## Contents

*Editor's foreword*  
*Preface*  

**Part I  Principles and elementary applications**  
1 Plausible reasoning  
   1.1 Deductive and plausible reasoning  
   1.2 Analogies with physical theories  
   1.3 The thinking computer  
   1.4 Introducing the robot  
   1.5 Boolean algebra  
   1.6 Adequate sets of operations  
   1.7 The basic desiderata  
   1.8 Comments  
   1.8.1 Common language vs. formal logic  
   1.8.2 Nitpicking  
2 The quantitative rules  
   2.1 The product rule  
   2.2 The sum rule  
   2.3 Qualitative properties  
   2.4 Numerical values  
   2.5 Notation and finite-sets policy  
   2.6 Comments  
   2.6.1 ‘Subjective’ vs. ‘objective’  
   2.6.2 Gödel’s theorem  
   2.6.3 Venn diagrams  
   2.6.4 The ‘Kolmogorov axioms’  
3 Elementary sampling theory  
   3.1 Sampling without replacement  
   3.2 Logic vs. propensity  
   3.3 Reasoning from less precise information  
   3.4 Expectations  
   3.5 Other forms and extensions

*page xvii*  
*xix*
Contents

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

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.12</td>
<td>Comparison of psi and chi-squared</td>
<td>300</td>
</tr>
<tr>
<td>9.13</td>
<td>The chi-squared test</td>
<td>302</td>
</tr>
<tr>
<td>9.14</td>
<td>Generalization</td>
<td>304</td>
</tr>
<tr>
<td>9.15</td>
<td>Halley’s mortality table</td>
<td>305</td>
</tr>
<tr>
<td>9.16</td>
<td>Comments</td>
<td>310</td>
</tr>
<tr>
<td>9.16.1</td>
<td>The irrationalists</td>
<td>310</td>
</tr>
<tr>
<td>9.16.2</td>
<td>Superstitions</td>
<td>312</td>
</tr>
<tr>
<td>10</td>
<td>Physics of ‘random experiments’</td>
<td>314</td>
</tr>
<tr>
<td>10.1</td>
<td>An interesting correlation</td>
<td>314</td>
</tr>
<tr>
<td>10.2</td>
<td>Historical background</td>
<td>315</td>
</tr>
<tr>
<td>10.3</td>
<td>How to cheat at coin and die tossing</td>
<td>317</td>
</tr>
<tr>
<td>10.3.1</td>
<td>Experimental evidence</td>
<td>320</td>
</tr>
<tr>
<td>10.4</td>
<td>Bridge hands</td>
<td>321</td>
</tr>
<tr>
<td>10.5</td>
<td>General random experiments</td>
<td>324</td>
</tr>
<tr>
<td>10.6</td>
<td>Induction revisited</td>
<td>326</td>
</tr>
<tr>
<td>10.7</td>
<td>But what about quantum theory?</td>
<td>327</td>
</tr>
<tr>
<td>10.8</td>
<td>Mechanics under the clouds</td>
<td>329</td>
</tr>
<tr>
<td>10.9</td>
<td>More on coins and symmetry</td>
<td>331</td>
</tr>
<tr>
<td>10.10</td>
<td>Independence of tosses</td>
<td>335</td>
</tr>
<tr>
<td>10.11</td>
<td>The arrogance of the uninformed</td>
<td>338</td>
</tr>
<tr>
<td>11</td>
<td>Discrete prior probabilities: the entropy principle</td>
<td>343</td>
</tr>
<tr>
<td>11.1</td>
<td>A new kind of prior information</td>
<td>343</td>
</tr>
<tr>
<td>11.2</td>
<td>Minimum $\sum p_i^2$</td>
<td>345</td>
</tr>
<tr>
<td>11.3</td>
<td>Entropy: Shannon’s theorem</td>
<td>346</td>
</tr>
<tr>
<td>11.4</td>
<td>The Wallis derivation</td>
<td>351</td>
</tr>
<tr>
<td>11.5</td>
<td>An example</td>
<td>354</td>
</tr>
<tr>
<td>11.6</td>
<td>Generalization: a more rigorous proof</td>
<td>355</td>
</tr>
<tr>
<td>11.7</td>
<td>Formal properties of maximum entropy distributions</td>
<td>358</td>
</tr>
<tr>
<td>11.8</td>
<td>Conceptual problems – frequency correspondence</td>
<td>365</td>
</tr>
<tr>
<td>11.9</td>
<td>Comments</td>
<td>370</td>
</tr>
<tr>
<td>12</td>
<td>Ignorance priors and transformation groups</td>
<td>372</td>
</tr>
<tr>
<td>12.1</td>
<td>What are we trying to do?</td>
<td>372</td>
</tr>
<tr>
<td>12.2</td>
<td>Ignorance priors</td>
<td>374</td>
</tr>
<tr>
<td>12.3</td>
<td>Continuous distributions</td>
<td>374</td>
</tr>
<tr>
<td>12.4</td>
<td>Transformation groups</td>
<td>378</td>
</tr>
<tr>
<td>12.4.1</td>
<td>Location and scale parameters</td>
<td>378</td>
</tr>
<tr>
<td>12.4.2</td>
<td>A Poisson rate</td>
<td>382</td>
</tr>
<tr>
<td>12.4.3</td>
<td>Unknown probability for success</td>
<td>382</td>
</tr>
<tr>
<td>12.4.4</td>
<td>Bertrand’s problem</td>
<td>386</td>
</tr>
<tr>
<td>12.5</td>
<td>Comments</td>
<td>394</td>
</tr>
</tbody>
</table>
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
15.9 Discussion
   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
   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
   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
       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
   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
Contents

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