CONTENTS

Online resources xvii
Acknowledgements xix
About the author xxi

1 How best to use this book 1

1.1 The purpose of this book 2
1.2 Who is this book for? 2
1.3 Prerequisites 2
1.4 Book outline 3
1.5 Route planner – suggested journeys through Bayesland 4
1.6 Video 5
1.7 Problem sets 6
1.8 R and Stan 6
1.9 Why don't more people use Bayesian statistics? 6
1.10 What are the tangible (non-academic) benefits of Bayesian statistics? 7
1.11 Suggested further reading 8

I An introduction to Bayesian inference 9

Part I mission statement 9
Part I goals 9

2 The subjective worlds of Frequentist and Bayesian statistics 11

2.1 Chapter mission statement 12
2.2 Chapter goals 12
2.3 Bayes' rule – allowing us to go from the effect back to its cause 12
2.4 The purpose of statistical inference 17
2.5 The world according to Frequentists 17
2.6 The world according to Bayesians 18
2.7 Do parameters actually exist and have a point value? 19
2.8 Frequentist and Bayesian inference 20
2.8.1 The Frequentist and Bayesian murder trials 22
2.8.2 Radio control towers 22
2.9 Bayesian inference via Bayes' rule 23
2.9.1 Likelihoods 23
2.9.2 Priors 24
2.9.3 The denominator 25
2.9.4 Posteriors: the goal of Bayesian inference 25

2.10 Implicit versus explicit subjectivity 25
2.11 Chapter summary 27
2.12 Chapter outcomes 28
2.13 Appendix 28
2.13.1 The Frequentist and Bayesian murder trials 28
2.14 Problem sets 28

3 Probability – the nuts and bolts of Bayesian inference 33

3.1 Chapter mission statement 34
3.2 Chapter goals 34
3.3 Probability distributions: helping us to explicitly state our ignorance 34
3.3.1 What makes a probability distribution valid? 35
3.3.2 Probabilities versus probability densities: interpreting discrete and continuous probability distributions 36
3.3.3 The mean of a distribution 40
3.3.4 Generalising probability distributions to two dimensions 42
3.3.5 Marginal distributions 44
3.3.6 Conditional distributions 48
3.4 Independence 50
3.5 Central limit theorems 53
3.6 A derivation of Bayes' rule 54
3.6.1 The intuition behind the formula 55
3.6.2 Breast cancer screening 55
3.7 The Bayesian inference process from the Bayesian formula 57
3.8 Chapter summary 58
3.9 Chapter outcomes 59
3.10 Problem sets 59

II Understanding the Bayesian formula 63

Part II mission statement 63
Part II goals 63

4 Likelihoods 65

4.1 Chapter mission statement 66
4.2 Chapter goals 66
4.3 What is a likelihood? 67
4.4 Why use *likelihood* rather than *probability*? 68
4.5 What are models and why do we need them? 70
4.6 How to choose an appropriate likelihood 71
4.6.1 Example: an individual's disease status 71
4.6.2 Example: disease prevalence of a group 73
4.6.3 Example: intelligence test scores for a group of people 76
4.7 Exchangeability versus random sampling 77

4.8 Maximum likelihood: a short introduction 78
4.8.1 Estimating disease prevalence 78
4.8.2 Estimating the mean and variance in intelligence scores 79
4.8.3 Maximum likelihood in simple steps 80
4.8.4 Inference in maximum likelihood 80
4.9 Chapter summary 81
4.10 Chapter outcomes 82
4.11 Problem sets 82

5 Priors **87**

5.1 Chapter mission statement 88
5.2 Chapter goals 88
5.3 What are priors and what do they represent? 88
5.3.1 Why do we need priors at all? 90
5.3.2 Why don't we normalise the likelihood by assuming a unity prior? 90
5.4 The explicit subjectivity of priors 91
5.5 Combining a prior and likelihood to form a posterior 92
5.5.1 The fish game 92
5.5.2 Disease proportions revisited 94
5.6 Constructing priors 96
5.6.1 Uninformative priors 96
5.6.2 Informative priors 98
5.6.3 Eliciting priors 99
5.7 A strong model is less sensitive to prior choice 100
5.7.1 Caveat: zero priors always affect the posterior 102
5.8 Chapter summary 103
5.9 Chapter outcomes 103
5.10 Appendix 104
5.10.1 Bayes' rule for the fish problem 104
5.10.2 The probabilities of having a disease 104
5.11 Problem sets 104

6 The devil is in the denominator **109**

6.1 Chapter mission statement 110
6.2 Chapter goals 110
6.3 An introduction to the denominator 110
6.3.1 The denominator as a normalising factor 110
6.3.2 Example: individual patient disease status 111
6.3.3 Example: the proportion of people who vote for the Conservative Party in a UK general election 112
6.3.4 The denominator as a probability distribution 114
6.4 The difficulty with the denominator 115
6.5 How to dispense with the difficulty: Bayesian computation 116
6.6 Chapter summary 117
6.7 Chapter outcomes 118
6.8 Problem sets 118

7 The posterior - the goal of Bayesian inference 121

7.1 Chapter mission statement 122
7.2 Chapter goals 122
7.3 Expressing parameter uncertainty in posteriors 122
7.4 Bayesian statistics: updating our pre-data uncertainty 124
7.4.1 Example: Bayesian coastguard 124
7.5 The intuition behind Bayes' rule for inference 125
7.6 Point parameter estimates 128
7.7 Intervals of uncertainty 129
7.7.1 Failings of the Frequentist confidence interval 129
7.7.2 Credible intervals 130
7.7.3 Reconciling the difference between confidence and credible intervals 132
7.8 From posterior to predictions by sampling 135
7.9 Chapter summary 137
7.10 Chapter outcomes 137
7.11 Problem sets 138

III Analytic Bayesian methods 141

Part III mission statement 141
Part III goals 141

8 An introduction to distributions for the mathematically uninclined 143

8.1 Chapter mission statement 144
8.2 Chapter goals 144
8.3 The interrelation among distributions 144
8.4 Distributions for likelihoods 146
8.4.1 Bernoulli 146
8.4.2 Binomial 148
8.4.3 Poisson 150
8.4.4 Negative binomial 152
8.4.5 Beta-binomial 155
8.4.6 Normal 157
8.4.7 Student-*t* 159
8.4.8 Exponential 162
8.4.9 Gamma 163
8.4.10 Multinomial 166
8.4.11 Multivariate normal and multivariate Student-*t* 169
8.5 Prior distributions 171
8.5.1 Distributions for probabilities and proportions 172
8.5.2 Distributions for means and regression coefficients 177
8.5.3 Distributions for non-negative parameters 181
8.5.4 Distributions for covariance and correlation matrices 185
8.6 Choosing a likelihood made easy 193

8.7 Table of common likelihoods, their uses and possible priors 194
8.8 Chapter summary 194
8.9 Chapter outcomes 194
8.10 Problem sets 198

9 Conjugate priors 201

9.1 Chapter mission statement 202
9.2 Chapter goals 202
9.3 What are conjugate priors and why are they useful? 202
9.4 Gamma–Poisson example: counting cars 205
9.5 Normal example: giraffe height 207
9.6 Table of conjugate priors 208
9.7 The lessons and limits of a conjugate analysis 208
9.8 Chapter summary 209
9.9 Chapter outcomes 210
9.10 Problem sets 210

10 Evaluation of model fit and hypothesis testing 215

10.1 Chapter mission statement 216
10.2 Chapter goals 216
10.3 Posterior predictive checks 216
10.3.1 Recap – posterior predictive distributions 216
10.3.2 Graphical PPCs and Bayesian *p* values 217
10.4 Why do we call it a *p value*? 224
10.5 Statistics measuring predictive accuracy: AIC, DIC, WAIC and LOO-CV 225
10.5.1 Out-of-sample prediction and overfitting 225
10.5.2 How to measure a model's predictive capability? 226
10.5.3 The ideal measure of a model's predictive accuracy 227
10.5.4 Estimating out-of-sample predictive accuracy from in-sample data 230
10.5.5 AIC 231
10.5.6 DIC 232
10.5.7 WAIC 232
10.5.8 LOO-CV 233
10.5.9 A practical summary of measures of predictive accuracy in simple terms 234
10.6 Marginal likelihoods and Bayes factors 235
10.7 Choosing one model or a number of them? 237
10.8 Sensitivity analysis 237
10.9 Chapter summary 239
10.10 Chapter outcomes 239
10.11 Problem sets 240

11 Making Bayesian analysis objective? 245

11.1 Chapter mission statement 246
11.2 Chapter goals 246

11.3 The illusion of the uninformative uniform prior 246
11.4 Jeffreys priors 247
11.4.1 Another definition of *uninformative* 247
11.4.2 Being critical of Jeffreys priors 250
11.5 Reference priors 251
11.6 Empirical Bayes 253
11.7 A move towards weakly informative priors 255
11.8 Chapter summary 255
11.9 Chapter outcomes 256
11.10 Problem sets 256

IV A practical guide to doing real-life Bayesian analysis: Computational Bayes 259

Part IV mission statement 259
Part IV goals 259

12 Leaving conjugates behind: Markov chain Monte Carlo 263

12.1 Chapter mission statement 264
12.2 Chapter goals 264
12.3 The difficulty with real-life Bayesian inference 265
12.4 Discrete approximation to continuous posteriors 266
12.5 The posterior through quadrature 268
12.6 Integrating using independent samples: an introduction to Monte Carlo 270
12.6.1 Mad cows revisited 272
12.7 Why is independent sampling easier said than done? 273
12.8 Ideal sampling from a posterior using only the numerator of Bayes' rule 276
12.8.1 The un-normalised posterior: a window onto the real deal 278
12.9 Moving from independent to dependent sampling 278
12.9.1 An analogy to MCMC: mapping mountains 280
12.10 What is the catch with dependent samplers? 281
12.11 Chapter summary 283
12.12 Chapter outcomes 284
12.13 Problem sets 284

13 Random Walk Metropolis 291

13.1 Chapter mission statement 292
13.2 Chapter goals 292
13.3 Sustainable fishing 293
13.4 Prospecting for iron 295
13.5 Defining the Random Walk Metropolis algorithm 297
13.6 When does Metropolis work? 299
13.6.1 Mathematical underpinnings of MCMC 300
13.6.2 Desirable qualities of a Markov chain 301

13.6.3 Detailed balance 302
13.6.4 The intuition behind the accept–reject rule of Metropolis and detailed balance 303
13.7 Efficiency of convergence: the importance of choosing the right proposal scale 304
13.7.1 MCMC as finding and exploring the typical set 307
13.7.2 Speeding up convergence: tuning the proposal distribution 307
13.8 Metropolis–Hastings 309
13.9 Judging convergence 310
13.9.1 Bob's bees in a house 311
13.9.2 Using multiple chains to monitor convergence 312
13.9.3 Using within- and between-chain variation to estimate convergence 315
13.9.4 Two types of non-convergence 316
13.9.5 Warm-up 318
13.10 Effective sample size revisited 319
13.10.1 Thinning samples to increase effective sample size 320
13.11 Chapter summary 321
13.12 Chapter outcomes 322
13.13 Problem sets 322

14 Gibbs sampling 329

14.1 Chapter mission statement 330
14.2 Chapter goals 330
14.3 Back to prospecting for iron 331
14.4 Defining the Gibbs algorithm 333
14.4.1 Crime and punishment (unemployment) 334
14.5 Gibbs' earth: the intuition behind the Gibbs algorithm 335
14.6 The benefits and problems with Gibbs and Random Walk Metropolis 336
14.7 A change of parameters to speed up exploration 337
14.8 Chapter summary 338
14.9 Chapter outcomes 339
14.10 Problem sets 339

15 Hamiltonian Monte Carlo 345

15.1 Chapter mission statement 346
15.2 Chapter goals 346
15.3 Hamiltonian Monte Carlo as a sledge 346
15.4 NLP space 348
15.5 Solving for the sledge motion over NLP space 349
15.5.1 Divergent iterations 351
15.6 How to shove the sledge 352
15.7 The acceptance probability of HMC 352
15.8 The complete HMC algorithm 353
15.9 The performance of HMC versus Random Walk Metropolis and Gibbs 354

15.10 Optimal step length of HMC: introducing the 'No-U-Turn Sampler' 354
15.11 Chapter summary 357
15.12 Chapter outcomes 358
15.13 Problem sets 358

16 Stan **365**

16.1 Chapter mission statement 366
16.2 Chapter goals 366
16.3 Why Stan, and how to get it? 367
16.3.1 When to use black box MCMC rather than coding up the algorithms yourself 367
16.3.2 What is Stan? 368
16.3.3 Why choose Stan? 368
16.4 Getting set up with Stan using RStan 369
16.5 Our first words in Stan 370
16.5.1 The main blocks of a Stan program 370
16.5.2 Executing a Stan program from R 375
16.5.3 Interpreting the results of a Stan MCMC run 376
16.5.4 The other blocks 378
16.5.5 Efficiency and memory considerations for each block 383
16.5.6 Loops and conditional statements 385
16.6 Essential Stan reading 386
16.7 What to do when things go wrong 406
16.7.1 Coding errors 406
16.7.2 Sampling issues 409
16.8 How to get further help 415
16.9 Chapter summary 416
16.10 Chapter outcomes 416
16.11 Problem sets 417

V Hierarchical models and regression **425**

Part V mission statement 425
Part V goals 425

17 Hierarchical models **429**

17.1 Chapter mission statement 430
17.2 Chapter goals 430
17.3 The spectrum from fully pooled to heterogeneous models 431
17.3.1 Fully pooled model 431
17.3.2 Heterogeneous model 434
17.3.3 Hierarchical model 435
17.3.4 The additional benefits of hierarchical models 437
17.4 Non-centred parameterisations in hierarchical models 438
17.5 Case study: forecasting the result of the UK's referendum on EU membership 440

17.6 The importance of fake data simulation for complex models 442
17.7 Chapter summary 444
17.8 Chapter outcomes 445
17.9 Problem sets 445

18 Linear regression models 453

18.1 Chapter mission statement 454
18.2 Chapter goals 454
18.3 Example: high school test scores in England 454
18.4 Pooled model 456
18.5 Interactions 457
18.6 Heterogeneous coefficient model 459
18.7 Hierarchical model 460
18.8 Incorporating LEA-level data 463
18.9 Chapter summary 465
18.10 Chapter outcomes 466
18.11 Problem sets 466

19 Generalised linear models and other animals 469

19.1 Chapter mission statement 470
19.2 Chapter goals 470
19.3 Example: electoral participation in European countries 471
19.3.1 Multilevel interactions 476
19.3.2 Comparing the predictive fit of each of the models 479
19.4 Discrete parameter models in Stan 480
19.4.1 Example: how many times was the coin flipped? 481
19.5 Chapter summary 484
19.6 Chapter outcomes 485
19.7 Problem sets 485

Bibliography 489
Index 493