

Statistics lectures have been a source of much bewilderment and frustration for generations of students. This book attempts to remedy the situation by expounding a logical and unified approach to the whole subject of data analysis.

This text is intended as a tutorial guide for senior undergraduates and research students in science and engineering. After explaining the basic principles of Bayesian probability theory, their use is illustrated with a variety of examples ranging from elementary parameter estimation to image processing. Other topics covered include reliability analysis, multivariate optimization, least-squares and maximum likelihood, error-propagation, hypothesis testing, maximum entropy and experimental design.

The Second Edition of this successful tutorial book contains a new chapter on extensions to the ubiquitous least-squares procedure, allowing for the straightforward handling of outliers and unknown correlated noise, and a cutting-edge contribution from John Skilling on a novel numerical technique for Bayesian computation called 'nested sampling'.

D. S. Sivia is a staff scientist at the Rutherford Appleton Laboratory and a college lecturer at St Catherine's College, and St John's College, Oxford University.

J. Skilling is an expert in numerical techniques and a founding director of Maximum Entropy Data Consultants. He taught at Cambridge University for over twenty years.

'This small (less than 200 pages) but much-needed book contains a wealth of worked-out numerical examples of Bayesian treatments of data, expounded from a theoretical standpoint identical to ours. It should be considered an adjunct to the present work, supplying a great deal of practical advice for the beginner, at an elementary level that will be grasped readily by every science or engineering student.'

E. T. Jaynes in *Probability Theory: The Logic of Science*

Cover image: The solid line shows the probability distribution for the inferred value of a certain quantity, given three separate measurements of it, where allowance has been made for the possibility of outliers; the dotted line indicates the corresponding result from the standard least-squares analysis.

ALSO PUBLISHED BY OXFORD UNIVERSITY PRESS

G. Cowan: Statistical Data Analysis

W. Krauth: Statistical Mechanics: Algorithms and Computations

J. Sethna: Statistical Mechanics: Entropy, Order Parameters and Complexity

OXFORD
UNIVERSITY PRESS

www.oup.com

ISBN 978-0-19-856832-2



9 780198 568322

PART I THE ESSENTIALS

1. The basics	3
1.1 Introduction: deductive logic versus plausible reasoning	3
1.2 Probability: Cox and the rules for consistent reasoning	4
1.3 Corollaries: Bayes' theorem and marginalization	5
1.4 Some history: Bayes, Laplace and orthodox statistics	8
1.5 Outline of book	12
2. Parameter estimation I	14
2.1 Example 1: is this a fair coin?	14
2.1.1 Different priors	17
2.1.2 Sequential or one-step data analysis?	19
2.2 Reliabilities: best estimates, error-bars and confidence intervals	20
2.2.1 The coin example	23
2.2.2 Asymmetric posterior pdfs	24
2.2.3 Multimodal posterior pdfs	25
2.3 Example 2: Gaussian noise and averages	26
2.3.1 Data with different-sized error-bars	29
2.4 Example 3: the lighthouse problem	29
2.4.1 The central limit theorem	33
3. Parameter estimation II	35
3.1 Example 4: amplitude of a signal in the presence of background	35
3.1.1 Marginal distributions	39
3.1.2 Binning the data	42
3.2 Reliabilities: best estimates, correlations and error-bars	43
3.2.1 Generalization of the quadratic approximation	49
3.2.2 Asymmetric and multimodal posterior pdfs	50
3.3 Example 5: Gaussian noise revisited	52
3.3.1 The Student- t and χ^2 distributions	54
3.4 Algorithms: a numerical interlude	55
3.4.1 Brute force and ignorance	56
3.4.2 The joys of linearity	57
3.4.3 Iterative linearization	58
3.4.4 Hard problems	60
3.5 Approximations: maximum likelihood and least-squares	61
3.5.1 Fitting a straight line	65

3.6	Error-propagation: changing variables	68
3.6.1	A useful short cut	73
3.6.2	Taking the square root of a number	74
4.	Model selection	78
4.1	Introduction: the story of Mr A and Mr B	78
4.1.1	Comparison with parameter estimation	83
4.1.2	Hypothesis testing	84
4.2	Example 6: how many lines are there?	85
4.2.1	An algorithm	89
4.2.2	Simulated data	91
4.2.3	Real data	93
4.3	Other examples: means, variance, dating and so on	94
4.3.1	The analysis of means and variance	94
4.3.2	Luminescence dating	98
4.3.3	Interlude: what not to compute	100
5.	Assigning probabilities	103
5.1	Ignorance: indifference and transformation groups	103
5.1.1	The binomial distribution	107
5.1.2	Location and scale parameters	108
5.2	Testable information: the principle of maximum entropy	110
5.2.1	The monkey argument	113
5.2.2	The Lebesgue measure	115
5.3	MaxEnt examples: some common pdfs	117
5.3.1	Averages and exponentials	117
5.3.2	Variance and the Gaussian distribution	118
5.3.3	MaxEnt and the binomial distribution	120
5.3.4	Counting and Poisson statistics	121
5.4	Approximations: interconnections and simplifications	121
5.5	Hangups: priors versus likelihoods	124
5.5.1	Improper pdfs	124
5.5.2	Conjugate and reference priors	125
 PART II ADVANCED TOPICS		
6.	Non-parametric estimation	129
6.1	Introduction: free-form solutions	129
6.1.1	Singular value decomposition	130
6.1.2	A parametric free-form solution?	135
6.2	MaxEnt: images, monkeys and a non-uniform prior	136
6.2.1	Regularization	138
6.3	Smoothness: fuzzy pixels and spatial correlations	140
6.3.1	Interpolation	141
6.4	Generalizations: some extensions and comments	142

6.4.1	Summary of the basic strategy	144
6.4.2	Inference or inversion?	145
6.4.3	Advanced examples	148
7.	Experimental design	149
7.1	Introduction: general issues	149
7.2	Example 7: optimizing resolution functions	151
7.2.1	An isolated sharp peak	152
7.2.2	A free-form solution	156
7.3	Calibration, model selection and binning	161
7.4	Information gain: quantifying the worth of an experiment	163
8.	Least-squares extensions	165
8.1	Introduction: constraints and restraints	165
8.2	Noise scaling: a simple global adjustment	166
8.3	Outliers: dealing with erratic data	167
8.3.1	A conservative formulation	168
8.3.2	The good-and-bad data model	171
8.3.3	The Cauchy formulation	172
8.4	Background removal	173
8.5	Correlated noise: avoiding over-counting	174
8.5.1	Nearest-neighbour correlations	175
8.5.2	An elementary example	176
8.5.3	Time series	177
8.6	Log-normal: least-squares for magnitude data	179
9.	Nested sampling	181
9.1	Introduction: the computational problem	181
9.1.1	Evidence and posterior	182
9.2	Nested sampling: the basic idea	184
9.2.1	Iterating a sequence of objects	185
9.2.2	Terminating the iterations	186
9.2.3	Numerical uncertainty of computed results	187
9.2.4	Programming nested sampling in 'C'	188
9.3	Generating a new object by random sampling	190
9.3.1	Markov chain Monte Carlo (MCMC) exploration	191
9.3.2	Programming the lighthouse problem in 'C'	192
9.4	Monte Carlo sampling of the posterior	195
9.4.1	Posterior distribution	196
9.4.2	Equally-weighted posterior samples: staircase sampling	197
9.4.3	The lighthouse posterior	198
9.4.4	Metropolis exploration of the posterior	199
9.5	How many objects are needed?	200
9.5.1	Bi-modal likelihood with a single 'gate'	200
9.5.2	Multi-modal likelihoods with several 'gates'	201

9.6	Simulated annealing	203
9.6.1	The problem of phase changes	203
9.6.2	Example: order/disorder in a pseudo-crystal	204
9.6.3	Programming the pseudo-crystal in 'C'	206
10.	Quantification	209
10.1	Exploring an intrinsically non-uniform prior	209
10.1.1	Binary trees for controlling MCMC transitions	210
10.2	Example: ON/OFF switching	212
10.2.1	The master engine: flipping switches individually	212
10.2.2	Programming which components are present	212
10.2.3	Another engine: exchanging neighbouring switches	215
10.2.4	The control of multiple engines	216
10.3	Estimating quantities	216
10.3.1	Programming the estimation of quantities in 'C'	218
10.4	Final remarks	223
A.	Gaussian integrals	224
A.1	The univariate case	224
A.2	The bivariate extension	225
A.3	The multivariate generalization	226
B.	Cox's derivation of probability	229
B.1	Lemma 1: associativity equation	232
B.2	Lemma 2: negation	235
	Bibliography	237
	Index	241