



ADDISON
WESLEY
DATA &
ANALYTICS
SERIES



BAYESIAN METHODS FOR Hackers

Probabilistic
Programming and
Bayesian Inference

CAMERON DAVIDSON-PILON

Bayesian Methods for Hackers

Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference

Table of Contents

Contents

Foreword

Preface

Acknowledgments

About the Author

1 The Philosophy of Bayesian Inference

1.1 Introduction

1.1.1 The Bayesian State of Mind

1.1.2 Bayesian Inference in Practice

1.1.3 Are Frequentist Methods Incorrect?

1.1.4 A Note on Big Data

1.2 Our Bayesian Framework

1.2.1 Example: Mandatory Coin-Flip

1.2.2 Example: Librarian or Farmer?

1.3 Probability Distributions

1.3.1 Discrete Case

1.3.2 Continuous Case

1.3.3 But What Is ?

1.4 Using Computers to Perform Bayesian Inference for Us

1.4.1 Example: Inferring Behavior from Text-Message Data

1.4.2 Introducing Our First Hammer: PyMC

Table of Contents

1.4.3 Interpretation

1.4.4 What Good Are Samples from the Posterior, Anyway?

1.5 Conclusion

1.6 Appendix

1.6.1 Determining Statistically if the Two θ s Are Indeed Different?

1.6.2 Extending to Two Switchpoints

1.7 Exercises

1.7.1 Answers

1.8 References

2 A Little More on PyMC

2.1 Introduction

2.1.1 Parent and Child Relationships

2.1.2 PyMC Variables

2.1.3 Including Observations in the Model

2.1.4 Finally. . .

2.2 Modeling Approaches

2.2.1 Same Story, Different Ending

2.2.2 Example: Bayesian A/B Testing

2.2.3 A Simple Case

2.2.4 A and B Together

2.2.5 Example: An Algorithm for Human Deceit

2.2.6 The Binomial Distribution

2.2.7 Example: Cheating Among Students

2.2.8 Alternative PyMC Model

2.2.9 More PyMC Tricks

2.2.10 Example: Challenger Space Shuttle Disaster

2.2.11 The Normal Distribution

2.2.12 What Happened the Day of the Challenger Disaster?

2.3 Is Our Model Appropriate?

Table of Contents

2.3.1 Separation Plots

2.4 Conclusion

2.5 Appendix

2.6 Exercises

2.6.1 Answers

2.7 References

3 Opening the Black Box of MCMC

3.1 The Bayesian Landscape

3.1.1 Exploring the Landscape Using MCMC

3.1.2 Algorithms to Perform MCMC

3.1.3 Other Approximation Solutions to the Posterior

3.1.4 Example: Unsupervised Clustering Using a Mixture Model

3.1.5 Dont Mix Posterior Samples

3.1.6 Using MAP to Improve Convergence

3.2 Diagnosing Convergence

3.2.1 Autocorrelation

3.2.2 Thinning

3.2.3 `pymc.Matplot.plot()`

3.3 Useful Tips for MCMC

3.3.1 Intelligent Starting Values

3.3.2 Priors

3.3.3 The Folk Theorem of Statistical Computing

3.4 Conclusion

3.5 Reference

4 The Greatest Theorem Never Told

4.1 Introduction

4.2 The Law of Large Numbers

4.2.1 Intuition

Table of Contents

4.2.2 Example: Convergence of Poisson Random Variables

4.2.3 How Do We Compute $\text{Var}(Z)$?

4.2.4 Expected Values and Probabilities

4.2.5 What Does All This Have to Do with Bayesian Statistics?

4.3 The Disorder of Small Numbers

4.3.1 Example: Aggregated Geographic Data

4.3.2 Example: Kaggle's U.S. Census Return Rate Challenge

4.3.3 Example: How to Sort Reddit Comments

4.3.4 Sorting!

4.3.5 But This Is Too Slow for Real-Time!

4.3.6 Extension to Starred Rating Systems

4.4 Conclusion

4.5 Appendix

4.5.1 Derivation of Sorting Comments Formula

4.6 Exercises

4.6.1 Answers

4.7 References

5 Would You Rather Lose an Arm or a Leg?

5.1 Introduction

5.2 Loss Functions

5.2.1 Loss Functions in the Real World

5.2.2 Example: Optimizing for the Showcase on The Price Is Right

5.3 Machine Learning via Bayesian Methods

5.3.1 Example: Financial Prediction

5.3.2 Example: Kaggle Contest on Observing Dark Worlds

5.3.3 The Data

5.3.4 Priors

5.3.5 Training and PyMC Implementation

5.4 Conclusion

Table of Contents

5.5 References

6 Getting Our Priorities Straight

6.1 Introduction

6.2 Subjective versus Objective Priors

6.2.1 Objective Priors

6.2.2 Subjective Priors

6.2.3 Decisions, Decisions . . .

6.2.4 Empirical Bayes

6.3 Useful Priors to Know About

6.3.1 The Gamma Distribution

6.3.2 The Wishart Distribution

6.3.3 The Beta Distribution

6.4 Example: Bayesian Multi-Armed Bandits

6.4.1 Applications

6.4.2 A Proposed Solution

6.4.3 A Measure of Good

6.4.4 Extending the Algorithm

6.5 Eliciting Prior Distributions from Domain Experts

6.5.1 Trial Roulette Method

6.5.2 Example: Stock Returns

6.5.3 Pro Tips for the Wishart Distribution 184

6.6 Conjugate Priors

6.7 Jeffreys Priors

6.8 Effect of the Prior as N Increases

6.9 Conclusion

6.10 Appendix

6.10.1 Bayesian Perspective of Penalized Linear Regressions

6.10.2 Picking a Degenerate Prior

Table of Contents

6.11 References

7 Bayesian A/B Testing

7.1 Introduction

7.2 Conversion Testing Recap

7.3 Adding a Linear Loss Function

7.3.1 Expected Revenue Analysis

7.3.2 Extending to an A/B Experiment

7.4 Going Beyond Conversions: t-test

7.4.1 The Setup of the t-test

7.5 Estimating the Increase

7.5.1 Creating Point Estimates

7.6 Conclusion

7.7 References

Glossary

A

B

C

D

E

F

G

L

M

O

P

S

T

Table of Contents

W
Index