

SECOND EDITION

Think Bayes

Bayesian Statistics in Python

Allen B. Downey

Beijing • Boston • Farnham • Sebastopol • Tokyo

O'REILLY®

Table of Contents

Preface.....	ix
1. Probability.....	1
Linda the Banker	1
Probability	2
Fraction of Bankers	3
The Probability Function	4
Political Views and Parties	4
Conjunction	5
Conditional Probability	6
Conditional Probability Is Not Commutative	7
Condition and Conjunction	8
Laws of Probability	8
Theorem 1	9
Theorem 2	10
Theorem 3	10
The Law of Total Probability	11
Summary	13
Exercises	14
2. Bayes's Theorem.....	17
The Cookie Problem	17
Diachronic Bayes	19
Bayes Tables	20
The Dice Problem	22
The Monty Hall Problem	23
Summary	25
Exercises	26

3. Distributions.....	29
Distributions	29
Probability Mass Functions	29
The Cookie Problem Revisited	32
101 Bowls	34
The Dice Problem	38
Updating Dice	39
Summary	40
Exercises	41
4. Estimating Proportions.....	43
The Euro Problem	43
The Binomial Distribution	44
Bayesian Estimation	47
Triangle Prior	49
The Binomial Likelihood Function	51
Bayesian Statistics	52
Summary	53
Exercises	54
5. Estimating Counts.....	57
The Train Problem	57
Sensitivity to the Prior	60
Power Law Prior	61
Credible Intervals	63
The German Tank Problem	64
Informative Priors	65
Summary	66
Exercises	66
6. Odds and Addends.....	69
Odds	69
Bayes's Rule	70
Oliver's Blood	71
Addends	73
Gluten Sensitivity	76
The Forward Problem	77
The Inverse Problem	78
Summary	80
More Exercises	81

7. Minimum, Maximum, and Mixture.....	83
Cumulative Distribution Functions	83
Best Three of Four	86
Maximum	88
Minimum	89
Mixture	90
General Mixtures	93
Summary	96
Exercises	97
8. Poisson Processes.....	99
The World Cup Problem	99
The Poisson Distribution	100
The Gamma Distribution	101
The Update	103
Probability of Superiority	105
Predicting the Rematch	106
The Exponential Distribution	108
Summary	110
Exercises	110
9. Decision Analysis.....	113
The Price Is Right Problem	113
The Prior	114
Kernel Density Estimation	115
Distribution of Error	116
Update	118
Probability of Winning	120
Decision Analysis	122
Maximizing Expected Gain	124
Summary	126
Discussion	126
More Exercises	127
10. Testing.....	129
Estimation	129
Evidence	131
Uniformly Distributed Bias	132
Bayesian Hypothesis Testing	134
Bayesian Bandits	134
Prior Beliefs	135
The Update	136

Multiple Bandits	137
Explore and Exploit	138
The Strategy	140
Summary	142
More Exercises	142
11. Comparison.....	145
Outer Operations	145
How Tall Is A?	147
Joint Distribution	148
Visualizing the Joint Distribution	149
Likelihood	151
The Update	152
Marginal Distributions	153
Conditional Posteriors	156
Dependence and Independence	157
Summary	158
Exercises	158
12. Classification.....	161
Penguin Data	161
Normal Models	163
The Update	164
Naive Bayesian Classification	166
Joint Distributions	168
Multivariate Normal Distribution	170
A Less Naive Classifier	172
Summary	173
Exercises	173
13. Inference.....	175
Improving Reading Ability	175
Estimating Parameters	177
Likelihood	178
Posterior Marginal Distributions	180
Distribution of Differences	181
Using Summary Statistics	184
Update with Summary Statistics	186
Comparing Marginals	187
Summary	188
Exercises	189

14. Survival Analysis.....	191
The Weibull Distribution	191
Incomplete Data	194
Using Incomplete Data	196
Light Bulbs	199
Posterior Means	201
Posterior Predictive Distribution	202
Summary	204
Exercises	204
15. Mark and Recapture.....	207
The Grizzly Bear Problem	207
The Update	209
Two-Parameter Model	211
The Prior	212
The Update	213
The Lincoln Index Problem	215
Three-Parameter Model	217
Summary	220
Exercises	221
16. Logistic Regression.....	223
Log Odds	223
The Space Shuttle Problem	226
Prior Distribution	229
Likelihood	230
The Update	231
Marginal Distributions	232
Transforming Distributions	233
Predictive Distributions	235
Empirical Bayes	237
Summary	238
More Exercises	238
17. Regression.....	241
More Snow?	241
Regression Model	243
Least Squares Regression	244
Priors	245
Likelihood	246
The Update	247
Marathon World Record	250

The Priors	252
Prediction	254
Summary	255
Exercises	255
18. Conjugate Priors.....	257
The World Cup Problem Revisited	257
The Conjugate Prior	258
What the Actual?	260
Binomial Likelihood	261
Lions and Tigers and Bears	263
The Dirichlet Distribution	264
Summary	266
Exercises	267
19. MCMC.....	269
The World Cup Problem	269
Grid Approximation	270
Prior Predictive Distribution	270
Introducing PyMC3	271
Sampling the Prior	272
When Do We Get to Inference?	274
Posterior Predictive Distribution	275
Happiness	276
Simple Regression	277
Multiple Regression	280
Summary	282
Exercises	283
20. Approximate Bayesian Computation.....	287
The Kidney Tumor Problem	287
A Simple Growth Model	288
A More General Model	289
Simulation	291
Approximate Bayesian Computation	294
Counting Cells	295
Cell Counting with ABC	298
When Do We Get to the Approximate Part?	299
Summary	302
Exercises	303
Index.....	305