Statistics

Emphasizing the use of WinBUGS and R to analyze real data, **Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians** presents statistical tools to address scientific questions. It highlights foundational issues in statistics, the importance of making accurate predictions, and the need for scientists and statisticians to collaborate in analyzing data.

The first five chapters of the book contain core material that spans basic Bayesian ideas, calculations, and inference, including modeling one and two sample data from traditional sampling models. The text then covers Monte Carlo methods, such as Markov chain Monte Carlo (MCMC) simulation. After discussing linear structures in regression, it presents binomial regression, normal regression, analysis of variance, and Poisson regression, before extending these methods to handle correlated data. The authors also examine survival analysis and binary diagnostic testing. The last chapter on nonparametric inference explores density estimation and flexible regression modeling of mean functions.

Features

- Covers a large number of statistical models
- Emphasizes elicitation of real prior information
- Explores numerical approximations via simulation
- Uses WinBUGS and R for computations
- Includes numerous exercises and real-world examples
- Provides data, programming code, and other materials on the book's website

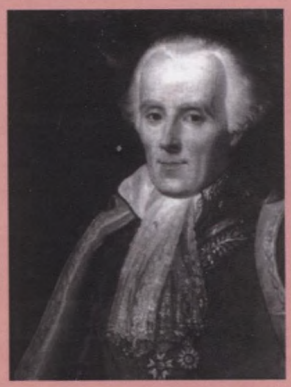
About the Authors

Ronald Christensen is Professor of Statistics at the University of New Mexico, Fellow of the American Statistical Association (ASA) and Institute of Mathematical Statistics, and former Chair of the ASA Section on Bayesian Statistical Science.

Wesley Johnson is Professor of Statistics at the University of California, Irvine, Fellow of the ASA, and Chair-Elect of the ASA Section on Bayesian Statistical Science.

Adam Branscum is Associate Professor of Public Health at Oregon State University.

Timothy E. Hanson is Associate Professor of Statistics at the University of South Carolina.



Pierre-Simon Laplace

CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K10199

ISBN: 978-1-4398-0354-7
90000
9 781439 803547

Preface xv

1 Prologue 1
1.1 Probability of a Defective: Binomial Data 2
1.2 Brass Alloy Zinc Content: Normal Data 3
1.3 Armadillo Hunting: Poisson Data 4
1.4 Abortion in Dairy Cattle: Survival Data 5
1.5 Ache Hunting with Age Trends 6
1.6 Lung Cancer Treatment: Log-Normal Regression 7
1.7 Survival with Random Effects: Ache Hunting 8

2 Fundamental Ideas I 13
2.1 Simple Probability Computations 14
2.2 Science, Priors, and Prediction 18
2.3 Statistical Models 22
2.4 Posterior Analysis 30
2.5 Commonly Used Distributions 33

3 Integration Versus Simulation 37
3.1 Introduction 37
3.2 WinBUGS I: Getting Started 41
3.3 Method of Composition 48
3.4 Monte Carlo Integration* 49
3.5 Posterior Computations in R 51

4 Fundamental Ideas II 53
4.1 Statistical Testing 53
4.1.1 Checking Bayesian Models 57
4.1.2 Predictive *P*-Values 59
4.1.3 Lindley-Jeffreys Paradox 60
4.2 Exchangeability 61
4.3 Likelihood Functions 63
4.4 Sufficient Statistics 66
4.5 Analysis Using Predictive Distributions 67
4.6 Flat Priors 69
4.6.1 Data Translated Likelihoods 71
4.7 Jeffreys' Priors 72
4.7.1 Multiple Parameter Jeffreys' Prior* 73
4.8 Bayes Factors* 74
4.8.1 General Parametric Testing 74
4.8.2 Nested Models 75
4.8.3 Simulating Bayes Factors 75

4.9 Other Model Selection Criteria 78
4.9.1 Bayesian Information Criterion 79
4.9.2 LPML 81
4.9.3 Deviance Information Criterion 82
4.9.4 Final Comments 83
4.10 Normal Approximations to Posteriors* 84
4.11 Bayesian Consistency and Inconsistency 88
4.12 Hierarchical Models 89
4.13 Some Final Comments on Likelihoods* 93
4.14 Identifiability and Noninformative Data 94

5 Comparing Populations 97
5.1 Inference for Proportions 97
5.1.1 Prior Distributions 99
5.1.1.1 Reference Priors 99
5.1.1.2 Informative Beta Priors 99
5.1.1.3 Rare Events 100
5.1.1.4 Non-Beta Priors 102
5.1.2 Effect Measures 103
5.1.3 Independent Binomials 105
5.1.4 Case-Control Sampling 107
5.2 Inference for Normal Populations 111
5.2.1 Reference Priors 111
5.2.2 Conjugate Priors 114
5.2.3 Independence Priors 115
5.2.4 Some Curious Distributional Results* 120
5.2.5 Two-Sample Normal Model 121
5.3 Inference for Rates 128
5.3.1 One-Sample Poisson Data 129
5.3.2 Informative Priors 131
5.3.3 Reference Priors 133
5.3.4 Two-Sample Poisson Data 134
5.4 Sample Size Determination* 136

6 Simulations 139
6.1 Generating Random Samples 139
6.2 Traditional Monte Carlo Methods 142
6.2.1 Acceptance-Rejection Sampling 142
6.2.2 Importance Sampling 143
6.3 Markov Chain Monte Carlo 145
6.3.1 Markov Chains 147
6.3.2 Gibbs Sampling 150
6.3.2.1 Proof that $p(\theta)$ is the Stationary Distribution in the Two-Block Case* 154
6.3.3 Metropolis Algorithm 154
6.3.3.1 Proof that $p(\theta)$ is the Stationary Distribution* 156
6.3.4 Slice Sampling 158
6.3.5 Checking MCMC Samples 159

7 Basic Concepts of Regression **161**
7.1 Introduction 161
7.2 Data Notation and Format 162
7.3 Predictive Models: An Overview 164
7.4 Modeling with Linear Structures 166
7.4.1 Continuous Predictors 166
7.4.2 Binary Predictors 166
7.4.3 Multi-Category Predictors 167
7.4.4 Predictor Selection 169
7.4.5 Several Categorical Covariates 171
7.4.6 Confounding 172
7.4.7 Effect Modification/Interaction 174
7.4.7.1 Two Categorical Predictors 175
7.4.7.2 One Continuous and One Categorical Predictor 176
7.4.7.3 Two Continuous Predictors 177
7.5 Illustration: FEV Data 178

8 Binomial Regression **181**
8.1 The Sampling Model 181
8.2 Binomial Regression Analysis 186
8.2.1 Predictive Probabilities 188
8.2.2 Inference for Regression Coefficients 190
8.2.3 Inference for LD_α 195
8.3 Model Checking 195
8.3.1 Box's Method 195
8.3.2 Link Selection 196
8.4 Prior Distributions 197
8.4.1 Simple Regression 197
8.4.2 General Regression 199
8.4.2.1 Prior Elicitation 203
8.4.2.2 Data Augmentation Priors 203
8.4.2.3 Standardized Variables 204
8.4.3 Reference Priors 207
8.4.4 Partial Prior Information 209
8.4.5 Partial Priors: Theoretical Considerations* 212
8.5 Mixed Models 213
8.5.1 Prior Elicitation 217
8.5.2 Mixed Model Likelihood 219
8.5.3 Gibbs Sampling and Centering* 219

9 Linear Regression **223**
9.1 The Sampling Model 223
9.2 Reference Priors 226
9.2.1 Least Squares Estimation 227
9.2.2 Posterior Analysis 229
9.2.3 A Proper Reference Prior 230
9.3 Conjugate Priors 231
9.4 Independence Priors 233
9.4.1 Prior on β 234
9.4.2 Prior on τ 236
9.4.3 Partial Prior Information 237
9.4.4 Inference and Displays 238

9.4.5 Gibbs Sampling* 239
9.4.6 WinBUGS and R Code 241
9.5 ANOVA 243
9.5.1 Independence Prior 243
9.5.1.1 Allocation and Diagnosis 248
9.5.2 Hierarchical Priors and Models 251
9.6 Model Diagnostics 252
9.7 Model Selection 257
9.8 Nonlinear Regression* 259

10 Correlated Data **263**
10.1 Introduction 263
10.2 Mixed Models 265
10.2.1 Random Intercept Model 267
10.2.2 Random Slopes and Random Intercepts 275
10.3 Multivariate Normal Models 278
10.3.1 Parameterized Covariance Matrices 279
10.3.1.1 Analytic Formulas for CS and AR(1) Precision Matrices 283
10.4 Multivariate Normal Regression 283
10.5 Posterior Sampling and Missing Data 285

11 Count Data **287**
11.1 Poisson Regression 287
11.1.1 Poisson Regression for Rates 289
11.2 Over-Dispersion and Mixtures of Poissons 294
11.2.1 Zero-Inflated Poisson Data 298
11.2.2 SAS Analysis of Foot-and-Mouth Disease Data 298
11.3 Longitudinal Data 300

12 Time to Event Data **301**
12.1 Introduction 301
12.1.1 Survival and Hazard Functions 302
12.1.2 Censoring 303
12.1.3 The Likelihood 304
12.2 One-Sample Models 305
12.2.1 Distributional Models 306
12.2.2 Posterior Analysis 307
12.2.3 Log-Normal Data 307
12.2.4 Exponential Data 307
12.2.5 WinBUGS for Censored Data 308
12.2.6 Weibull Data 309
12.2.7 Prediction 311
12.2.8 Interval Censoring 312
12.3 Two-Sample Data 314
12.3.1 Two-Sample Exponential Model 314
12.3.2 Two-Sample Weibull Model 319
12.3.3 Two-Sample Log-Normal Model 320
12.4 Plotting Survival and Hazard Functions 322

13 Time to Event Regression 325
13.1 Accelerated Failure Time Models 325
13.1.1 Abortion Data 333
13.1.2 Prior Elicitation for AFTs 334
13.1.2.1 Specifying the Marginal Prior for β 335
13.1.2.2 Partial Prior Information for β 338
13.1.2.3 Uncertainty About τ 339
13.1.3 Case Deletion Diagnostics for AFT Models 340
13.1.3.1 Predictive Influence 342
13.1.4 Bayes Factor Model Selection 343
13.1.5 Sensitivity Analysis 343
13.1.6 Final Comments 344
13.2 Proportional Hazards Modeling 345
13.2.1 The Proportional Hazards (PH) Model 345
13.2.2 A Baseline Hazard Model 347
13.2.3 The Likelihood 347
13.2.3.1 Noninformative Data* 349
13.2.4 Priors for β 349
13.2.5 Priors for λ 351
13.2.6 Our Data Model 352
13.2.7 WinBUGS Code 353
13.2.8 Posterior Analysis for Leukemia Data 355
13.2.9 SAS Analysis of Leukemia Data 356
13.2.10 Another Example 358
13.3 Survival with Random Effects 363

14 Binary Diagnostic Tests 365
14.1 Basic Ideas 366
14.2 One Test, One Population 368
14.2.1 Gold-Standard Data 369
14.2.2 No Gold-Standard Data 371
14.3 Two Tests, Two Populations 374
14.3.1 Methods for Conditionally Independent Tests 374
14.4 Prevalence Distributions 379

15 Nonparametric Models 385
15.1 Flexible Density Shapes 386
15.1.1 Finite Mixtures 386
15.1.1.1 Identifiability Issues* 391
15.1.2 Dirichlet Process Mixtures: Infinite Mixtures 392
15.1.3 Mixtures of Polya Trees 396
15.2 Flexible Regression Functions 402
15.3 Proportional Hazards Modeling 414

Appendix A: Matrices and Vectors 419
A.1 Matrix Addition and Subtraction 420
A.2 Scalar Multiplication 420
A.3 Matrix Multiplication 420
A.4 Special Matrices 422
A.5 Linear Dependence and Rank 423
A.6 Inverse Matrices 424
A.7 A List of Useful Properties 426

A.8 Eigenvalues and Eigenvectors 426
A.9 Properties of Determinants 428
A.10 Calculus and Taylor's Theorem 428
A.11 Partitioned Matrices 428

Appendix B: Probability **431**
B.1 Univariate Probability 431
B.2 Multivariate Probability 432
B.2.1 Joint Distribution of Two Vectors 434
B.2.2 Conditional Distributions 434
B.2.3 Independence 436
B.2.4 Moment Generating Functions 437
B.2.5 Change of Variables 437
B.3 Models and Conditional Independence 438

Appendix C: Getting Started in R **443**
C.1 Getting R 443
C.2 Some R Basics 443
C.3 User-Contributed Packages 446
C.4 Reading Data 447
C.5 Graphing 447
C.6 Interface Between R and WinBUGS 456
C.7 Writing New R Functions 456

References **459**

Author Index **467**

Subject Index **473**