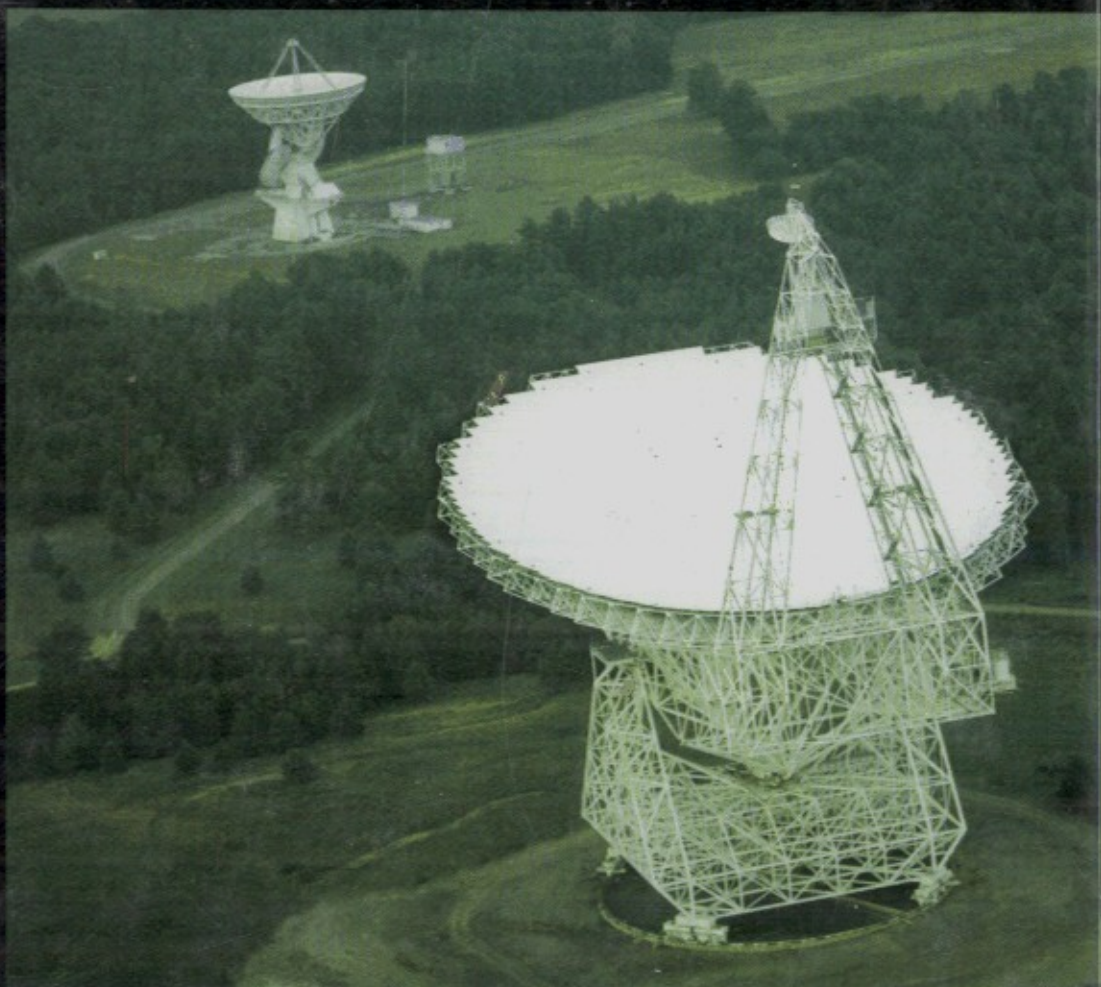


PHIL GREGORY

Bayesian Logical Data Analysis for the Physical Sciences

A Comparative Approach with
Mathematica® Support



CAMBRIDGE

Contents

<i>Preface</i>	page xiii
Software support	xv
<i>Acknowledgements</i>	xvii
1 Role of probability theory in science	1
1.1 Scientific inference	1
1.2 Inference requires a probability theory	2
1.2.1 The two rules for manipulating probabilities	4
1.3 Usual form of Bayes' theorem	5
1.3.1 Discrete hypothesis space	5
1.3.2 Continuous hypothesis space	6
1.3.3 Bayes' theorem – model of the learning process	7
1.3.4 Example of the use of Bayes' theorem	8
1.4 Probability and frequency	10
1.4.1 Example: incorporating frequency information	11
1.5 Marginalization	12
1.6 The two basic problems in statistical inference	15
1.7 Advantages of the Bayesian approach	16
1.8 Problems	17
2 Probability theory as extended logic	21
2.1 Overview	21
2.2 Fundamentals of logic	21
2.2.1 Logical propositions	21
2.2.2 Compound propositions	22
2.2.3 Truth tables and Boolean algebra	22
2.2.4 Deductive inference	24
2.2.5 Inductive or plausible inference	25
2.3 Brief history	25
2.4 An adequate set of operations	26
2.4.1 Examination of a logic function	27
2.5 Operations for plausible inference	29

2.5.1	The desiderata of Bayesian probability theory	30
2.5.2	Development of the product rule	30
2.5.3	Development of sum rule	34
2.5.4	Qualitative properties of product and sum rules	36
2.6	Uniqueness of the product and sum rules	37
2.7	Summary	39
2.8	Problems	39
3	The how-to of Bayesian inference	41
3.1	Overview	41
3.2	Basics	41
3.3	Parameter estimation	43
3.4	Nuisance parameters	45
3.5	Model comparison and Occam's razor	45
3.6	Sample spectral line problem	50
3.6.1	Background information	50
3.7	Odds ratio	52
3.7.1	Choice of prior $p(T M_1, I)$	53
3.7.2	Calculation of $p(D M_1, T, I)$	55
3.7.3	Calculation of $p(D M_2, I)$	58
3.7.4	Odds, uniform prior	58
3.7.5	Odds, Jeffreys prior	58
3.8	Parameter estimation problem	59
3.8.1	Sensitivity of odds to $T_{\max}$	59
3.9	Lessons	61
3.10	Ignorance priors	63
3.11	Systematic errors	65
3.11.1	Systematic error example	66
3.12	Problems	69
4	Assigning probabilities	72
4.1	Introduction	72
4.2	Binomial distribution	72
4.2.1	Bernoulli's law of large numbers	75
4.2.2	The gambler's coin problem	75
4.2.3	Bayesian analysis of an opinion poll	77
4.3	Multinomial distribution	79
4.4	Can you really answer that question?	80
4.5	Logical versus causal connections	82
4.6	Exchangeable distributions	83
4.7	Poisson distribution	85

4.7.1	Bayesian and frequentist comparison	87
4.8	Constructing likelihood functions	89
4.8.1	Deterministic model	90
4.8.2	Probabilistic model	91
4.9	Summary	93
4.10	Problems	94
5	Frequentist statistical inference	96
5.1	Overview	96
5.2	The concept of a random variable	96
5.3	Sampling theory	97
5.4	Probability distributions	98
5.5	Descriptive properties of distributions	100
5.5.1	Relative line shape measures for distributions	101
5.5.2	Standard random variable	102
5.5.3	Other measures of central tendency and dispersion	103
5.5.4	Median baseline subtraction	104
5.6	Moment generating functions	105
5.7	Some discrete probability distributions	107
5.7.1	Binomial distribution	107
5.7.2	The Poisson distribution	109
5.7.3	Negative binomial distribution	112
5.8	Continuous probability distributions	113
5.8.1	Normal distribution	113
5.8.2	Uniform distribution	116
5.8.3	Gamma distribution	116
5.8.4	Beta distribution	117
5.8.5	Negative exponential distribution	118
5.9	Central Limit Theorem	119
5.10	Bayesian demonstration of the Central Limit Theorem	120
5.11	Distribution of the sample mean	124
5.11.1	Signal averaging example	125
5.12	Transformation of a random variable	125
5.13	Random and pseudo-random numbers	127
5.13.1	Pseudo-random number generators	131
5.13.2	Tests for randomness	132
5.14	Summary	136
5.15	Problems	137
6	What is a statistic?	139
6.1	Introduction	139
6.2	The χ^2 distribution	141

6.3	Sample variance S^2	143
6.4	The Student's t distribution	147
6.5	F distribution (F -test)	150
6.6	Confidence intervals	152
6.6.1	Variance σ^2 known	152
6.6.2	Confidence intervals for μ , unknown variance	156
6.6.3	Confidence intervals: difference of two means	158
6.6.4	Confidence intervals for σ^2	159
6.6.5	Confidence intervals: ratio of two variances	159
6.7	Summary	160
6.8	Problems	161
7	Frequentist hypothesis testing	162
7.1	Overview	162
7.2	Basic idea	162
7.2.1	Hypothesis testing with the χ^2 statistic	163
7.2.2	Hypothesis test on the difference of two means	167
7.2.3	One-sided and two-sided hypothesis tests	170
7.3	Are two distributions the same?	172
7.3.1	Pearson χ^2 goodness-of-fit test	173
7.3.2	Comparison of two-binned data sets	177
7.4	Problem with frequentist hypothesis testing	177
7.4.1	Bayesian resolution to optional stopping problem	179
7.5	Problems	181
8	Maximum entropy probabilities	184
8.1	Overview	184
8.2	The maximum entropy principle	185
8.3	Shannon's theorem	186
8.4	Alternative justification of MaxEnt	187
8.5	Generalizing MaxEnt	190
8.5.1	Incorporating a prior	190
8.5.2	Continuous probability distributions	191
8.6	How to apply the MaxEnt principle	191
8.6.1	Lagrange multipliers of variational calculus	191
8.7	MaxEnt distributions	192
8.7.1	General properties	192
8.7.2	Uniform distribution	194
8.7.3	Exponential distribution	195
8.7.4	Normal and truncated Gaussian distributions	197
8.7.5	Multivariate Gaussian distribution	202

8.8	MaxEnt image reconstruction	203
8.8.1	The kangaroo justification	203
8.8.2	MaxEnt for uncertain constraints	206
8.9	Pixon multiresolution image reconstruction	208
8.10	Problems	211
9	Bayesian inference with Gaussian errors	212
9.1	Overview	212
9.2	Bayesian estimate of a mean	212
9.2.1	Mean: known noise σ	213
9.2.2	Mean: known noise, unequal σ	217
9.2.3	Mean: unknown noise σ	218
9.2.4	Bayesian estimate of σ	224
9.3	Is the signal variable?	227
9.4	Comparison of two independent samples	228
9.4.1	Do the samples differ?	230
9.4.2	How do the samples differ?	233
9.4.3	Results	233
9.4.4	The difference in means	236
9.4.5	Ratio of the standard deviations	237
9.4.6	Effect of the prior ranges	239
9.5	Summary	240
9.6	Problems	241
10	Linear model fitting (Gaussian errors)	243
10.1	Overview	243
10.2	Parameter estimation	244
10.2.1	Most probable amplitudes	249
10.2.2	More powerful matrix formulation	253
10.3	Regression analysis	256
10.4	The posterior is a Gaussian	257
10.4.1	Joint credible regions	260
10.5	Model parameter errors	264
10.5.1	Marginalization and the covariance matrix	264
10.5.2	Correlation coefficient	268
10.5.3	More on model parameter errors	272
10.6	Correlated data errors	273
10.7	Model comparison with Gaussian posteriors	275
10.8	Frequentist testing and errors	279
10.8.1	Other model comparison methods	281
10.9	Summary	283
10.10	Problems	284

11 Nonlinear model fitting	287
11.1 Introduction	287
11.2 Asymptotic normal approximation	288
11.3 Laplacian approximations	291
11.3.1 Bayes factor	291
11.3.2 Marginal parameter posteriors	293
11.4 Finding the most probable parameters	294
11.4.1 Simulated annealing	296
11.4.2 Genetic algorithm	297
11.5 Iterative linearization	298
11.5.1 Levenberg–Marquardt method	300
11.5.2 Marquardt’s recipe	301
11.6 <i>Mathematica</i> example	302
11.6.1 Model comparison	304
11.6.2 Marginal and projected distributions	306
11.7 Errors in both coordinates	307
11.8 Summary	309
11.9 Problems	309
12 Markov chain Monte Carlo	312
12.1 Overview	312
12.2 Metropolis–Hastings algorithm	313
12.3 Why does Metropolis–Hastings work?	319
12.4 Simulated tempering	321
12.5 Parallel tempering	321
12.6 Example	322
12.7 Model comparison	326
12.8 Towards an automated MCMC	330
12.9 Extrasolar planet example	331
12.9.1 Model probabilities	335
12.9.2 Results	337
12.10 MCMC robust summary statistic	342
12.11 Summary	346
12.12 Problems	349
13 Bayesian revolution in spectral analysis	352
13.1 Overview	352
13.2 New insights on the periodogram	352
13.2.1 How to compute $p(f D, I)$	356
13.3 Strong prior signal model	358
13.4 No specific prior signal model	360

13.4.1	X-ray astronomy example	362
13.4.2	Radio astronomy example	363
13.5	Generalized Lomb–Scargle periodogram	365
13.5.1	Relationship to Lomb–Scargle periodogram	367
13.5.2	Example	367
13.6	Non-uniform sampling	370
13.7	Problems	373
14	Bayesian inference with Poisson sampling	376
14.1	Overview	376
14.2	Infer a Poisson rate	377
14.2.1	Summary of posterior	378
14.3	Signal + known background	379
14.4	Analysis of ON/OFF measurements	380
14.4.1	Estimating the source rate	381
14.4.2	Source detection question	384
14.5	Time-varying Poisson rate	386
14.6	Problems	388
Appendix A	Singular value decomposition	389
Appendix B	Discrete Fourier Transforms	392
B.1	Overview	392
B.2	Orthogonal and orthonormal functions	392
B.3	Fourier series and integral transform	394
B.3.1	Fourier series	395
B.3.2	Fourier transform	396
B.4	Convolution and correlation	398
B.4.1	Convolution theorem	399
B.4.2	Correlation theorem	400
B.4.3	Importance of convolution in science	401
B.5	Waveform sampling	403
B.6	Nyquist sampling theorem	404
B.6.1	Astronomy example	406
B.7	Discrete Fourier Transform	407
B.7.1	Graphical development	407
B.7.2	Mathematical development of the DFT	409
B.7.3	Inverse DFT	410
B.8	Applying the DFT	411
B.8.1	DFT as an approximate Fourier transform	411
B.8.2	Inverse discrete Fourier transform	413

B.9	The Fast Fourier Transform	415
B.10	Discrete convolution and correlation	417
B.10.1	Deconvolving a noisy signal	418
B.10.2	Deconvolution with an optimal Wiener filter	420
B.10.3	Treatment of end effects by zero padding	421
B.11	Accurate amplitudes by zero padding	422
B.12	Power-spectrum estimation	424
B.12.1	Parseval's theorem and power spectral density	424
B.12.2	Periodogram power-spectrum estimation	425
B.12.3	Correlation spectrum estimation	426
B.13	Discrete power spectral density estimation	428
B.13.1	Discrete form of Parseval's theorem	428
B.13.2	One-sided discrete power spectral density	429
B.13.3	Variance of periodogram estimate	429
B.13.4	Yule's stochastic spectrum estimation model	431
B.13.5	Reduction of periodogram variance	431
B.14	Problems	432
Appendix C	Difference in two samples	434
C.1	Outline	434
C.2	Probabilities of the four hypotheses	434
C.2.1	Evaluation of $p(C, S D_1, D_2, I)$	434
C.2.2	Evaluation of $p(C, \bar{S} D_1, D_2, I)$	436
C.2.3	Evaluation of $p(\bar{C}, S D_1, D_2, I)$	438
C.2.4	Evaluation of $p(\bar{C}, \bar{S} D_1, D_2, I)$	439
C.3	The difference in the means	439
C.3.1	The two-sample problem	440
C.3.2	The Behrens–Fisher problem	441
C.4	The ratio of the standard deviations	442
C.4.1	Estimating the ratio, given the means are the same	442
C.4.2	Estimating the ratio, given the means are different	443
Appendix D	Poisson ON/OFF details	445
D.1	Derivation of $p(s N_{\text{on}}, I)$	445
D.1.1	Evaluation of Num	446
D.1.2	Evaluation of Den	447
D.2	Derivation of the Bayes factor $B_{\{s+b, b\}}$	448
Appendix E	Multivariate Gaussian from maximum entropy	450
<i>References</i>		455
<i>Index</i>		461