

Contents

| | |
|--|-----------------|
| | 1 |
| | 2 |
| | 3 |
| | 4 |
| | 5 |
| | 6 |
| | 7 |
| | 8 |
| | 9 |
| | 10 |
| | 11 |
| | 12 |
| | 13 |
| | 14 |
| | 15 |
| List of Figures | xiii 16 |
| | 17 |
| List of Tables | xix 18 |
| | 19 |
| Preface | xxi 20 |
| | 21 |
| Acknowledgments | xxv 22 |
| | 23 |
| Introduction | xxvii 24 |
| | 25 |
| | 26 |
| Part I Introducing Bayesian Analysis | 1 27 |
| | 28 |
| 1 The foundations of Bayesian inference | 3 29 |
| 1.1 What is probability? | 3 30 |
| 1.1.1 Probability in classical statistics | 4 31 |
| 1.1.2 Subjective probability | 5 32 |
| 1.2 Subjective probability in Bayesian statistics | 7 33 |
| 1.3 Bayes theorem, discrete case | 8 34 |
| 1.4 Bayes theorem, continuous parameter | 13 35 |
| 1.4.1 Conjugate priors | 15 36 |
| 1.4.2 Bayesian updating with irregular priors | 16 37 |
| 1.4.3 Cromwell's Rule | 18 38 |
| 1.4.4 Bayesian updating as information accumulation | 19 39 |
| 1.5 Parameters as random variables, beliefs as distributions | 21 40 |
| 1.6 Communicating the results of a Bayesian analysis | 22 41 |
| 1.6.1 Bayesian point estimation | 23 42 |
| 1.6.2 Credible regions | 26 43 |
| 1.7 Asymptotic properties of posterior distributions | 29 44 |
| 1.8 Bayesian hypothesis testing | 31 45 |
| 1.8.1 Model choice | 36 46 |
| 1.8.2 Bayes factors | 37 47 |

viii CONTENTS

| | | | |
|----------------|--|------------|-----------|
| 1.9 | From subjective beliefs to parameters and models | 38 | 1 |
| 1.9.1 | Exchangeability | 39 | 2 |
| 1.9.2 | Implications and extensions of de Finetti's Representation Theorem | 42 | 4 |
| 1.9.3 | Finite exchangeability | 43 | 5 |
| 1.9.4 | Exchangeability and prediction | 43 | 6 |
| 1.9.5 | Conditional exchangeability and multiparameter models | 44 | 7 |
| 1.9.6 | Exchangeability of parameters: hierarchical modeling | 45 | 8 |
| 1.10 | Historical note | 46 | 9 |
| | | | 10 |
| 2 | Getting started: Bayesian analysis for simple models | 49 | 11 |
| 2.1 | Learning about probabilities, rates and proportions | 49 | 12 |
| 2.1.1 | Conjugate priors for probabilities, rates and proportions | 51 | 13 |
| 2.1.2 | Bayes estimates as weighted averages of priors and data | 58 | 14 |
| 2.1.3 | Parameterizations and priors | 61 | 15 |
| 2.1.4 | The variance of the posterior density | 64 | 16 |
| 2.2 | Associations between binary variables | 67 | 17 |
| 2.3 | Learning from counts | 73 | 18 |
| 2.3.1 | Predictive inference with count data | 78 | 19 |
| 2.4 | Learning about a normal mean and variance | 80 | 20 |
| 2.4.1 | Variance known | 80 | 21 |
| 2.4.2 | Mean and variance unknown | 83 | 22 |
| 2.4.3 | Conditionally conjugate prior | 92 | 23 |
| 2.4.4 | An improper, reference prior | 93 | 24 |
| 2.4.5 | Conflict between likelihood and prior | 98 | 25 |
| 2.4.6 | Non-conjugate priors | 98 | 26 |
| 2.5 | Regression models | 99 | 27 |
| 2.5.1 | Bayesian regression analysis | 102 | 28 |
| 2.5.2 | Likelihood function | 103 | 29 |
| 2.5.3 | Conjugate prior | 104 | 30 |
| 2.5.4 | Improper, reference prior | 107 | 31 |
| 2.6 | Further reading | 124 | 32 |
| | | | 33 |
| | | | 34 |
| Part II | Simulation Based Bayesian Analysis | 129 | 35 |
| | | | 36 |
| 3 | Monte Carlo methods | 133 | 37 |
| 3.1 | Simulation consistency | 134 | 38 |
| 3.2 | Inference for functions of parameters | 140 | 39 |
| 3.3 | Marginalization via Monte Carlo integration | 142 | 40 |
| 3.4 | Sampling algorithms | 153 | 41 |
| 3.4.1 | Inverse-CDF method | 153 | 42 |
| 3.4.2 | Importance sampling | 156 | 43 |
| | | | 44 |
| | | | 45 |
| | | | 46 |
| | | | 47 |

CONTENTS ix

| | | | |
|----------|---|------------|-----------|
| 3.4.3 | Accept-reject sampling | 159 | 1 |
| 3.4.4 | Adaptive rejection sampling | 163 | 2 |
| 3.5 | Further reading | 167 | 3 |
| | | | 4 |
| 4 | Markov chains | 171 | 5 |
| 4.1 | Notation and definitions | 172 | 6 |
| 4.1.1 | State space | 173 | 7 |
| 4.1.2 | Transition kernel | 173 | 8 |
| 4.2 | Properties of Markov chains | 176 | 9 |
| 4.2.1 | Existence of a stationary distribution, discrete case | 177 | 10 |
| 4.2.2 | Existence of a stationary distribution, continuous case | 178 | 11 |
| 4.2.3 | Irreducibility | 179 | 12 |
| 4.2.4 | Recurrence | 182 | 13 |
| 4.2.5 | Invariant measure | 184 | 14 |
| 4.2.6 | Reversibility | 185 | 15 |
| 4.2.7 | Aperiodicity | 186 | 16 |
| 4.3 | Convergence of Markov chains | 187 | 17 |
| 4.3.1 | Speed of convergence | 189 | 18 |
| 4.4 | Limit theorems for Markov chains | 191 | 19 |
| 4.4.1 | Simulation inefficiency | 191 | 20 |
| 4.4.2 | Central limit theorems for Markov chains | 195 | 21 |
| 4.5 | Further reading | 196 | 22 |
| | | | 23 |
| 5 | Markov chain Monte Carlo | 201 | 24 |
| 5.1 | Metropolis-Hastings algorithm | 201 | 25 |
| 5.1.1 | Theory for the Metropolis-Hastings algorithm | 202 | 26 |
| 5.1.2 | Choosing the proposal density | 204 | 27 |
| 5.2 | Gibbs sampling | 214 | 28 |
| 5.2.1 | Theory for the Gibbs sampler | 218 | 29 |
| 5.2.2 | Connection to the Metropolis algorithm | 221 | 30 |
| 5.2.3 | Deriving conditional densities for the Gibbs sampler: statistical models as conditional independence graphs | 225 | 32 |
| 5.2.4 | Pathologies | 229 | 33 |
| 5.2.5 | Data augmentation | 236 | 34 |
| 5.2.6 | Missing data problems | 237 | 35 |
| 5.2.7 | The slice sampler | 244 | 36 |
| | | | 37 |
| 6 | Implementing Markov chain Monte Carlo | 251 | 38 |
| 6.1 | Software for Markov chain Monte Carlo | 251 | 39 |
| 6.2 | Assessing convergence and run-length | 252 | 40 |
| 6.3 | Working with BUGS/JAGS from R | 256 | 41 |
| 6.4 | Tricks of the trade | 261 | 42 |
| 6.4.1 | Thinning | 261 | 43 |
| | | | 44 |
| | | | 45 |
| | | | 46 |
| | | | 47 |



x CONTENTS

| | | | |
|-----------------|---|------------|-----------|
| 6.4.2 | Blocking | 264 | 1 |
| 6.4.3 | Reparameterization | 270 | 2 |
| 6.5 | Other examples | 272 | 3 |
| 6.6 | Further reading | 292 | 4 |
| | | | 5 |
| | | | 6 |
| Part III | Advanced Applications in the Social Sciences | 299 | 7 |
| | | | 8 |
| 7 | Hierarchical Statistical Models | 301 | 9 |
| 7.1 | Data and parameters that vary by groups: the case for hierarchical modeling | 301 | 10 |
| | | | 11 |
| 7.1.1 | Exchangeable parameters generate hierarchical models | 305 | 12 |
| 7.1.2 | 'Borrowing strength' via exchangeability | 307 | 13 |
| 7.1.3 | Hierarchical modeling as a 'semi-pooling' estimator | 307 | 14 |
| 7.1.4 | Hierarchical modeling as a 'shrinkage' estimator | 308 | 15 |
| 7.1.5 | Computation via Markov chain Monte Carlo | 310 | 16 |
| 7.2 | ANOVA as a hierarchical model | 317 | 17 |
| 7.2.1 | One-way analysis of variance | 317 | 18 |
| 7.2.2 | Two-way ANOVA | 329 | 19 |
| 7.3 | Hierarchical models for longitudinal data | 345 | 20 |
| 7.4 | Hierarchical models for non-normal data | 354 | 21 |
| 7.5 | Multi-level models | 362 | 22 |
| | | | 23 |
| 8 | Bayesian analysis of choice making | 379 | 24 |
| 8.1 | Regression models for binary responses | 379 | 25 |
| 8.1.1 | Probit model via data augmentation | 380 | 26 |
| 8.1.2 | Probit model via marginal data augmentation | 389 | 27 |
| 8.1.3 | Logit model | 393 | 28 |
| 8.1.4 | Binomial model for grouped binary data | 395 | 29 |
| 8.2 | Ordered outcomes | 397 | 30 |
| 8.2.1 | Identification | 399 | 31 |
| 8.3 | Multinomial outcomes | 415 | 32 |
| 8.3.1 | Multinomial logit | 415 | 33 |
| 8.3.2 | Independence of irrelevant alternatives | 423 | 34 |
| 8.4 | Multinomial probit | 424 | 35 |
| 8.4.1 | Bayesian analysis via MCMC | 426 | 36 |
| | | | 37 |
| 9 | Bayesian approaches to measurement | 435 | 38 |
| 9.1 | Bayesian inference for latent states | 435 | 39 |
| 9.1.1 | A formal role for prior information | 436 | 40 |
| 9.1.2 | Inference for many parameters | 436 | 41 |
| 9.2 | Factor analysis | 438 | 42 |
| 9.2.1 | Likelihood and prior densities | 439 | 43 |
| 9.2.2 | Identification | 440 | 44 |
| 9.2.3 | Posterior density | 442 | 45 |
| 9.2.4 | Inference over rank orderings of the latent variable | 448 | 46 |
| 9.2.5 | Incorporating additional information via hierarchical modeling | 449 | 47 |



CONTENTS xi

| | | | |
|---------------------|--|------------|-----------|
| 9.3 | Item-response models | 454 | 1 |
| 9.4 | Dynamic measurement models | 471 | 2 |
| 9.4.1 | State-space models for ‘pooling the polls’ | 473 | 3 |
| 9.4.2 | Bayesian inference | 474 | 4 |
| | | | 5 |
| | | | 6 |
| Part IV | Appendices | 489 | 7 |
| | | | 8 |
| Appendix A: | Working with vectors and matrices | 491 | 9 |
| | | | 10 |
| Appendix B: | Probability review | 497 | 11 |
| B.1 | Foundations of probability | 497 | 12 |
| B.2 | Probability densities and mass functions | 498 | 13 |
| B.2.1 | Probability mass functions for discrete random quantities | 501 | 14 |
| B.2.2 | Probability density functions for continuous random quantities | 503 | 15 |
| | | 510 | 16 |
| B.3 | Convergence of sequences of random variables | 510 | 17 |
| | | | 18 |
| Appendix C: | Proofs of selected propositions | 513 | 19 |
| C.1 | Products of normal densities | 513 | 20 |
| C.2 | Conjugate analysis of normal data | 516 | 21 |
| C.3 | Asymptotic normality of the posterior density | 533 | 22 |
| | | | 23 |
| Topic index | | 553 | 24 |
| | | | 25 |
| Author index | | 555 | 26 |
| | | | 27 |
| | | | 28 |
| | | | 29 |
| | | | 30 |
| | | | 31 |
| | | | 32 |
| | | | 33 |
| | | | 34 |
| | | | 35 |
| | | | 36 |
| | | | 37 |
| | | | 38 |
| | | | 39 |
| | | | 40 |
| | | | 41 |
| | | | 42 |
| | | | 43 |
| | | | 44 |
| | | | 45 |
| | | | 46 |
| | | | 47 |

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

UNCORRECTED PROOFS