |
| 4.3.5 | Collinearity and its Effects | 154 |
| 4.4 | Model Building and Model Checking | 155 |
| 4.4.1 | The Bias/Variance Tradeoff and Model Complexity | 156 |
| 4.4.2 | Model Comparison: Penalized Likelihood Indices and Bayes Factors | 157 |
| 4.4.3 | Model Selection Strategies | 158 |
| 4.4.4 | Example: Modeling Selling Price of a Home | 159 |
| 4.4.5 | Cross-Validation and its Use in Model Checking | 161 |
| 4.4.6 | Posterior Predictive Distribution and Its Use for Model Checking | 162 |
| 4.5 | Regularization Methods of Model-Fitting* | 164 |
| 4.5.1 | The Lasso Method Shrinks Many Estimated Effects to 0 | 164 |
| 4.5.2 | Example: Predicting GPA with Student Survey Data | 165 |
| 4.6 | Linear Modeling Using Matrix Algebra* | 169 |
| 4.6.1 | The Linear Model Expressed with a Model Matrix | 169 |
| 4.6.2 | Properties of Least Squares Estimators of Linear Model Parameters | 170 |
| 4.6.3 | Multivariate Normal Distribution and Normal Linear Model | 171 |
| 4.6.4 | Bayesian Normal Linear Model: Variance Known | 171 |
| 4.6.5 | Bayesian Normal Linear Model: Variance Unknown | 172 |
| 4.7 | Chapter Summary | 173 |
| 4.8 | Exercises | 174 |
| 5 | Bayesian Inference for Generalized Linear Models | 181 |
| 5.1 | Introduction to Generalized Linear Models | 182 |
| 5.1.1 | GLMs for Normal, Binomial, and Poisson Response Variables | 183 |
| 5.1.2 | Transforming Data for Ordinary Linear Modeling or GLMs | 183 |
| 5.1.3 | Frequentist Fitting of Generalized Linear Models* | 184 |
| 5.1.4 | Bayesian Fitting of Generalized Linear Models | 185 |
| 5.2 | Generalized Linear Models for Continuous Response Variables | 186 |
| 5.2.1 | GLM Assuming a Gamma Distribution for the Response Variable | 186 |
| 5.2.2 | Example: Normal and Gamma GLMs for House Selling Prices | 187 |
| 5.2.3 | Modeling Nonlinear Relationships | 189 |
| 5.2.4 | Example: Exponential Regression for Covid-19 Cases over Time | 190 |
| 5.2.5 | Cauchy Distribution for a Continuous Response Variable* | 192 |
| 5.3 | Logistic Regression for Binary Data | 193 |
| 5.3.1 | Logistic Regression: Model Expressions and Interpretations | 194 |
| 5.3.2 | Bayesian Inference for Logistic Regression | 196 |
| 5.3.3 | Example: Modeling of Endometrial Cancer Risk Factors | 197 |
| 5.3.4 | Complete and Quasi-Complete Separation and Infinite ML Estimates | 197 |
| 5.3.5 | Example: Bayesian Modeling of Endometrial Cancer Risk Factors | 199 |
| 5.3.6 | Fitted Values and Posterior Intervals for Probabilities | 202 |
| 5.3.7 | Determining an Informative Prior Distribution for a GLM | 202 |
| 5.4 | Loglinear Models for Count Response Outcomes | 204 |
| 5.4.1 | Poisson Loglinear Models | 204 |
| 5.4.2 | Example: Poisson Modeling of Horseshoe Crab Satellite Counts | 205 |

| | | |
|----------|---|------------|
| 5.4.3 | Overdispersion in Poisson Modeling of Count Response Data | 207 |
| 5.4.4 | Negative Binomial Models for Count Response Data | 208 |
| 5.4.5 | Example: Negative Binomial Modeling of Horseshoe Crab Satellite Counts | 209 |
| 5.5 | Modeling Rates | 210 |
| 5.5.1 | Including an Offset in a Loglinear Model | 210 |
| 5.5.2 | Example: Lung Cancer Survival Counts | 210 |
| 5.6 | Hierarchical Bayesian Treatment of Prior Distributions* | 212 |
| 5.6.1 | Prior Distributions for Hyperparameters of Prior Distributions | 212 |
| 5.6.2 | A Bayesian Hierarchical Model for Multiple Means | 213 |
| 5.6.3 | Example: Comparing Mean Telephone Holding Times, Revisited | 215 |
| 5.7 | Multilevel Models* | 216 |
| 5.7.1 | Example: Modeling Performance of Students and their Schools | 217 |
| 5.7.2 | Example: Smoking Prevention and Cessation Study | 218 |
| 5.8 | Generalized Linear Mixed Models for Repeated Measurement* | 220 |
| 5.8.1 | Generalized Linear Mixed Models: GLMs with Cluster-Specific Ran- dom Effects | 221 |
| 5.8.2 | Example: Repeated Responses on Similar Survey Questions | 222 |
| 5.9 | Chapter Summary | 224 |
| 5.10 | Exercises | 225 |
| 6 | Bayesian MCMC Posterior Computation and Diagnostics | 233 |
| 6.1 | Monte Carlo Simulation of Markov Chains | 233 |
| 6.1.1 | Review of the Monte Carlo Method | 234 |
| 6.1.2 | Markov Chains and the Markov Property | 234 |
| 6.1.3 | MCMC: Combining Markov Chains with Monte Carlo | 236 |
| 6.2 | Markov Chain Monte Carlo Model-Fitting Diagnostics | 238 |
| 6.2.1 | Effective Sample Size | 239 |
| 6.2.2 | R-hat for Comparing Parallel Markov Chains | 242 |
| 6.2.3 | Examining Trace Plots | 242 |
| 6.2.4 | Autocorrelation | 243 |
| 6.3 | Metropolis and Metropolis–Hastings MCMC Algorithms* | 245 |
| 6.3.1 | The Metropolis Algorithm | 246 |
| 6.3.2 | Example: Metropolis Algorithm for Estimating Normal Mean with Known Variance | 246 |
| 6.3.3 | More Details about the Metropolis Algorithm | 248 |
| 6.3.4 | The Metropolis–Hastings Algorithm | 250 |
| 6.4 | Other MCMC Algorithms* | 252 |
| 6.4.1 | Gibbs Sampling | 252 |
| 6.4.2 | Hamiltonian Monte Carlo | 252 |
| 6.4.3 | Approximate Bayesian Computation | 254 |
| 6.4.4 | Yet Other Algorithms | 255 |
| 6.5 | Chapter Summary | 255 |
| 6.6 | Exercises | 256 |
| 7 | Choosing and Extending Bayesian Models | 259 |
| 7.1 | Bayesian Model Choice and Model Averaging | 260 |
| 7.1.1 | Bayes Factors for Model Choice | 261 |
| 7.1.2 | Example: Choosing Between Two Regression Models | 261 |
| 7.1.3 | Example: Comparing Non-Nested Models Using BF on House Price Data | 263 |

| | | |
|--|---|------------|
| 7.1.4 | Posterior Model Probabilities | 264 |
| 7.1.5 | Posterior Predictive Checks | 265 |
| 7.1.6 | Bayesian Model Averaging | 266 |
| 7.1.7 | Example: Combining Predictions from Regression Models | 266 |
| 7.2 | Bayesian Robustness | 269 |
| 7.2.1 | Example of Informal Sensitivity Analysis | 269 |
| 7.2.2 | Example of Local Robustness | 270 |
| 7.2.3 | Example of Global Robustness | 272 |
| 7.2.4 | Simulated Data Examples: Comparing Priors with Increasing Outliers | 273 |
| 7.2.5 | Real Data Example: Logistic Regression with Heavy-Tailed Priors | 276 |
| 7.3 | Classification Methods* | 278 |
| 7.3.1 | Logistic Regression and Classification | 279 |
| 7.3.2 | Naive Bayesian Classification | 279 |
| 7.3.3 | Example: Predicting Iris Flower Species | 280 |
| 7.3.4 | Classification Using Linear Discriminant Analysis | 282 |
| 7.3.5 | Other Classification Methods | 283 |
| 7.3.6 | Pros and Cons of Various Classification Methods | 285 |
| 7.4 | Network Models* | 286 |
| 7.4.1 | Basic Definitions and Notation | 287 |
| 7.4.2 | A First Model for Network Data: The Erdős–Rényi Random Graph | 289 |
| 7.4.3 | Stochastic Block Models | 291 |
| 7.4.4 | Other Network Models | 293 |
| 7.5 | Dealing with Missing Data* | 295 |
| 7.5.1 | Data Missing Completely at Random, at Random, or Not at Random | 296 |
| 7.5.2 | Multiple Imputation Using the Monte Carlo Method | 296 |
| 7.5.3 | Example: Mental Impairment Study with Missing Data | 297 |
| 7.6 | Chapter Summary | 299 |
| 7.7 | Exercises | 300 |
| Appendix A Using R for Bayesian Data Analysis | | 305 |
| A.0 | Introduction to R | 305 |
| A.1 | Chapter 1: Fundamentals for Bayesian Statistics Using R | 306 |
| A.1.1 | R Software Packages for Bayesian Data Analysis | 306 |
| A.1.2 | R Functions for Probability Distributions | 307 |
| A.1.3 | Posterior Intervals | 308 |
| A.2 | Chapter 2: R for Bayesian Inference for Proportions | 309 |
| A.2.1 | Bayes Factors | 309 |
| A.2.2 | The <code>brms</code> Package: Installation and First Steps | 310 |
| A.3 | Chapter 3: R for Bayesian Inference for Means | 312 |
| A.3.1 | Constructing Histograms | 312 |
| A.3.2 | Comparing Means with an Inverse Gamma Prior Distribution for the Standard Deviation | 313 |
| A.3.3 | Alternative Parameterization for Comparing Means | 314 |
| A.3.4 | Comparing Two Means Permitting Unequal Variances | 315 |
| A.3.5 | Example: Comparing Driver Reaction Times According to Cell Phone Use | 316 |
| A.4 | Chapter 4: R for Bayesian Inference for Linear Models | 319 |
| A.4.1 | Linear Models in <code>brms</code> | 319 |
| A.4.2 | Model Validation and Evaluating Residuals for Bayesian Models | 321 |
| A.4.3 | Visualization of Estimated Expected Values and of Predictions | 324 |
| A.4.4 | Bayesian Regularized Regression (a Bayesian Lasso)* | 326 |

| | | |
|---|---|------------|
| A.5 | Chapter 5: R for Bayesian Inference for Generalized Linear Models | 329 |
| A.5.1 | Fitting GLMs Using the <code>brms</code> Package | 329 |
| A.5.2 | Logistic-Regression Based Posterior Intervals for Probabilities | 329 |
| A.5.3 | Random Effects in a Hierarchical Model | 331 |
| A.6 | Chapter 6: R for Bayesian Posterior Computation and Diagnostics | 333 |
| A.6.1 | The Gibbs Sampler* | 334 |
| A.6.2 | Model-Fitting Diagnostics in <code>coda</code> | 335 |
| A.6.3 | Exporting <code>brms</code> Output | 338 |
| A.6.4 | Rank Trace Plots | 339 |
| A.7 | Chapter 7: R for Choosing and Extending Bayesian Models | 341 |
| A.7.1 | Posterior Predictive Distribution for Model Checking | 341 |
| A.7.2 | Evaluating Classification with Receiver Operating Characteristic (ROC) Curves | 342 |
| Appendix B Using Python for Bayesian Data Analysis | | 345 |
| B.0 | Basics of Python | 345 |
| B.0.1 | Python Installation and Preliminaries | 345 |
| B.1 | Chapter 1: Fundamentals for Bayesian Statistics Using Python | 347 |
| B.1.1 | Python Packages for Bayesian Data Analysis | 347 |
| B.1.2 | Quantiles and Probability Density Function of Beta Distribution | 348 |
| B.1.3 | Highest Posterior Density (HPD) Interval for a Binomial Parameter | 349 |
| B.2 | Chapter 2: Bayesian Inference for Proportions in Python | 350 |
| B.2.1 | Bayes Estimates for a Proportion and Several Prior Distributions | 350 |
| B.2.2 | Posterior Intervals for a Proportion | 350 |
| B.2.3 | Posterior Probability Analogs of P -values | 351 |
| B.2.4 | Bayesian Prediction of Future Observations | 351 |
| B.2.5 | Posterior Estimation for Difference of Proportions | 352 |
| B.2.6 | Analogs of Significance Tests about Difference of Proportions | 353 |
| B.2.7 | Multiple Comparisons of Binomial Parameters | 353 |
| B.2.8 | Dirichlet–Multinomial Model for Probability Vector | 353 |
| B.2.9 | Bayes Factor for Dependence in Contingency Tables | 354 |
| B.2.10 | Empirical Bayes Approach for Estimating Many Proportions | 355 |
| B.2.11 | Logit-Normal Prior Distributions for Proportions | 356 |
| B.3 | Chapter 3: Bayesian Inference for Means in Python | 357 |
| B.3.1 | Descriptive and Inferential Analysis for a Normal Mean | 357 |
| B.3.2 | Approximating a Posterior Distribution by Monte Carlo | 359 |
| B.3.3 | Bayesian Inference for a Mean Using Improper Prior Distributions | 361 |
| B.3.4 | Future Predictions and Prediction Intervals | 361 |
| B.3.5 | Comparing Two Means | 362 |
| B.3.6 | Comparing Means Permitting Unequal Variances | 365 |
| B.3.7 | Comparing Several Means | 367 |
| B.3.8 | Analyzing Means Assuming a Poisson Distribution | 368 |
| B.3.9 | Inference Using a Cauchy Prior Distribution for the Mean | 369 |
| B.4 | Chapter 4: Bayesian Inference for Linear Models in Python | 370 |
| B.4.1 | Frequentist Fitting of a Linear Model | 370 |
| B.4.2 | Bayesian Fitting of Normal Linear Model | 371 |
| B.4.3 | Including Categorical Explanatory Variables in a Linear Model | 372 |
| B.4.4 | Permitting Interaction in a Linear Model | 372 |
| B.4.5 | Collinearity and its Effects | 374 |
| B.4.6 | Model Building and Model Selection | 374 |
| B.4.7 | Regularization Fitting of a Linear Model with the Lasso | 375 |

| | | |
|--------|---|------------|
| B.5 | Chapter 5: Bayesian Inference for Generalized Linear Models in Python . . . | 377 |
| B.5.1 | Generalized Linear Models for Normal and Gamma Distributed Responses | 378 |
| B.5.2 | Models for an Exponential Relationship | 379 |
| B.5.3 | Cauchy Distribution for a Continuous Response Variable | 381 |
| B.5.4 | Separation and Infinite ML Estimates for Logistic Regression | 381 |
| B.5.5 | Bayesian Fitting of Logistic Regression Models | 382 |
| B.5.6 | Logistic Regression Fitted Values and Posterior Intervals for Probabilities | 383 |
| B.5.7 | Poisson Loglinear Modeling for Count Response Data | 384 |
| B.5.8 | GLMs Assuming a Negative Binomial Response Distribution | 384 |
| B.5.9 | Generalized Linear Modeling of Rates | 385 |
| B.5.10 | Hierarchical Bayesian Modeling of Several Means | 386 |
| B.5.11 | Multilevel Models and other Generalized Linear Mixed Models | 387 |
| B.6 | Chapter 6: Bayesian Posterior Computation and Diagnostics in Python . . . | 389 |
| B.6.1 | Managing MCMC Sampling and Producing Diagnostics | 389 |
| B.6.2 | Trace Plots | 391 |
| B.6.3 | Autocorrelation | 392 |
| B.6.4 | MCMC Algorithm for Estimating Normal Mean with Known Variance | 392 |
| B.6.5 | Tuning the Acceptance Probability of an MCMC Algorithm | 393 |
| B.7 | Chapter 7: Choosing and Extending Bayesian Models in Python | 394 |
| B.7.1 | Model Choice and Posterior Predictive Distributions | 394 |
| B.7.2 | Pseudo Bayesian Model Averaging | 396 |
| B.7.3 | Comparing Prior Distributions in Case of Outliers | 397 |
| B.7.4 | Logistic Regression with Heavy-Tailed Priors | 399 |
| B.7.5 | Classification | 400 |
| B.7.6 | Networks | 402 |
| B.7.7 | Multiple Imputation to Adjust for Missing Data | 402 |
| | Appendix C Solutions for exercises | 405 |
| | Bibliography | 425 |
| | Bibliography | 425 |
| | Example Index | 428 |
| | Subject Index | 430 |

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

