

Contents

Foreword xiii

Preface xv

Acknowledgments xvii

About the Author xix

1 The Philosophy of Bayesian Inference 1

- 1.1 Introduction 1
 - 1.1.1 The Bayesian State of Mind 1
 - 1.1.2 Bayesian Inference in Practice 3
 - 1.1.3 Are Frequentist Methods Incorrect? 4
 - 1.1.4 A Note on “Big Data” 4
- 1.2 Our Bayesian Framework 5
 - 1.2.1 Example: Mandatory Coin-Flip 5
 - 1.2.2 Example: Librarian or Farmer? 6
- 1.3 Probability Distributions 8
 - 1.3.1 Discrete Case 9
 - 1.3.2 Continuous Case 10
 - 1.3.3 But What Is λ ? 12
- 1.4 Using Computers to Perform Bayesian Inference for Us 12
 - 1.4.1 Example: Inferring Behavior from Text-Message Data 12
 - 1.4.2 Introducing Our First Hammer: PyMC 14
 - 1.4.3 Interpretation 18
 - 1.4.4 What Good Are Samples from the Posterior, Anyway? 18
- 1.5 Conclusion 20
- 1.6 Appendix 20
 - 1.6.1 Determining Statistically if the Two λ s Are Indeed Different? 20
 - 1.6.2 Extending to Two Switchpoints 22
- 1.7 Exercises 24
 - 1.7.1 Answers 24
- 1.8 References 25

2	A Little More on PyMC	27
2.1	Introduction	27
2.1.1	Parent and Child Relationships	27
2.1.2	PyMC Variables	28
2.1.3	Including Observations in the Model	31
2.1.4	Finally...	33
2.2	Modeling Approaches	33
2.2.1	Same Story, Different Ending	35
2.2.2	Example: Bayesian A/B Testing	38
2.2.3	A Simple Case	38
2.2.4	A and B Together	41
2.2.5	Example: An Algorithm for Human Deceit	45
2.2.6	The Binomial Distribution	45
2.2.7	Example: Cheating Among Students	46
2.2.8	Alternative PyMC Model	50
2.2.9	More PyMC Tricks	51
2.2.10	Example: <i>Challenger</i> Space Shuttle Disaster	52
2.2.11	The Normal Distribution	55
2.2.12	What Happened the Day of the <i>Challenger</i> Disaster?	61
2.3	Is Our Model Appropriate?	61
2.3.1	Separation Plots	64
2.4	Conclusion	68
2.5	Appendix	68
2.6	Exercises	69
2.6.1	Answers	69
2.7	References	69
3	Opening the Black Box of MCMC	71
3.1	The Bayesian Landscape	71
3.1.1	Exploring the Landscape Using MCMC	76
3.1.2	Algorithms to Perform MCMC	78
3.1.3	Other Approximation Solutions to the Posterior	79
3.1.4	Example: Unsupervised Clustering Using a Mixture Model	79

3.1.5	Don't Mix Posterior Samples	88
3.1.6	Using MAP to Improve Convergence	91
3.2	Diagnosing Convergence	92
3.2.1	Autocorrelation	92
3.2.2	Thinning	95
3.2.3	<code>pymc.Matplot.plot()</code>	97
3.3	Useful Tips for MCMC	98
3.3.1	Intelligent Starting Values	98
3.3.2	Priors	99
3.3.3	The Folk Theorem of Statistical Computing	99
3.4	Conclusion	99
3.5	Reference	99

4 The Greatest Theorem Never Told 101

4.1	Introduction	101
4.2	The Law of Large Numbers	101
4.2.1	Intuition	101
4.2.2	Example: Convergence of Poisson Random Variables	102
4.2.3	How Do We Compute $Var(Z)$?	106
4.2.4	Expected Values and Probabilities	106
4.2.5	What Does All This Have to Do with Bayesian Statistics?	107
4.3	The Disorder of Small Numbers	107
4.3.1	Example: Aggregated Geographic Data	107
4.3.2	Example: Kaggle's <i>U.S. Census Return Rate Challenge</i>	109
4.3.3	Example: How to Sort Reddit Comments	111
4.3.4	Sorting!	115
4.3.5	But This Is Too Slow for Real-Time!	117
4.3.6	Extension to Starred Rating Systems	122
4.4	Conclusion	122
4.5	Appendix	122
4.5.1	Derivation of Sorting Comments Formula	122
4.6	Exercises	123
4.6.1	Answers	124
4.7	References	125

5	Would You Rather Lose an Arm or a Leg?	127
5.1	Introduction	127
5.2	Loss Functions	127
5.2.1	Loss Functions in the Real World	129
5.2.2	Example: Optimizing for the Showcase on <i>The Price Is Right</i>	131
5.3	Machine Learning via Bayesian Methods	139
5.3.1	Example: Financial Prediction	139
5.3.2	Example: Kaggle Contest on Observing Dark Worlds	144
5.3.3	The Data	145
5.3.4	Priors	146
5.3.5	Training and PyMC Implementation	148
5.4	Conclusion	156
5.5	References	156
6	Getting Our Priorities Straight	157
6.1	Introduction	157
6.2	Subjective versus Objective Priors	157
6.2.1	Objective Priors	157
6.2.2	Subjective Priors	158
6.2.3	Decisions, Decisions ...	159
6.2.4	Empirical Bayes	160
6.3	Useful Priors to Know About	161
6.3.1	The Gamma Distribution	161
6.3.2	The Wishart Distribution	161
6.3.3	The Beta Distribution	163
6.4	Example: Bayesian Multi-Armed Bandits	164
6.4.1	Applications	165
6.4.2	A Proposed Solution	165
6.4.3	A Measure of Good	169
6.4.4	Extending the Algorithm	173
6.5	Eliciting Prior Distributions from Domain Experts	176
6.5.1	Trial Roulette Method	176
6.5.2	Example: Stock Returns	177
6.5.3	Pro Tips for the Wishart Distribution	184
6.6	Conjugate Priors	185
6.7	Jeffreys Priors	185

- 6.8 Effect of the Prior as N Increases 187
- 6.9 Conclusion 189
- 6.10 Appendix 190
 - 6.10.1 Bayesian Perspective of Penalized Linear Regressions 190
 - 6.10.2 Picking a Degenerate Prior 192
- 6.11 References 193

7 Bayesian A/B Testing 195

- 7.1 Introduction 195
- 7.2 Conversion Testing Recap 195
- 7.3 Adding a Linear Loss Function 198
 - 7.3.1 Expected Revenue Analysis 198
 - 7.3.2 Extending to an A/B Experiment 202
- 7.4 Going Beyond Conversions: t-test 204
 - 7.4.1 The Setup of the t-test 204
- 7.5 Estimating the Increase 207
 - 7.5.1 Creating Point Estimates 210
- 7.6 Conclusion 211
- 7.7 References 212

Glossary 213

Index 217