
ELEMENTS OF INFORMATION THEORY

Second Edition

THOMAS M. COVER

JOY A. THOMAS

 **WILEY-
INTERSCIENCE**

A JOHN WILEY & SONS, INC., PUBLICATION

CONTENTS

Preface to the Second Edition	xv
Preface to the First Edition	xvii
Acknowledgments for the Second Edition	xxi
Acknowledgments for the First Edition	xxiii
1 Introduction and Preview	1
1.1 Preview of the Book	5
2 Entropy, Relative Entropy, and Mutual Information	13
2.1 Entropy	13
2.2 Joint Entropy and Conditional Entropy	16
2.3 Relative Entropy and Mutual Information	19
2.4 Relationship Between Entropy and Mutual Information	20
2.5 Chain Rules for Entropy, Relative Entropy, and Mutual Information	22
2.6 Jensen's Inequality and Its Consequences	25
2.7 Log Sum Inequality and Its Applications	30
2.8 Data-Processing Inequality	34
2.9 Sufficient Statistics	35
2.10 Fano's Inequality	37
Summary	41
Problems	43
Historical Notes	54

3	Asymptotic Equipartition Property	57
3.1	Asymptotic Equipartition Property Theorem	58
3.2	Consequences of the AEP: Data Compression	60
3.3	High-Probability Sets and the Typical Set	62
	Summary	64
	Problems	64
	Historical Notes	69
4	Entropy Rates of a Stochastic Process	71
4.1	Markov Chains	71
4.2	Entropy Rate	74
4.3	Example: Entropy Rate of a Random Walk on a Weighted Graph	78
4.4	Second Law of Thermodynamics	81
4.5	Functions of Markov Chains	84
	Summary	87
	Problems	88
	Historical Notes	100
5	Data Compression	103
5.1	Examples of Codes	103
5.2	Kraft Inequality	107
5.3	Optimal Codes	110
5.4	Bounds on the Optimal Code Length	112
5.5	Kraft Inequality for Uniquely Decodable Codes	115
5.6	Huffman Codes	118
5.7	Some Comments on Huffman Codes	120
5.8	Optimality of Huffman Codes	123
5.9	Shannon–Fano–Elias Coding	127
5.10	Competitive Optimality of the Shannon Code	130
5.11	Generation of Discrete Distributions from Fair Coins	134
	Summary	141
	Problems	142
	Historical Notes	157

6	Gambling and Data Compression	159
6.1	The Horse Race	159
6.2	Gambling and Side Information	164
6.3	Dependent Horse Races and Entropy Rate	166
6.4	The Entropy of English	168
6.5	Data Compression and Gambling	171
6.6	Gambling Estimate of the Entropy of English	173
	Summary	175
	Problems	176
	Historical Notes	182
7	Channel Capacity	183
7.1	Examples of Channel Capacity	184
7.1.1	Noiseless Binary Channel	184
7.1.2	Noisy Channel with Nonoverlapping Outputs	185
7.1.3	Noisy Typewriter	186
7.1.4	Binary Symmetric Channel	187
7.1.5	Binary Erasure Channel	188
7.2	Symmetric Channels	189
7.3	Properties of Channel Capacity	191
7.4	Preview of the Channel Coding Theorem	191
7.5	Definitions	192
7.6	Jointly Typical Sequences	195
7.7	Channel Coding Theorem	199
7.8	Zero-Error Codes	205
7.9	Fano's Inequality and the Converse to the Coding Theorem	206
7.10	Equality in the Converse to the Channel Coding Theorem	208
7.11	Hamming Codes	210
7.12	Feedback Capacity	216
7.13	Source-Channel Separation Theorem	218
	Summary	222
	Problems	223
	Historical Notes	240

8	Differential Entropy	243
8.1	Definitions	243
8.2	AEP for Continuous Random Variables	245
8.3	Relation of Differential Entropy to Discrete Entropy	247
8.4	Joint and Conditional Differential Entropy	249
8.5	Relative Entropy and Mutual Information	250
8.6	Properties of Differential Entropy, Relative Entropy, and Mutual Information	252
	Summary	256
	Problems	256
	Historical Notes	259
9	Gaussian Channel	261
9.1	Gaussian Channel: Definitions	263
9.2	Converse to the Coding Theorem for Gaussian Channels	268
9.3	Bandlimited Channels	270
9.4	Parallel Gaussian Channels	274
9.5	Channels with Colored Gaussian Noise	277
9.6	Gaussian Channels with Feedback	280
	Summary	289
	Problems	290
	Historical Notes	299
10	Rate Distortion Theory	301
10.1	Quantization	301
10.2	Definitions	303
10.3	Calculation of the Rate Distortion Function	307
	10.3.1 Binary Source	307
	10.3.2 Gaussian Source	310
	10.3.3 Simultaneous Description of Independent Gaussian Random Variables	312
10.4	Converse to the Rate Distortion Theorem	315
10.5	Achievability of the Rate Distortion Function	318
10.6	Strongly Typical Sequences and Rate Distortion	325
10.7	Characterization of the Rate Distortion Function	329

10.8	Computation of Channel Capacity and the Rate Distortion Function	332
	Summary	335
	Problems	336
	Historical Notes	345
11	Information Theory and Statistics	347
11.1	Method of Types	347
11.2	Law of Large Numbers	355
11.3	Universal Source Coding	357
11.4	Large Deviation Theory	360
11.5	Examples of Sanov's Theorem	364
11.6	Conditional Limit Theorem	366
11.7	Hypothesis Testing	375
11.8	Chernoff–Stein Lemma	380
11.9	Chernoff Information	384
11.10	Fisher Information and the Cramér–Rao Inequality	392
	Summary	397
	Problems	399
	Historical Notes	408
12	Maximum Entropy	409
12.1	Maximum Entropy Distributions	409
12.2	Examples	411
12.3	Anomalous Maximum Entropy Problem	413
12.4	Spectrum Estimation	415
12.5	Entropy Rates of a Gaussian Process	416
12.6	Burg's Maximum Entropy Theorem	417
	Summary	420
	Problems	421
	Historical Notes	425
13	Universal Source Coding	427
13.1	Universal Codes and Channel Capacity	428
13.2	Universal Coding for Binary Sequences	433
13.3	Arithmetic Coding	436

- 13.4 Lempel–Ziv Coding 440
 - 13.4.1 Sliding Window Lempel–Ziv Algorithm 441
 - 13.4.2 Tree-Structured Lempel–Ziv Algorithms 442
- 13.5 Optimality of Lempel–Ziv Algorithms 443
 - 13.5.1 Sliding Window Lempel–Ziv Algorithms 443
 - 13.5.2 Optimality of Tree-Structured Lempel–Ziv Compression 448
- Summary 456
- Problems 457
- Historical Notes 461

14 Kolmogorov Complexity 463

- 14.1 Models of Computation 464
- 14.2 Kolmogorov Complexity: Definitions and Examples 466
- 14.3 Kolmogorov Complexity and Entropy 473
- 14.4 Kolmogorov Complexity of Integers 475
- 14.5 Algorithmically Random and Incompressible Sequences 476
- 14.6 Universal Probability 480
- 14.7 Kolmogorov complexity 482
- 14.8 Ω 484
- 14.9 Universal Gambling 487
- 14.10 Occam’s Razor 488
- 14.11 Kolmogorov Complexity and Universal Probability 490
- 14.12 Kolmogorov Sufficient Statistic 496
- 14.13 Minimum Description Length Principle 500
- Summary 501
- Problems 503
- Historical Notes 507

15 Network Information Theory 509

- 15.1 Gaussian Multiple-User Channels 513

- 15.1.1 Single-User Gaussian Channel 513
- 15.1.2 Gaussian Multiple-Access Channel with m Users 514
- 15.1.3 Gaussian Broadcast Channel 515
- 15.1.4 Gaussian Relay Channel 516
- 15.1.5 Gaussian Interference Channel 518
- 15.1.6 Gaussian Two-Way Channel 519
- 15.2 Jointly Typical Sequences 520
- 15.3 Multiple-Access Channel 524
 - 15.3.1 Achievability of the Capacity Region for the Multiple-Access Channel 530
 - 15.3.2 Comments on the Capacity Region for the Multiple-Access Channel 532
 - 15.3.3 Convexity of the Capacity Region of the Multiple-Access Channel 534
 - 15.3.4 Converse for the Multiple-Access Channel 538
 - 15.3.5 m -User Multiple-Access Channels 543
 - 15.3.6 Gaussian Multiple-Access Channels 544
- 15.4 Encoding of Correlated Sources 549
 - 15.4.1 Achievability of the Slepian–Wolf Theorem 551
 - 15.4.2 Converse for the Slepian–Wolf Theorem 555
 - 15.4.3 Slepian–Wolf Theorem for Many Sources 556
 - 15.4.4 Interpretation of Slepian–Wolf Coding 557
- 15.5 Duality Between Slepian–Wolf Encoding and Multiple-Access Channels 558
- 15.6 Broadcast Channel 560
 - 15.6.1 Definitions for a Broadcast Channel 563
 - 15.6.2 Degraded Broadcast Channels 564
 - 15.6.3 Capacity Region for the Degraded Broadcast Channel 565
- 15.7 Relay Channel 571
- 15.8 Source Coding with Side Information 575
- 15.9 Rate Distortion with Side Information 580

15.10 General Multiterminal Networks 587

Summary 594

Problems 596

Historical Notes 609

16 Information Theory and Portfolio Theory **613**

16.1 The Stock Market: Some Definitions 613

16.2 Kuhn–Tucker Characterization of the Log-Optimal Portfolio 617

16.3 Asymptotic Optimality of the Log-Optimal Portfolio 619

16.4 Side Information and the Growth Rate 621

16.5 Investment in Stationary Markets 623

16.6 Competitive Optimality of the Log-Optimal Portfolio 627

16.7 Universal Portfolios 629

 16.7.1 Finite-Horizon Universal Portfolios 631

 16.7.2 Horizon-Free Universal Portfolios 638

16.8 Shannon–McMillan–Breiman Theorem
(General AEP) 644

Summary 650

Problems 652

Historical Notes 655

17 Inequalities in Information Theory **657**

17.1 Basic Inequalities of Information Theory 657

17.2 Differential Entropy 660

17.3 Bounds on Entropy and Relative Entropy 663

17.4 Inequalities for Types 665

17.5 Combinatorial Bounds on Entropy 666

17.6 Entropy Rates of Subsets 667

17.7 Entropy and Fisher Information 671

17.8 Entropy Power Inequality and Brunn–Minkowski
Inequality 674

17.9 Inequalities for Determinants 679

17.10 Inequalities for Ratios of Determinants	683
Summary	686
Problems	686
Historical Notes	687
Bibliography	689
List of Symbols	723
Index	727