

Contents

		page	
	<i>Editor's foreword</i>		xvii
	<i>Preface</i>		xix
Part I	Principles and elementary applications		
	1 Plausible reasoning		3
	1.1 Deductive and plausible reasoning		3
	1.2 Analogies with physical theories		6
	1.3 The thinking computer		7
	1.4 Introducing the robot		8
	1.5 Boolean algebra		9
	1.6 Adequate sets of operations		12
	1.7 The basic desiderata		17
	1.8 Comments		19
	1.8.1 Common language vs. formal logic		21
	1.8.2 Nitpicking		23
	2 The quantitative rules		24
	2.1 The product rule		24
	2.2 The sum rule		30
	2.3 Qualitative properties		35
	2.4 Numerical values		37
	2.5 Notation and finite-sets policy		43
	2.6 Comments		44
	2.6.1 'Subjective' vs. 'objective'		44
	2.6.2 Gödel's theorem		45
	2.6.3 Venn diagrams		47
	2.6.4 The 'Kolmogorov axioms'		49
	3 Elementary sampling theory		51
	3.1 Sampling without replacement		52
	3.2 Logic vs. propensity		60
	3.3 Reasoning from less precise information		64
	3.4 Expectations		66
	3.5 Other forms and extensions		68

	3.6	Probability as a mathematical tool	68
	3.7	The binomial distribution	69
	3.8	Sampling with replacement	72
	3.8.1	Digression: a sermon on reality vs. models	73
	3.9	Correction for correlations	75
	3.10	Simplification	81
	3.11	Comments	82
	3.11.1	A look ahead	84
	4	Elementary hypothesis testing	86
	4.1	Prior probabilities	87
	4.2	Testing binary hypotheses with binary data	90
iiix	4.3	Nonextensibility beyond the binary case	97
xix	4.4	Multiple hypothesis testing	98
	4.4.1	Digression on another derivation	101
3	4.5	Continuous probability distribution functions	107
3	4.6	Testing an infinite number of hypotheses	109
6	4.6.1	Historical digression	112
7	4.7	Simple and compound (or composite) hypotheses	115
8	4.8	Comments	116
9	4.8.1	Etymology	116
12	4.8.2	What have we accomplished?	117
17	5	Queer uses for probability theory	119
19	5.1	Extrasensory perception	119
21	5.2	Mrs Stewart's telepathic powers	120
23	5.2.1	Digression on the normal approximation	122
24	5.2.2	Back to Mrs Stewart	122
24	5.3	Converging and diverging views	126
30	5.4	Visual perception – evolution into Bayesianity?	132
32	5.5	The discovery of Neptune	133
37	5.5.1	Digression on alternative hypotheses	135
43	5.5.2	Back to Newton	137
44	5.6	Horse racing and weather forecasting	140
44	5.6.1	Discussion	142
45	5.7	Paradoxes of intuition	143
47	5.8	Bayesian jurisprudence	144
49	5.9	Comments	146
51	5.9.1	What is queer?	148
52	6	Elementary parameter estimation	149
60	6.1	Inversion of the urn distributions	149
64	6.2	Both N and R unknown	150
66	6.3	Uniform prior	152
68	6.4	Predictive distributions	154

6.5	Truncated uniform priors	157
6.6	A concave prior	158
6.7	The binomial monkey prior	160
6.8	Metamorphosis into continuous parameter estimation	163
6.9	Estimation with a binomial sampling distribution	163
6.9.1	Digression on optional stopping	166
6.10	Compound estimation problems	167
6.11	A simple Bayesian estimate: quantitative prior information	168
6.11.1	From posterior distribution function to estimate	172
6.12	Effects of qualitative prior information	177
6.13	Choice of a prior	178
6.14	On with the calculation!	179
6.15	The Jeffreys prior	181
6.16	The point of it all	183
6.17	Interval estimation	186
6.18	Calculation of variance	186
6.19	Generalization and asymptotic forms	188
6.20	Rectangular sampling distribution	190
6.21	Small samples	192
6.22	Mathematical trickery	193
6.23	Comments	195
7	The central, Gaussian or normal distribution	198
7.1	The gravitating phenomenon	199
7.2	The Herschel–Maxwell derivation	200
7.3	The Gauss derivation	202
7.4	Historical importance of Gauss’s result	203
7.5	The Landon derivation	205
7.6	Why the ubiquitous use of Gaussian distributions?	207
7.7	Why the ubiquitous success?	210
7.8	What estimator should we use?	211
7.9	Error cancellation	213
7.10	The near irrelevance of sampling frequency distributions	215
7.11	The remarkable efficiency of information transfer	216
7.12	Other sampling distributions	218
7.13	Nuisance parameters as safety devices	219
7.14	More general properties	220
7.15	Convolution of Gaussians	221
7.16	The central limit theorem	222
7.17	Accuracy of computations	224
7.18	Galton’s discovery	227
7.19	Population dynamics and Darwinian evolution	229
7.20	Evolution of humming-birds and flowers	231

157	7.21	Application to economics	233
158	7.22	The great inequality of Jupiter and Saturn	234
160	7.23	Resolution of distributions into Gaussians	235
163	7.24	Hermite polynomial solutions	236
163	7.25	Fourier transform relations	238
166	7.26	There is hope after all	239
167	7.27	Comments	240
168	7.27.1	Terminology again	240
172	8	Sufficiency, ancillarity, and all that	243
177	8.1	Sufficiency	243
178	8.2	Fisher sufficiency	245
179	8.2.1	Examples	246
181	8.2.2	The Blackwell–Rao theorem	247
183	8.3	Generalized sufficiency	248
186	8.4	Sufficiency plus nuisance parameters	249
186	8.5	The likelihood principle	250
188	8.6	Ancillarity	253
190	8.7	Generalized ancillary information	254
192	8.8	Asymptotic likelihood: Fisher information	256
193	8.9	Combining evidence from different sources	257
195	8.10	Pooling the data	260
198	8.10.1	Fine-grained propositions	261
199	8.11	Sam's broken thermometer	262
200	8.12	Comments	264
202	8.12.1	The fallacy of sample re-use	264
203	8.12.2	A folk theorem	266
205	8.12.3	Effect of prior information	267
207	8.12.4	Clever tricks and gamesmanship	267
210	9	Repetitive experiments: probability and frequency	270
211	9.1	Physical experiments	271
213	9.2	The poorly informed robot	274
215	9.3	Induction	276
216	9.4	Are there general inductive rules?	277
218	9.5	Multiplicity factors	280
219	9.6	Partition function algorithms	281
220	9.6.1	Solution by inspection	282
221	9.7	Entropy algorithms	285
222	9.8	Another way of looking at it	289
224	9.9	Entropy maximization	290
227	9.10	Probability and frequency	292
229	9.11	Significance tests	293
231	9.11.1	Implied alternatives	296

397	9.12	Comparison of psi and chi-squared	300
397	9.13	The chi-squared test	302
398	9.14	Generalization	304
400	9.15	Halley's mortality table	305
402	9.16	Comments	310
402	9.16.1	The irrationalists	310
404	9.16.2	Superstitions	312
406	10	Physics of 'random experiments'	314
410	10.1	An interesting correlation	314
412	10.2	Historical background	315
415	10.3	How to cheat at coin and die tossing	317
417	10.3.1	Experimental evidence	320
418	10.4	Bridge hands	321
418	10.5	General random experiments	324
421	10.6	Induction revisited	326
423	10.7	But what about quantum theory?	327
423	10.8	Mechanics under the clouds	329
424	10.9	More on coins and symmetry	331
426	10.10	Independence of tosses	335
426	10.11	The arrogance of the uninformed	338
	Part II	Advanced applications	
	11	Discrete prior probabilities: the entropy principle	343
430	11.1	A new kind of prior information	343
432	11.2	Minimum $\sum p_i^2$	345
437	11.3	Entropy: Shannon's theorem	346
438	11.4	The Wallis derivation	351
439	11.5	An example	354
440	11.6	Generalization: a more rigorous proof	355
443	11.7	Formal properties of maximum entropy distributions	358
445	11.8	Conceptual problems – frequency correspondence	365
449	11.9	Comments	370
451	12	Ignorance priors and transformation groups	372
451	12.1	What are we trying to do?	372
452	12.2	Ignorance priors	374
453	12.3	Continuous distributions	374
456	12.4	Transformation groups	378
459	12.4.1	Location and scale parameters	378
464	12.4.2	A Poisson rate	382
467	12.4.3	Unknown probability for success	382
470	12.4.4	Bertrand's problem	386
474	12.5	Comments	394

002	13	Decision theory, historical background	397
302	13.1	Inference vs. decision	397
304	13.2	Daniel Bernoulli's suggestion	398
305	13.3	The rationale of insurance	400
310	13.4	Entropy and utility	402
310	13.5	The honest weatherman	402
312	13.6	Reactions to Daniel Bernoulli and Laplace	404
314	13.7	Wald's decision theory	406
314	13.8	Parameter estimation for minimum loss	410
315	13.9	Reformulation of the problem	412
317	13.10	Effect of varying loss functions	415
320	13.11	General decision theory	417
321	13.12	Comments	418
324	13.12.1	'Objectivity' of decision theory	418
326	13.12.2	Loss functions in human society	421
327	13.12.3	A new look at the Jeffreys prior	423
329	13.12.4	Decision theory is not fundamental	423
331	13.12.5	Another dimension?	424
332	14	Simple applications of decision theory	426
338	14.1	Definitions and preliminaries	426
343	14.2	Sufficiency and information	428
343	14.3	Loss functions and criteria of optimum performance	430
345	14.4	A discrete example	432
346	14.5	How would our robot do it?	437
351	14.6	Historical remarks	438
354	14.6.1	The classical matched filter	439
355	14.7	The widget problem	440
358	14.7.1	Solution for Stage 2	443
362	14.7.2	Solution for Stage 3	445
370	14.7.3	Solution for Stage 4	449
372	14.8	Comments	450
372	15	Paradoxes of probability theory	451
374	15.1	How do paradoxes survive and grow?	451
374	15.2	Summing a series the easy way	452
378	15.3	Nonconglomerability	453
378	15.4	The tumbling tetrahedra	456
382	15.5	Solution for a finite number of tosses	459
382	15.6	Finite vs. countable additivity	464
382	15.7	The Borel–Kolmogorov paradox	467
386	15.8	The marginalization paradox	470
394	15.8.1	On to greater disasters	474

15.9	Discussion	478
15.9.1	The DSZ Example #5	480
15.9.2	Summary	483
15.10	A useful result after all?	484
15.11	How to mass-produce paradoxes	485
15.12	Comments	486
16	Orthodox methods: historical background	490
16.1	The early problems	490
16.2	Sociology of orthodox statistics	492
16.3	Ronald Fisher, Harold Jeffreys, and Jerzy Neyman	493
16.4	Pre-data and post-data considerations	499
16.5	The sampling distribution for an estimator	500
16.6	Pro-causal and anti-causal bias	503
16.7	What is real, the probability or the phenomenon?	505
16.8	Comments	506
16.8.1	Communication difficulties	507
17	Principles and pathology of orthodox statistics	509
17.1	Information loss	510
17.2	Unbiased estimators	511
17.3	Pathology of an unbiased estimate	516
17.4	The fundamental inequality of the sampling variance	518
17.5	Periodicity: the weather in Central Park	520
17.5.1	The folly of pre-filtering data	521
17.6	A Bayesian analysis	527
17.7	The folly of randomization	531
17.8	Fisher: common sense at Rothamsted	532
17.8.1	The Bayesian safety device	532
17.9	Missing data	533
17.10	Trend and seasonality in time series	534
17.10.1	Orthodox methods	535
17.10.2	The Bayesian method	536
17.10.3	Comparison of Bayesian and orthodox estimates	540
17.10.4	An improved orthodox estimate	541
17.10.5	The orthodox criterion of performance	544
17.11	The general case	545
17.12	Comments	550
18	The A_p distribution and rule of succession	553
18.1	Memory storage for old robots	553
18.2	Relevance	555
18.3	A surprising consequence	557
18.4	Outer and inner robots	559

18.5	An application	561
18.6	Laplace's rule of succession	563
18.7	Jeffreys' objection	566
18.8	Bass or carp?	567
18.9	So where does this leave the rule?	568
18.10	Generalization	568
18.11	Confirmation and weight of evidence	571
18.11.1	Is indifference based on knowledge or ignorance?	573
18.12	Carnap's inductive methods	574
18.13	Probability and frequency in exchangeable sequences	576
18.14	Prediction of frequencies	576
18.15	One-dimensional neutron multiplication	579
18.15.1	The frequentist solution	579
18.15.2	The Laplace solution	581
18.16	The de Finetti theorem	586
18.17	Comments	588
19	Physical measurements	589
19.1	Reduction of equations of condition	589
19.2	Reformulation as a decision problem	592
19.2.1	Sermon on Gaussian error distributions	592
19.3	The underdetermined case: K is singular	594
19.4	The overdetermined case: K can be made nonsingular	595
19.5	Numerical evaluation of the result	596
19.6	Accuracy of the estimates	597
19.7	Comments	599
19.7.1	A paradox	599
20	Model comparison	601
20.1	Formulation of the problem	602
20.2	The fair judge and the cruel realist	603
20.2.1	Parameters known in advance	604
20.2.2	Parameters unknown	604
20.3	But where is the idea of simplicity?	605
20.4	An example: linear response models	607
20.4.1	Digression: the old sermon still another time	608
20.5	Comments	613
20.5.1	Final causes	614
21	Outliers and robustness	615
21.1	The experimenter's dilemma	615
21.2	Robustness	617
21.3	The two-model model	619
21.4	Exchangeable selection	620
21.5	The general Bayesian solution	622

21.6	Pure outliers	624
21.7	One receding datum	625
22	Introduction to communication theory	627
22.1	Origins of the theory	627
22.2	The noiseless channel	628
22.3	The information source	634
22.4	Does the English language have statistical properties?	636
22.5	Optimum encoding: letter frequencies known	638
22.6	Better encoding from knowledge of digram frequencies	641
22.7	Relation to a stochastic model	644
22.8	The noisy channel	648
Appendix A	Other approaches to probability theory	651
A.1	The Kolmogorov system of probability	651
A.2	The de Finetti system of probability	655
A.3	Comparative probability	656
A.4	Holdouts against universal comparability	658
A.5	Speculations about lattice theories	659
Appendix B	Mathematical formalities and style	661
B.1	Notation and logical hierarchy	661
B.2	Our 'cautious approach' policy	662
B.3	Willy Feller on measure theory	663
B.4	Kronecker vs. Weierstrasz	665
B.5	What is a legitimate mathematical function?	666
B.5.1	Delta-functions	668
B.5.2	Nondifferentiable functions	668
B.5.3	Bogus nondifferentiable functions	669
B.6	Counting infinite sets?	671
B.7	The Hausdorff sphere paradox and mathematical diseases	672
B.8	What am I supposed to publish?	674
B.9	Mathematical courtesy	675
Appendix C	Convolutions and cumulants	677
C.1	Relation of cumulants and moments	679
C.2	Examples	680
	References	683
	Bibliography	705
	Author index	721
	Subject index	724