Contents

Preface xiii

Part I: Fundamentals of Bayesian Inference 1

1 Probability and inference 3
1.1 The three steps of Bayesian data analysis 3
1.2 General notation for statistical inference 4
1.3 Bayesian inference 6
1.4 Discrete probability examples: genetics and spell checking 8
1.5 Probability as a measure of uncertainty 11
1.6 Example of probability assignment: football point spreads 13
1.7 Example: estimating the accuracy of record linkage 16
1.8 Some useful results from probability theory 19
1.9 Computation and software 22
1.10 Bayesian inference in applied statistics 24
1.11 Bibliographic note 25
1.12 Exercises 27

2 Single-parameter models 29
2.1 Estimating a probability from binomial data 29
2.2 Posterior as compromise between data and prior information 32
2.3 Summarizing posterior inference 32
2.4 Informative prior distributions 34
2.5 Estimating a normal mean with known variance 39
2.6 Other standard single-parameter models 42
2.7 Example: informative prior distribution for cancer rates 47
2.8 Noninformative prior distributions 51
2.9 Weakly informative prior distributions 55
2.10 Bibliographic note 56
2.11 Exercises 57

3 Introduction to multiparameter models 63
3.1 Averaging over 'nuisance parameters' 63
3.2 Normal data with a noninformative prior distribution 64
3.3 Normal data with a conjugate prior distribution 67
3.4 Multinomial model for categorical data 69
3.5 Multivariate normal model with known variance 70
3.6 Multivariate normal with unknown mean and variance 72
3.7 Example: analysis of a bioassay experiment 74
3.8 Summary of elementary modeling and computation 78
3.9 Bibliographic note 78
3.10 Exercises 79

4 Asymptotics and connections to non-Bayesian approaches **83**
4.1 Normal approximations to the posterior distribution 83
4.2 Large-sample theory 87
4.3 Counterexamples to the theorems 89
4.4 Frequency evaluations of Bayesian inferences 91
4.5 Bayesian interpretations of other statistical methods 92
4.6 Bibliographic note 97
4.7 Exercises 98

5 Hierarchical models **101**
5.1 Constructing a parameterized prior distribution 102
5.2 Exchangeability and setting up hierarchical models 104
5.3 Fully Bayesian analysis of conjugate hierarchical models 108
5.4 Estimating exchangeable parameters from a normal model 113
5.5 Example: parallel experiments in eight schools 119
5.6 Hierarchical modeling applied to a meta-analysis 124
5.7 Weakly informative priors for hierarchical variance parameters 128
5.8 Bibliographic note 132
5.9 Exercises 134

Part II: Fundamentals of Bayesian Data Analysis **139**

6 Model checking **141**
6.1 The place of model checking in applied Bayesian statistics 141
6.2 Do the inferences from the model make sense? 142
6.3 Posterior predictive checking 143
6.4 Graphical posterior predictive checks 153
6.5 Model checking for the educational testing example 159
6.6 Bibliographic note 161
6.7 Exercises 163

7 Evaluating, comparing, and expanding models **165**
7.1 Measures of predictive accuracy 166
7.2 Information criteria and cross-validation 169
7.3 Model comparison based on predictive performance 178
7.4 Model comparison using Bayes factors 182
7.5 Continuous model expansion 184
7.6 Implicit assumptions and model expansion: an example 187
7.7 Bibliographic note 192
7.8 Exercises 193

8 Modeling accounting for data collection **197**
8.1 Bayesian inference requires a model for data collection 197
8.2 Data-collection models and ignorability 199
8.3 Sample surveys 205
8.4 Designed experiments 214
8.5 Sensitivity and the role of randomization 218
8.6 Observational studies 220
8.7 Censoring and truncation 224
8.8 Discussion 229
8.9 Bibliographic note 229
8.10 Exercises 230

9 **Decision analysis** **237**
9.1 Bayesian decision theory in different contexts 237
9.2 Using regression predictions: incentives for telephone surveys 239
9.3 Multistage decision making: medical screening 245
9.4 Hierarchical decision analysis for radon measurement 246
9.5 Personal vs. institutional decision analysis 256
9.6 Bibliographic note 257
9.7 Exercises 257

Part III: Advanced Computation **259**

10 **Introduction to Bayesian computation** **261**
10.1 Numerical integration 261
10.2 Distributional approximations 262
10.3 Direct simulation and rejection sampling 263
10.4 Importance sampling 265
10.5 How many simulation draws are needed? 267
10.6 Computing environments 268
10.7 Debugging Bayesian computing 270
10.8 Bibliographic note 271
10.9 Exercises 272

11 **Basics of Markov chain simulation** **275**
11.1 Gibbs sampler 276
11.2 Metropolis and Metropolis-Hastings algorithms 278
11.3 Using Gibbs and Metropolis as building blocks 280
11.4 Inference and assessing convergence 281
11.5 Effective number of simulation draws 286
11.6 Example: hierarchical normal model 288
11.7 Bibliographic note 291
11.8 Exercises 291

12 **Computationally efficient Markov chain simulation** **293**
12.1 Efficient Gibbs samplers 293
12.2 Efficient Metropolis jumping rules 295
12.3 Further extensions to Gibbs and Metropolis 297
12.4 Hamiltonian Monte Carlo 300
12.5 Hamiltonian dynamics for a simple hierarchical model 305
12.6 Stan: developing a computing environment 307
12.7 Bibliographic note 308
12.8 Exercises 309

13 **Modal and distributional approximations** **311**
13.1 Finding posterior modes 311
13.2 Boundary-avoiding priors for modal summaries 313
13.3 Normal and related mixture approximations 318
13.4 Finding marginal posterior modes using EM 320
13.5 Approximating conditional and marginal posterior densities 325
13.6 Example: hierarchical normal model (continued) 326
13.7 Variational inference 331
13.8 Expectation propagation 338
13.9 Other approximations 343

13.10 Unknown normalizing factors 345
13.11 Bibliographic note 348
13.12 Exercises 349

Part IV: Regression Models **351**

14 Introduction to regression models **353**
14.1 Conditional modeling 353
14.2 Bayesian analysis of the classical regression model 354
14.3 Regression for causal inference: incumbency in congressional elections 358
14.4 Goals of regression analysis 364
14.5 Assembling the matrix of explanatory variables 365
14.6 Regularization and dimension reduction for multiple predictors 367
14.7 Unequal variances and correlations 369
14.8 Including numerical prior information 376
14.9 Bibliographic note 378
14.10 Exercises 378

15 Hierarchical linear models **381**
15.1 Regression coefficients exchangeable in batches 382
15.2 Example: forecasting U.S. presidential elections 383
15.3 Interpreting a normal prior distribution as additional data 388
15.4 Varying intercepts and slopes 390
15.5 Computation: batching and transformation 392
15.6 Analysis of variance and the batching of coefficients 395
15.7 Hierarchical models for batches of variance components 398
15.8 Bibliographic note 400
15.9 Exercises 402

16 Generalized linear models **405**
16.1 Standard generalized linear model likelihoods 406
16.2 Working with generalized linear models 407
16.3 Weakly informative priors for logistic regression 412
16.4 Example: hierarchical Poisson regression for police stops 420
16.5 Example: hierarchical logistic regression for political opinions 422
16.6 Models for multivariate and multinomial responses 423
16.7 Loglinear models for multivariate discrete data 428
16.8 Bibliographic note 431
16.9 Exercises 432

17 Models for robust inference **435**
17.1 Aspects of robustness 435
17.2 Overdispersed versions of standard probability models 437
17.3 Posterior inference and computation 439
17.4 Robust inference and sensitivity analysis for the eight schools 441
17.5 Robust regression using t-distributed errors 444
17.6 Bibliographic note 445
17.7 Exercises 446

18 Models for missing data **449**
18.1 Notation 449
18.2 Multiple imputation 451
18.3 Missing data in the multivariate normal and t models 454
18.4 Example: multiple imputation for a series of polls 456
18.5 Missing values with counted data 462
18.6 Example: an opinion poll in Slovenia 463
18.7 Bibliographic note 466
18.8 Exercises 467

Part V: Nonlinear and Nonparametric Models **469**

19 Parametric nonlinear models **471**
19.1 Example: serial dilution assay 471
19.2 Example: population toxicokinetics 477
19.3 Bibliographic note 485
19.4 Exercises 486

20 Basis function models **487**
20.1 Splines and weighted sums of basis functions 487
20.2 Basis selection and shrinkage of coefficients 490
20.3 Non-normal models and multivariate regression surfaces 494
20.4 Bibliographic note 498
20.5 Exercises 498

21 Gaussian process models **501**
21.1 Gaussian process regression 501
21.2 Example: birthdays and birthdates 505
21.3 Latent Gaussian process models 510
21.4 Functional data analysis 512
21.5 Density estimation and regression 513
21.6 Bibliographic note 516
21.7 Exercises 516

22 Finite mixture models **519**
22.1 Setting up and interpreting mixture models 519
22.2 Example: reaction times and schizophrenia 524
22.3 Label switching and posterior computation 533
22.4 Unspecified number of mixture components 536
22.5 Mixture models for classification and regression 539
22.6 Bibliographic note 542
22.7 Exercises 543

23 Dirichlet process models **545**
23.1 Bayesian histograms 545
23.2 Dirichlet process prior distributions 546
23.3 Dirichlet process mixtures 549
23.4 Beyond density estimation 557
23.5 Hierarchical dependence 560
23.6 Density regression 568
23.7 Bibliographic note 571
23.8 Exercises 573

A Standard probability distributions **575**
A.1 Continuous distributions 575
A.2 Discrete distributions 583
A.3 Bibliographic note 584

B Outline of proofs of limit theorems **585**
B.1 Bibliographic note 588

C Computation in R and Stan **589**
C.1 Getting started with R and Stan 589
C.2 Fitting a hierarchical model in Stan 589
C.3 Direct simulation, Gibbs, and Metropolis in R 594
C.4 Programming Hamiltonian Monte Carlo in R 601
C.5 Further comments on computation 605
C.6 Bibliographic note 606

References **607**

Author Index **641**

Subject Index **649**