Contents

Editor's foreword .............................. xvii
Preface ........................................... 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
## Contents

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

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

9.12 Comparison of psi and chi-squared 300
9.13 The chi-squared test 302
9.14 Generalization 304
9.15 Halley’s mortality table 305
9.16 Comments 310
  9.16.1 The irrationalists 310
  9.16.2 Superstitions 312

10 Physics of ‘random experiments’ 314
  10.1 An interesting correlation 314
  10.2 Historical background 315
  10.3 How to cheat at coin and die tossing 317
    10.3.1 Experimental evidence 320
  10.4 Bridge hands 321
  10.5 General random experiments 324
  10.6 Induction revisited 326
  10.7 But what about quantum theory? 327
  10.8 Mechanics under the clouds 329
  10.9 More on coins and symmetry 331
  10.10 Independence of tosses 335
  10.11 The arrogance of the uninformed 338

Part II Advanced applications

11 Discrete prior probabilities: the entropy principle 343
  11.1 A new kind of prior information 343
  11.2 Minimum $\sum p_i^2$ 345
  11.3 Entropy: Shannon’s theorem 346
  11.4 The Wallis derivation 351
  11.5 An example 354
  11.6 Generalization: a more rigorous proof 355
  11.7 Formal properties of maximum entropy distributions 358
  11.8 Conceptual problems – frequency correspondence 365
  11.9 Comments 370

12 Ignorance priors and transformation groups 372
  12.1 What are we trying to do? 372
  12.2 Ignorance priors 374
  12.3 Continuous distributions 374
  12.4 Transformation groups 378
    12.4.1 Location and scale parameters 378
    12.4.2 A Poisson rate 382
    12.4.3 Unknown probability for success 382
    12.4.4 Bertrand’s problem 386
  12.5 Comments 394
13 Decision theory, historical background  397
  13.1 Inference vs. decision  397
  13.2 Daniel Bernoulli's suggestion  398
  13.3 The rationale of insurance  400
  13.4 Entropy and utility  402
  13.5 The honest weatherman  402
  13.6 Reactions to Daniel Bernoulli and Laplace  404
  13.7 Wald's decision theory  406
  13.8 Parameter estimation for minimum loss  410
  13.9 Reformulation of the problem  412
  13.10 Effect of varying loss functions  415
  13.11 General decision theory  417
  13.12 Comments  418
    13.12.1 'Objectivity' of decision theory  418
    13.12.2 Loss functions in human society  421
    13.12.3 A new look at the Jeffreys prior  423
    13.12.4 Decision theory is not fundamental  423
    13.12.5 Another dimension?  424
14 Simple applications of decision theory  426
  14.1 Definitions and preliminaries  426
  14.2 Sufficiency and information  428
  14.3 Loss functions and criteria of optimum performance  430
  14.4 A discrete example  432
  14.5 How would our robot do it?  437
  14.6 Historical remarks  438
    14.6.1 The classical matched filter  439
  14.7 The widget problem  440
    14.7.1 Solution for Stage 2  443
    14.7.2 Solution for Stage 3  445
    14.7.3 Solution for Stage 4  449
  14.8 Comments  450
15 Paradoxes of probability theory  451
  15.1 How do paradoxes survive and grow?  451
  15.2 Summing a series the easy way  452
  15.3 Nonconglomerability  453
  15.4 The tumbling tetrahedra  456
  15.5 Solution for a finite number of tosses  459
  15.6 Finite vs. countable additivity  464
  15.7 The Borel–Kolmogorov paradox  467
  15.8 The marginalization paradox  470
    15.8.1 On to greater disasters  474
Contents

15.9 Discussion
15.9.1 The DSZ Example #5
15.9.2 Summary
15.10 A useful result after all?
15.11 How to mass-produce paradoxes
15.12 Comments

16 Orthodox methods: historical background
16.1 The early problems
16.2 Sociology of orthodox statistics
16.3 Ronald Fisher, Harold Jeffreys, and Jerzy Neyman
16.4 Pre-data and post-data considerations
16.5 The sampling distribution for an estimator
16.6 Pro-causal and anti-causal bias
16.7 What is real, the probability or the phenomenon?
16.8 Comments
16.8.1 Communication difficulties

17 Principles and pathology of orthodox statistics
17.1 Information loss
17.2 Unbiased estimators
17.3 Pathology of an unbiased estimate
17.4 The fundamental inequality of the sampling variance
17.5 Periodicity: the weather in Central Park
17.5.1 The folly of pre-filtering data
17.6 A Bayesian analysis
17.7 The folly of randomization
17.8 Fisher: common sense at Rothamsted
17.8.1 The Bayesian safety device
17.9 Missing data
17.10 Trend and seasonality in time series
17.10.1 Orthodox methods
17.10.2 The Bayesian method
17.10.3 Comparison of Bayesian and orthodox estimates
17.10.4 An improved orthodox estimate
17.10.5 The orthodox criterion of performance
17.11 The general case
17.12 Comments

18 The $A_p$ distribution and rule of succession
18.1 Memory storage for old robots
18.2 Relevance
18.3 A surprising consequence
18.4 Outer and inner robots
Contents

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 597
19.6 Accuracy of the estimates 599
19.7 Comments 601
  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 610
  20.5.1 Final causes 610

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