Contents

List of Figures xix

1 Introduction and Preview 1

1.1 Preview of the book / 5

2 Entropy, Relative Entropy and Mutual Information 12

2.1 Entropy / 12
2.2 Joint entropy and conditional entropy / 15
2.3 Relative entropy and mutual information / 18
2.4 Relationship between entropy and mutual information / 19
2.5 Chain rules for entropy, relative entropy and mutual information / 21
2.6 Jensen's inequality and its consequences / 23
2.7 The log sum inequality and its applications / 29
2.8 Data processing inequality / 32
2.9 The second law of thermodynamics / 33
2.10 Sufficient statistics / 36
2.11 Fano's inequality / 38
Summary of Chapter 2 / 40
Problems for Chapter 2 / 42
Historical notes / 49

3 The Asymptotic Equipartition Property 50

3.1 The AEP / 51

3.2 Consequences of the AEP: data compression / 53
3.3 High probability sets and the typical set / 55
Summary of Chapter 3 / 56
Problems for Chapter 3 / 57
Historical notes / 59

4 Entropy Rates of a Stochastic Process 60

4.1 Markov chains / 60
4.2 Entropy rate / 63
4.3 Example: Entropy rate of a random walk on a weighted graph / 66
4.4 Hidden Markov models / 69
Summary of Chapter 4 / 71
Problems for Chapter 4 / 72
Historical notes / 77

5 Data Compression 78

5.1 Examples of codes / 79
5.2 Kraft inequality / 82
5.3 Optimal codes / 84
5.4 Bounds on the optimal codelength / 87
5.5 Kraft inequality for uniquely decodable codes / 90
5.6 Huffman codes / 92
5.7 Some comments on Huffman codes / 94
5.8 Optimality of Huffman codes / 97
5.9 Shannon-Fano-Elias coding / 101
5.10 Arithmetic coding / 104
5.11 Competitive optimality of the Shannon code / 107
5.12 Generation of discrete distributions from fair coins / 110
Summary of Chapter 5 / 117
Problems for Chapter 5 / 118
Historical notes / 124

6 Gambling and Data Compression 125

6.1 The horse race / 125
6.2 Gambling and side information / 130
6.3 Dependent horse races and entropy rate / 131
6.4 The entropy of English / 133
6.5 Data compression and gambling / 136

6.6 Gambling estimate of the entropy of English / 138
Summary of Chapter 6 / 140
Problems for Chapter 6 / 141
Historical notes / 143

7 Kolmogorov Complexity **144**

7.1 Models of computation / 146
7.2 Kolmogorov complexity: definitions and examples / 147
7.3 Kolmogorov complexity and entropy / 153
7.4 Kolmogorov complexity of integers / 155
7.5 Algorithmically random and incompressible sequences / 156
7.6 Universal probability / 160
7.7 The halting problem and the non-computability of Kolmogorov complexity / 162
7.8 Ω / 164
7.9 Universal gambling / 166
7.10 Occam's razor / 168
7.11 Kolmogorov complexity and universal probability / 169
7.12 The Kolmogorov sufficient statistic / 175
Summary of Chapter 7 / 178
Problems for Chapter 7 / 180
Historical notes / 182

8 Channel Capacity **183**

8.1 Examples of channel capacity / 184
8.2 Symmetric channels / 189
8.3 Properties of channel capacity / 190
8.4 Preview of the channel coding theorem / 191
8.5 Definitions / 192
8.6 Jointly typical sequences / 194
8.7 The channel coding theorem / 198
8.8 Zero-error codes / 203
8.9 Fano's inequality and the converse to the coding theorem / 204
8.10 Equality in the converse to the channel coding theorem / 207
8.11 Hamming codes / 209
8.12 Feedback capacity / 212

8.13 The joint source channel coding theorem / 215
Summary of Chapter 8 / 218
Problems for Chapter 8 / 220
Historical notes / 222

9 Differential Entropy **224**

9.1 Definitions / 224
9.2 The AEP for continuous random variables / 225
9.3 Relation of differential entropy to discrete entropy / 228
9.4 Joint and conditional differential entropy / 229
9.5 Relative entropy and mutual information / 231
9.6 Properties of differential entropy, relative entropy and mutual information / 232
9.7 Differential entropy bound on discrete entropy / 234
Summary of Chapter 9 / 236
Problems for Chapter 9 / 237
Historical notes / 238

10 The' Gaussian Channel **239**

10.1 The Gaussian channel: definitions / 241
10.2 Converse to the coding theorem for Gaussian channels / 245
10.3 Band-limited channels / 247
10.4 Parallel Gaussian channels / 250
10.5 Channels with colored Gaussian noise / 253
10.6 Gaussian channels with feedback / 256
Summary of Chapter 10 / 262
Problems for Chapter 10 / 263
Historical notes / 264

11 Maximum Entropy and Spectral Estimation **266**

11.1 Maximum entropy distributions / 266
11.2 Examples / 268
11.3 An anomalous maximum entropy problem / 270
11.4 Spectrum estimation / 272
11.5 Entropy rates of a Gaussian process / 273
11.6 Burg's maximum entropy theorem / 274
Summary of Chapter 11 / 277
Problems for Chapter 11 / 277
Historical notes / 278

12 Information Theory and Statistics *279*

12.1 The method of types / 279
12.2 The law of large numbers / 286
12.3 Universal source coding / 288
12.4 Large deviation theory / 291
12.5 Examples of Sanov's theorem / 294
12.6 The conditional limit theorem / 297
12.7 Hypothesis testing / 304
12.8 Stein's lemma / 309
12.9 Chernoff bound / 312
12.10 Lempel-Ziv coding / 319
12.11 Fisher information and the Cramer-Rao inequality / 326
Summary of Chapter 12 / 331
Problems for Chapter 12 / 333
Historical notes / 335

13 Rate Distortion Theory **336**

13.1 Quantization / 337
13.2 Definitions / 338
13.3 Calculation of the rate distortion function / 342
13.4 Converse to the rate distortion theorem / 349
13.5 Achievability of the rate distortion function / 351
13.6 Strongly typical sequences and rate distortion / 358
13.7 Characterization of the rate distortion function / 362
13.8 Computation of channel capacity and the rate distortion function / 367
Summary of Chapter 13 / 367
Problems for Chapter 13 / 368
Historical notes / 372

14 Network Information Theory **374**

14.1 Gaussian multiple user channels / 377
14.2 Jointly typical sequences / 384
14.3 The multiple access channel / 388
14.4 Encoding of correlated sources / 407
14.5 Duality between Slepian-Wolf encoding and multiple access channels / 416
14.6 The broadcast channel / 418
14.7 The relay channel / 428

14.8 Source coding with side information / 432
14.9 Rate distortion with side information / 438
14.10 General multiterminal networks / 444
Summary of Chapter 14 / 450
Problems for Chapter 14 / 452
Historical notes / 457

15 Information Theory and the Stock Market 459

15.1 The stock market: some definitions / 459
15.2 Kuhn-Tucker characterization of the log-optimal portfolio / 462
15.3 Asymptotic optimality of the log-optimal portfolio / 465
15.4 Side information and the doubling rate / 467
15.5 Investment in stationary markets / 469
15.6 Competitive optimality of the log-optimal portfolio / 471
15.7 The Shannon-McMillan-Breiman theorem / 474
Summary of Chapter 15 / 479
Problems for Chapter 15 / 480
Historical notes / 481

16 Inequalities in Information Theory 462

16.1 Basic inequalities of information theory / 482
16.2 Differential entropy / 485
16.3 Bounds on entropy and relative entropy / 488
16.4 Inequalities for types / 490
16.5 Entropy rates of subsets / 490
16.6 Entropy and Fisher information / 494
16.7 The entropy power inequality and the Brunn-Minkowski inequality / 497
16.8 Inequalities for determinants / 501
16.9 Inequalities for ratios of determinants / 505
Overall Summary / 508
Problems for Chapter 16 / 509
Historical notes / 509

Bibliography 510

List of Symbols 526

Index 529