

Elements of Information Theory

Thomas M. Cover, Joy A. Thomas

Copyright © 1991 John Wiley & Sons, Inc.

Print ISBN 0-471-06259-6 Online ISBN 0-471-20061-1

Elements of Information Theory

Elements of Information Theory
Thomas M. Cover, Joy A. Thomas
Copyright © 1991 John Wiley & Sons, Inc.
Print ISBN 0-471-06259-6 Online ISBN 0-471-20061-1

WILEY SERIES IN TELECOMMUNICATIONS

Donald L. Schilling, Editor
City College of New York

Digital Telephony, 2nd Edition
John Bellamy

Elements of Information Theory
Thomas M. Cover and Joy A. Thomas

Telecommunication System Engineering, 2nd Edition
Roger L. Freeman

Telecommunication Transmission Handbook, 3rd Edition
Roger L. Freeman

Introduction to Communications Engineering, 2nd Edition
Robert M. Gagliardi

Expert System Applications to Telecommunications
Jay Liebowitz

Synchronization in Digital Communications, Volume 1
Heinrich Meyr and Gerd Ascheid

Synchronization in Digital Communications, Volume 2
Heinrich Meyr and Gerd Ascheid (in preparation)

Computational Methods of Signal Recovery and Recognition
Richard J. Mammone (in preparation)

Business Earth Stations for Telecommunications
Walter L. Morgan and Denis Rouffet

Satellite Communications: The First Quarter Century of Service
David W. E. Rees

Worldwide Telecommunications Guide for the Business Manager
Walter L. Vignault

Elements of Information Theory

THOMAS M. COVER

*Stanford University
Stanford, California*

JOY A. THOMAS

*IBM T. J. Watson Research Center
Yorktown Heights, New York*



A Wiley-Interscience Publication

JOHN WILEY & SONS, INC.

New York / Chichester / Brisbane / Toronto / Singapore

Copyright © 1991 by John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic or mechanical, including uploading, downloading, printing, decompiling, recording or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional person should be sought.

ISBN 0-471-20061-1.

This title is also available in print as ISBN 0-471-06259-6

For more information about Wiley products, visit our web site at www.Wiley.com.

Library of Congress Cataloging in Publication Data:

Cover, T. M., 1938 —

Elements of Information theory / Thomas M. Cover, Joy A. Thomas.

p. cm. — (Wiley series in telecommunications)

“A Wiley-Interscience publication.”

Includes bibliographical references and index.

ISBN 0-471-06259-6

1. Information theory. I. Thomas, Joy A. II. Title.

III. Series.

Q360.C68 1991

003'.54 — dc20

90-45484

CIP

Printed in the United States of America

20 19 18 17 16 15 14 13

To my father
Tom Cover

To my parents
Joy Thomas

Preface

This is intended to be a simple and accessible book on information theory. As Einstein said, “*Everything should be made as simple as possible, but no simpler.*” Although we have not verified the quote (first found in a fortune cookie), this point of view drives our development throughout the book. There are a few key ideas and techniques that, when mastered, make the subject appear simple and provide great intuition on new questions.

This book has arisen from over ten years of lectures in a two-quarter sequence of a senior and first-year graduate level course in information theory, and is intended as an introduction to information theory for students of communication theory, computer science and statistics.

There are two points to be made about the simplicities inherent in information theory. First, certain quantities like entropy and mutual information arise as the answers to fundamental questions. For example, entropy is the minimum descriptive complexity of a random variable, and mutual information is the communication rate in the presence of noise. Also, as we shall point out, mutual information corresponds to the increase in the doubling rate of wealth given side information. Second, the answers to information theoretic questions have a natural algebraic structure. For example, there is a chain rule for entropies, and entropy and mutual information are related. Thus the answers to problems in data compression and communication admit extensive interpretation. We all know the feeling that follows when one investigates a problem, goes through a large amount of algebra and finally investigates the answer to find that the entire problem is illuminated, not by the analysis, but by the inspection of the answer. Perhaps the outstanding examples of this in physics are Newton’s laws and

Schrodinger's wave equation. Who could have foreseen the awesome philosophical interpretations of Schrödinger's wave equation?

In the text we often investigate properties of the answer before we look at the question. For example, in Chapter 2, we define entropy, relative entropy and mutual information and study the relationships and a few interpretations of them, showing how the answers fit together in various ways. Along the way we speculate on the meaning of the second law of thermodynamics. Does entropy always increase? The answer is yes and no. This is the sort of result that should please experts in the area but might be overlooked as standard by the novice.

In fact, that brings up a point that often occurs in teaching. It is fun to find new proofs or slightly new results that no one else knows. When one presents these ideas along with the established material in class, the response is "sure, sure, sure." But the excitement of teaching the material is greatly enhanced. Thus we have derived great pleasure from investigating a number of new ideas in this text book.

Examples of some of the new material in this text include the chapter on the relationship of information theory to gambling, the work on the universality of the second law of thermodynamics in the context of Markov chains, the joint typicality proofs of the channel capacity theorem, the competitive optimality of Huffman codes and the proof of Burg's theorem on maximum entropy spectral density estimation. Also the chapter on Kolmogorov complexity has no counterpart in other information theory texts. We have also taken delight in relating Fisher information, mutual information, and the Brunn-Minkowski and entropy power inequalities. To our surprise, many of the classical results on determinant inequalities are most easily proved using information theory.

Even though the field of information theory has grown considerably since Shannon's original paper, we have strived to emphasize its coherence. While it is clear that Shannon was motivated by problems in communication theory when he developed information theory, we treat information theory as a field of its own with applications to communication theory and statistics.

We were drawn to the field of information theory from backgrounds in communication theory, probability theory and statistics, because of the apparent impossibility of capturing the intangible concept of information.

Since most of the results in the book are given as theorems and proofs, we expect the elegance of the results to speak for themselves. In many cases we actually describe the properties of the solutions before introducing the problems. Again, the properties are interesting in themselves and provide a natural rhythm for the proofs that follow.

One innovation in the presentation is our use of long chains of inequalities, with no intervening text, followed immediately by the

explanations. By the time the reader comes to many of these proofs, we expect that he or she will be able to follow most of these steps without any explanation and will be able to pick out the needed explanations. These chains of inequalities serve as pop quizzes in which the reader can be reassured of having the knowledge needed to prove some important theorems. The natural flow of these proofs is so compelling that it prompted us to flout one of the cardinal rules of technical writing. And the absence of verbiage makes the logical necessity of the ideas evident and the key ideas perspicuous. We hope that by the end of the book the reader will share our appreciation of the elegance, simplicity and naturalness of information theory.

Throughout the book we use the method of weakly typical sequences, which has its origins in Shannon's original 1948 work but was formally developed in the early 1970s. The key idea here is the so-called asymptotic equipartition property, which can be roughly paraphrased as "Almost everything is almost equally probable."

Chapter 2, which is the true first chapter of the subject, includes the basic algebraic relationships of entropy, relative entropy and mutual information as well as a discussion of the second law of thermodynamics and sufficient statistics. The asymptotic equipartition property (AEP) is given central prominence in Chapter 3. This leads us to discuss the entropy rates of stochastic processes and data compression in Chapters 4 and 5. A gambling sojourn is taken in Chapter 6, where the duality of data compression and the growth rate of wealth is developed.

The fundamental idea of Kolmogorov complexity as an intellectual foundation for information theory is explored in Chapter 7. Here we replace the goal of finding a description that is good on the average with the goal of finding the universally shortest description. There is indeed a universal notion of the descriptive complexity of an object. Here also the wonderful number Ω is investigated. This number, which is the binary expansion of the probability that a Turing machine will halt, reveals many of the secrets of mathematics.

Channel capacity, which is the fundamental theorem in information theory, is established in Chapter 8. The necessary material on differential entropy is developed in Chapter 9, laying the groundwork for the extension of previous capacity theorems to continuous noise channels. The capacity of the fundamental Gaussian channel is investigated in Chapter 10.

The relationship between information theory and statistics, first studied by Kullback in the early 1950s, and relatively neglected since, is developed in Chapter 12. Rate distortion theory requires a little more background than its noiseless data compression counterpart, which accounts for its placement as late as Chapter 13 in the text.

The huge subject of network information theory, which is the study of the simultaneously achievable flows of information in the presence of

noise and interference, is developed in Chapter 14. Many new ideas come into play in network information theory. The primary new ingredients are interference and feedback. Chapter 15 considers the stock market, which is the generalization of the gambling processes considered in Chapter 6, and shows again the close correspondence of information theory and gambling.

Chapter 16, on inequalities in information theory, gives us a chance to recapitulate the interesting inequalities strewn throughout the book, put them in a new framework and then add some interesting new inequalities on the entropy rates of randomly drawn subsets. The beautiful relationship of the Brunn-Minkowski inequality for volumes of set sums, the entropy power inequality for the effective variance of the sum of independent random variables and the Fisher information inequalities are made explicit here.

We have made an attempt to keep the theory at a consistent level. The mathematical level is a reasonably high one, probably senior year or first-year graduate level, with a background of at least one good semester course in probability and a solid background in mathematics. We have, however, been able to avoid the use of measure theory. Measure theory comes up only briefly in the proof of the AEP for ergodic processes in Chapter 15. This fits in with our belief that the fundamentals of information theory are orthogonal to the techniques required to bring them to their full generalization.

Each chapter ends with a brief telegraphic summary of the key results. These summaries, in equation form, do not include the qualifying conditions. At the end of each we have included a variety of problems followed by brief historical notes describing the origins of the main results. The bibliography at the end of the book includes many of the key papers in the area and pointers to other books and survey papers on the subject.

The essential vitamins are contained in Chapters 2, 3, 4, 5, 8, 9, 10, 12, 13 and 14. This subset of chapters can be read without reference to the others and makes a good core of understanding. In our opinion, Chapter 7 on Kolmogorov complexity is also essential for a deep understanding of information theory. The rest, ranging from gambling to inequalities, is part of the terrain illuminated by this coherent and beautiful subject.

Every course has its first lecture, in which a sneak preview and overview of ideas is presented. Chapter 1 plays this role.

TOM COVER
JOY THOMAS

Acknowledgments

We wish to thank everyone who helped make this book what it is. In particular, Toby Berger, Masoud Salehi, Alon Orlitsky, Jim Mazo and Andrew Barron have made detailed comments on various drafts of the book which guided us in our final choice of content. We would like to thank Bob Gallager for an initial reading of the manuscript and his encouragement to publish it. We were pleased to use twelve of his problems in the text. Aaron Wyner donated his new proof with Ziv on the convergence of the Lempel-Ziv algorithm. We would also like to thank Norman Abramson, Ed van der Meulen, Jack Salz and Raymond Yeung for their suggestions.

Certain key visitors and research associates contributed as well, including Amir Dembo, Paul Algoet, Hirosuke Yamamoto, Ben Kawabata, Makoto Shimizu and Yoichiro Watanabe. We benefited from the advice of John Gill when he used this text in his class. Abbas El Gamal made invaluable contributions and helped begin this book years ago when we planned to write a research monograph on multiple user information theory. We would also like to thank the Ph.D. students in information theory as the book was being written: Laura Ekroot, Will Equitz, Don Kimber, Mitchell Trott, Andrew Nobel, Jim Roche, Erik Ordentlich, Elza Erkip and Vittorio Castelli. Also Mitchell Oslick, Chien-Wen Tseng and Michael Morrell were among the most active students in contributing questions and suggestions to the text. Marc Goldberg and Anil Kaul helped us produce some of the figures. Finally we would like to thank Kirsten Goodell and Kathy Adams for their support and help in some of the aspects of the preparation of the manuscript.

Joy Thomas would also like to thank Peter Franaszek, Steve Lavenberg, Fred Jelinek, David Nahamoo and Lalit Bahl for their encouragement and support during the final stages of production of this book.

TOM COVER
JOY THOMAS

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 Cramér-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 / 364 | |
| | 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 | 482 |
| 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 |

List of Figures

| | | |
|-----|---|-----|
| 1.1 | The relationship of information theory with other fields | 2 |
| 1.2 | Information theoretic extreme points of communication theory | 2 |
| 1.3 | Noiseless binary channel. | 7 |
| 1.4 | A noisy channel | 7 |
| 1.5 | Binary symmetric channel | 8 |
| 2.1 | $H(p)$ versus p | 15 |
| 2.2 | Relationship between entropy and mutual information | 20 |
| 2.3 | Examples of convex and concave functions | 24 |
| 3.1 | Typical sets and source coding | 53 |
| 3.2 | Source code using the typical set | 54 |
| 4.1 | Two-state Markov chain | 62 |
| 4.2 | Random walk on a graph | 66 |
| 5.1 | Classes of codes | 81 |
| 5.2 | Code tree for the Kraft inequality | 83 |
| 5.3 | Properties of optimal codes | 98 |
| 5.4 | Induction step for Huffman coding | 100 |
| 5.5 | Cumulative distribution function and Shannon-Fano-Elias coding | 101 |
| 5.6 | Tree of strings for arithmetic coding | 105 |
| 5.7 | The sgn function and a bound | 109 |
| 5.8 | Tree for generation of the distribution $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ | 111 |
| 5.9 | Tree to generate a $(\frac{2}{3}, \frac{1}{3})$ distribution | 114 |
| 7.1 | A Turing machine | 147 |

| | | |
|-------|---|-----|
| 7.2 | $H_0(p)$ versus p | 158 |
| 7.3 | Assignment of nodes | 173 |
| 7.4 | Kolmogorov sufficient statistic | 177 |
| 7.5 | Kolmogorov sufficient statistic for a Bernoulli sequence | 177 |
| 7.6 | Mona Lisa | 178 |
| 8.1 | A communication system | 184 |
| 8.2 | Noiseless binary channel. | 185 |
| 8.3 | Noisy channel with nonoverlapping outputs. | 185 |
| 8.4 | Noisy typewriter. | 186 |
| 8.5 | Binary symmetric channel. | 187 |
| 8.6 | Binary erasure channel | 188 |
| 8.7 | Channels after n uses | 192 |
| 8.8 | A communication channel | 192 |
| 8.9 | Jointly typical sequences | 197 |
| 8.10 | Lower bound on the probability of error | 207 |
| 8.11 | Discrete memoryless channel with feedback | 213 |
| 8.12 | Joint source and channel coding | 216 |
| 9.1 | Quantization of a continuous random variable | 228 |
| 9.2 | Distribution of \tilde{X} | 235 |
| 10.1 | The Gaussian channel | 240 |
| 10.2 | Sphere packing for the Gaussian channel | 243 |
| 10.3 | Parallel Gaussian channels | 251 |
| 10.4 | Water-filling for parallel channels | 253 |
| 10.5 | Water-filling in the spectral domain | 256 |
| 10.6 | Gaussian channel with feedback | 257 |
| 12.1 | Universal code and the probability simplex | 290 |
| 12.2 | Error exponent for the universal code | 291 |
| 12.3 | The probability simplex and Sanov's theorem | 293 |
| 12.4 | Pythagorean theorem for relative entropy | 297 |
| 12.5 | Triangle inequality for distance squared | 299 |
| 12.6 | The conditional limit theorem | 302 |
| 12.7 | Testing between two Gaussian distributions | 307 |
| 12.8 | The likelihood ratio test on the probability simplex | 308 |
| 12.9 | The probability simplex and Chernoff's bound | 313 |
| 12.10 | Relative entropy $D(P_\lambda \ P_1)$ and $D(P_\lambda \ P_2)$ as a function of λ | 314 |
| 12.11 | Distribution of yards gained in a run or a pass play | 317 |
| 12.12 | Probability simplex for a football game | 318 |
| 13.1 | One bit quantization of a Gaussian random variable | 337 |

| | | |
|-------|---|-----|
| 13.2 | Rate distortion encoder and decoder | 339 |
| 13.3 | Joint distribution for binary source | 343 |
| 13.4 | Rate distortion function for a binary source | 344 |
| 13.5 | Joint distribution for Gaussian source | 345 |
| 13.6 | Rate distortion function for a Gaussian source | 346 |
| 13.7 | Reverse water-filling for independent Gaussian random variables | 349 |
| 13.8 | Classes of source sequences in rate distortion theorem | 361 |
| 13.9 | Distance between convex sets | 365 |
| 13.10 | Joint distribution for upper bound on rate distortion function | 370 |
| 14.1 | A multiple access channel | 375 |
| 14.2 | A broadcast channel | 375 |
| 14.3 | A communication network | 376 |
| 14.4 | Network of water pipes | 376 |
| 14.5 | The Gaussian interference channel | 382 |
| 14.6 | The two-way channel | 383 |
| 14.7 | The multiple access channel | 388 |
| 14.8 | Capacity region for a multiple access channel | 389 |
| 14.9 | Independent binary symmetric channels | 390 |
| 14.10 | Capacity region for independent BSC's | 391 |
| 14.11 | Capacity region for binary multiplier channel | 391 |
| 14.12 | Equivalent single user channel for user 2 of a binary erasure multiple access channel | 392 |
| 14.13 | Capacity region for binary erasure multiple access channel | 392 |
| 14.14 | Achievable region of multiple access channel for a fixed input distribution | 395 |
| 14.15 | m -user multiple access channel | 403 |
| 14.16 | Gaussian multiple access channel | 403 |
| 14.17 | Gaussian multiple access channel capacity | 406 |
| 14.18 | Slepian-Wolf coding | 408 |
| 14.19 | Slepian-Wolf encoding: the jointly typical pairs are isolated by the product bins | 412 |
| 14.20 | Rate region for Slepian-Wolf encoding | 414 |
| 14.21 | Jointly typical fans | 416 |
| 14.22 | Multiple access channels | 417 |
| 14.23 | Correlated source encoding | 417 |
| 14.24 | Broadcast channel | 418 |
| 14.25 | Capacity region for two orthogonal broadcast channels | 419 |

| | | |
|-------|---|-----|
| 14.26 | Binary symmetric broadcast channel | 426 |
| 14.27 | Physically degraded binary symmetric broadcast channel | 426 |
| 14.28 | Capacity region of binary symmetric broadcast channel | 427 |
| 14.29 | Gaussian broadcast channel | 428 |
| 14.30 | The relay channel | 428 |
| 14.31 | Encoding with side information | 433 |
| 14.32 | Rate distortion with side information | 438 |
| 14.33 | Rate distortion for two correlated sources | 443 |
| 14.34 | A general multiterminal network | 444 |
| 14.35 | The relay channel | 448 |
| 14.36 | Transmission of correlated sources over a multiple access channel | 449 |
| 14.37 | Multiple access channel with cooperating senders | 452 |
| 14.38 | Capacity region of a broadcast channel | 456 |
| 14.39 | Broadcast channel—BSC and erasure channel | 456 |
| 15.1 | Sharpe-Markowitz theory: Set of achievable mean-variance pairs | 460 |
| 16.1 | The function $f(t) = -t \log t$ | 489 |

Elements of Information Theory

Index

Page numbers set in **boldface** indicate the primary references.

- Abramson, N. M., xi, 510
- Acceptance region, 305, 306, 309–311
- Achievable rate, **195**, 404, 406, 454
- Achievable rate distortion pair, 341
- Achievable rate region, 389, 408, 421
- Aczél, J., 511
- Adams, K., xi
- Adaptive source coding, 107
- Additive channel, 220, 221
- Additive white Gaussian noise (AWGN)
 channel, *see* Gaussian channel
- Adler, R. L., 124, 510
- AEP (asymptotic equipartition property), ix,
 x, 6, 11, **51**. *See also*
 Shannon–McMillan–Breiman theorem
- continuous random variables, 226, 227
- discrete random variables, **51**, 50–59, 65,
 133, 216–218
- joint, **195**, 201–204, **384**
- stationary ergodic processes, 474–480
- stock market, 471
- Ahlsvede, R., 10, 457, 458, 510
- Algoet, P., xi, 59, 481, 510
- Algorithm:
 arithmetic coding, 104–107, 124, 136–138
- Blahut–Arimoto, 191, 223, 366, 367, 373
- Durbin, 276
- Frank–Wolfe, 191
- generation of random variables, 110–117
- Huffman coding, 92–110
- Lempel–Ziv, 319–326
- Levinson, 275
- universal data compression, 107, 288–291,
 319–326
- Algorithmically random, 156, **157**, 166, 179,
 181–182
- Algorithmic complexity, 1, 3, 144, 146, **147**,
 162, 182
- Alphabet:
 continuous, 224, 239
- discrete, 13
- effective size, 46, 237
- input, **184**
- output, **184**
- Alphabetic code, 96
- Amari, S., 49, 510
- Approximation, Stirling's, 151, 181, 269, 282,
 284
- Approximations to English, 133–135
- Arimoto, S., 191, 223, 366, 367, 373, 510, 511.
 See also Blahut–Arimoto algorithm
- Arithmetic coding, 104–107, 124, 136–138
- Arithmetic mean geometric mean inequality,
 492
- ASCII, 147, 326
- Ash, R. B., 511
- Asymmetric distortion, 368
- Asymptotic equipartition property (AEP),
 see AEP
- Asymptotic optimality of log-optimal
 portfolio, 466
- Atmosphere, 270

- Atom, 114–116, 238
- Autocorrelation, 272, 276, 277
- Autoregressive process, 273
- Auxiliary random variable, 422, 426
- Average codeword length, 85
- Average distortion, 340, 356, 358, 361
- Average power, 239, 246
- Average probability of error, 194
- AWGN (additive white Gaussian noise), *see* Gaussian channel
- Axiomatic definition of entropy, 13, 14, 42, 43

- Bahl, L. R., xi, 523
- Band, 248, 349, 407
- Band-limited channel, 247–250, 262, 407
- Bandpass filter, 247
- Bandwidth, 249, 250, 262, 379, 406
- Barron, A., xi, 59, 276, 496, 511
- Baseball, 316
- Base of logarithm, 13
- Bayesian hypothesis testing, 314–316, 332
- BCH (Bose–Chaudhuri–Hocquenghem) codes, 212
- Beckner, W., 511
- Bell, R., 143, 481, 511
- Bell, T. C., 320, 335, 517
- Bellman, R., 511
- Bennett, C. H., 49, 511
- Benzel, R., 458, 511
- Berger, T., xi, 358, 371, 373, 457, 458, 511, 525
- Bergmans, P., 457, 512
- Berlekamp, E. R., 512, 523
- Bernoulli, J., 143
- Bernoulli random variable, 14, 43, **56**, 106, 154, 157, 159, 164, 166, 175, 177, 236, 291, 392, 454
 - entropy, 14
 - rate distortion function, 342, 367–369
- Berry's paradox, 163
- Betting, 126–133, 137–138, 166, 474
- Bias, **326**, 334, 335
- Biased, 305, 334
- Bierbaum, M., 454, 512
- Binary entropy function, **14**, 44, 150
 - graph of, 15
- Binary erasure channel, 187–189, 218
 - with feedback, 189, 214
 - multiple access, 391
- Binary multiplying channel, 457
- Binary random variable, *see* Bernoulli random variable
- Binary rate distortion function, 342
- Binary symmetric channel (BSC), **8**, **186**, 209, 212, 218, 220, 240, 343, 425–427, 456
- Binning, 410, 411, 442, 457
- Birkhoff's ergodic theorem, 474
- Bit, 13, 14
- Blachman, N., 497, 509, 512
- Blackwell, D., 512
- Blahut, R. E., 191, 223, 367, 373, 512
- Blahut–Arimoto algorithm, 191, 223, 366, 367, 373
- Block code:
 - channel coding, 193, 209, 211, 221
 - source coding, 53–55, 288
- Block length, 8, 104, 211, 212, 221, 222, 291, 356, 399, 445
- Boltzmann, L., 49. *See also* Maxwell–Boltzmann distribution
- Bookie, 128
- Borel–Cantelli lemma, 287, 467, 478
- Bose, R. C., 212, 512
- Bottleneck, 47
- Bounded convergence theorem, 329, 477, 496
- Bounded distortion, 342, 354
- Brain, 146
- Brascamp, H. J., 512
- Breiman, L., 59, 512. *See also* Shannon–McMillan–Breiman theorem
- Brillouin, L., 49, 512
- Broadcast channel, 10, 374, 377, 379, 382, 396, **420**, 418–428, 449, 451, 454–458
 - common information, 421
 - definitions, 420–422
 - degraded:
 - achievability, 422–424
 - capacity region, 422
 - converse, 455
 - physically degraded, **422**
 - stochastically degraded, **422**
 - examples, 418–420, 425–427
 - Gaussian, 379–380, 427–428
- Brunn–Minkowski inequality, viii, x, 482, 497, **498**, 500, 501, 509
- BSC (binary symmetric channel), 186, 208, 220
- Burg, J. P., 273, 278, 512
- Burg's theorem, viii, 274, 278
- Burst error correcting code, 212
- Buzo, A., 519

- Calculus, 78, 85, 86, 191, 267
- Capacity, ix, 2, 7–10, 184–223, 239–265, 377–458, 508
 - channel, *see* Channel, capacity
- Capacity region, 10, 374, 379, 380, 384, **389**, 390–458
 - broadcast channel, **421**, 422
 - multiple access channel, **389**, 396
- Capital assets pricing model, 460

- Carathéodory, 398
- Cardinality, 226, 397, 398, 402, 422, 426
- Cards, 36, 132, 133, 141
- Carleial, A. B., 458, 512
- Cascade of channels, 221, 377, 425
- Castelli, V., xi
- Cauchy distribution, 486
- Cauchy-Schwarz inequality, 327, 329
- Causal, 257, 258, 380
 - portfolio strategy, 465, 466
- Central limit theorem, 240, 291
- Central processing unit (CPU), 146
- Centroid, 338, 346
- Cesáro mean, 64, 470, 505
- Chain rule, 16, 21–24, 28, 32, 34, 39, 47, 65, 70, 204–206, 232, 275, 351, 400, 401, 414, 435, 441, 447, 469, 470, 480, 483, 485, 490, 491, 493
 - differential entropy, 232
 - entropy, 21
 - growth rate, 469
 - mutual information, 22
 - relative entropy, 23
- Chaitin, G. J., 3, 4, 182, 512, 513
- Channel, ix, 3, 7–10, 183, **184**, 185–223, 237, 239–265, 374–458, 508. *See also* Binary symmetric channel; Broadcast channel; Gaussian channel; Interference channel; Multiple access channel; Relay channel; Two-way channel
 - capacity:
 - computation, 191, 367
 - examples, 7, 8, 184–190
 - information capacity, 184
 - operational definition, 194
 - zero-error, 222, 223
 - cascade, 221, 377, 425
 - discrete memoryless:
 - capacity theorem, 198–206
 - converse, 206–212
 - definitions, 192
 - feedback, 212–214
 - symmetric, 189
- Channel code, **194**, 215–217
- Channel coding theorem, 198
- Channels with memory, 220, 253, 256, 449
- Channel transition matrix, 184, 189, 374
- Chebyshev's inequality, 57
- Chernoff, H., 312, 318, 513
- Chernoff bound, 309, 312–316, 318
- Chernoff information, 312, **314**, 315, 332
- Chessboard, 68, 75
- χ^2 (Chi-squared) distribution, 333, 486
- Choi, B. S., 278, 513, 514
- Chung, K. L., 59, 513
- Church's thesis, 146
- Cipher, 136
- Cleary, J. G., 124, 320, 335, 511, 524
- Closed system, 10
- Cloud of codewords, 422, 423
- Cocktail party, 379
- Code, 3, 6, 8, 10, 18, 53–55, 78–124, 136–137, 194–222, 242–258, 337–358, 374–458
 - alphabetic, 96
 - arithmetic, 104–107, 136–137
 - block, *see* Block code
 - channel, *see* Channel code
 - convolutional, 212
 - distributed source, *see* Distributed source code
 - error correcting, *see* Error correcting code
 - Hamming, *see* Hamming code
 - Huffman, *see* Huffman code
 - Morse, 78, 80
 - rate distortion, **341**
 - Reed-Solomon, 212
 - Shannon, *see* Shannon code
 - source, *see* Source code
- Codebook:
 - channel coding, **193**
 - rate distortion, **341**
- Codelength, 86–89, 94, 96, 107, 119
- Codepoints, 337
- Codeword, 8, 10, 54, 57, 78–124, 193–222, 239–256, 355–362, 378–456
- Coding, random, *see* Random coding
- Coin tosses, 13, 110
- Coin weighing, 45
- Common information, 421
- Communication channel, 1, 6, 7, 183, 186, 215, 219, 239, 488
- Communication system, 8, 49, 184, 193, 215
- Communication theory, vii, viii, 1, 4, 145
- Compact discs, 3, 212
- Compact set, 398
- Competitive optimality:
 - log-optimal portfolio, 471–474
 - Shannon code, 107–110
- Compression, *see* Data compression
- Computable, 147, 161, 163, 164, 170, 179
- Computable probability distribution, 161
- Computable statistical tests, 159
- Computation:
 - channel capacity, 191, 367
 - halting, 147
 - models of, 146
 - rate distortion function, 366–367
- Computers, 4–6, 144–181, 374
- Computer science, vii, 1, 3, 145, 162
- Concatenation, 80, 90

- Concavity, 14, **23**, 24–27, 29, 31, 40, 155, 191, 219, 237, 247, 267, 323, 369, 461–463, 479, 483, 488, 501, 505, 506. *See also*
 - Convexity
- Concavity of entropy, 14, 31
- Conditional differential entropy, **230**
- Conditional entropy, **16**
- Conditional limit theorem, 297–304, 316, 317, 332
- Conditionally typical set, 359, 370, 371
- Conditional mutual information, **22**, 44, 48, 396
- Conditional relative entropy, **23**
- Conditional type, 371
- Conditioning reduces entropy, 28, 483
- Consistent estimation, 3, 161, 165, 167, **327**
- Constrained sequences, 76, 77
- Continuous random variable, **224**, 226, 229, 235, 237, 273, 336, 337, 370. *See also*
 - Differential entropy; Quantization; Rate distortion theory
- AEP, 226
- Converse:
 - broadcast channel, 355
 - discrete memoryless channel, 206–212
 - with feedback, 212–214
 - Gaussian channel, 245–247
 - general multiterminal network, 445–447
 - multiple access channel, 399–402
 - rate distortion with side information, 440–442
 - rate distortion theorem, 349–351
 - Slepian–Wolf coding, 413–415
 - source coding with side information, 433–436
- Convex hull, 389, 393, 395, 396, 403, 448, 450
- Convexification, 454
- Convexity, **23**, 24–26, 29–31, 41, 49, 72, 309, 353, 362, 364, 396–398, 440–442, 454, 461, 462, 479, **482–484**. *See also*
 - Concavity
- capacity region:
 - broadcast channel, 454
 - multiple access channel, 396
- conditional rate distortion function, 439
- entropy and relative entropy, 30–32
- rate distortion function, 349
- Convex sets, 191, 267, 297, 299, 330, 362, 416, 454
 - distance between, 464
- Convolution, 498
- Convolutional code, 212
- Coppersmith, D., 124, 510
- Correlated random variables, 38, 238, 256, 264
 - encoding of, *see* Slepian–Wolf coding
- Correlation, 38, 46, 449
- Costa, M. H. M., 449, 513, 517
- Costello, D. J., 519
- Covariance matrix, 230, 254–256, 501–505
- Cover, T. M., x, 59, 124, 143, 182, 222, 265, 278, 432, 449, 450, 457, 458, 481, 509, 510, 511, 513–515, 523
- CPU (central processing unit), 146
- Cramér, H., 514
- Cramér–Rao bound, 325–329, 332, 335, 494
- Crosstalk, 250, 375
- Cryptography, 136
- Csiszár, I., 42, 49, 279, 288, 335, 358, 364–367, 371, 454, 458, 514, 518
- Cumulative distribution function, **101**, 102, 104, 106, **224**
- D*-adic, **87**
- Daróczy, Z., 511
- Data compression, vii, ix 1, 3–5, 9, 53, 60, 78, 117, 129, 136, 137, 215–217, 319, 331, 336, 374, 377, 407, 454, 459, 508
 - universal, 287, 319
- Davissón, L. D., 515
- de Bruijn’s identity, 494
- Decision theory, *see* Hypothesis testing
- Decoder, 104, 137, 138, 184, **192**, 203–220, 288, 291, 339, 354, 405–451, 488
- Decoding delay, 121
- Decoding function, **193**
- Decryption, 136
- Degradation, 430
- Degraded, *see* Broadcast channel, degraded;
 - Relay channel, degraded
- Dembo, A., xi, 498, 509, 514
- Demodulation, 3
- Dempster, A. P., 514
- Density, xii, **224**, 225–231, 267–271, 486–507
- Determinant, 230, 233, 237, 238, 255, 260
 - inequalities, 501–508
- Deterministic, 32, 137, 138, 193, 202, 375, 432, 457
- Deterministic function, 370, 454
 - entropy, 43
- Dice, 268, 269, 282, 295, 304, 305
- Differential entropy, ix, **224**, 225–238, 485–497
 - table of, 486–487
- Digital, 146, 215
- Digitized, 215
- Dimension, 45, 210

- Dirichlet region, 338
- Discrete channel, **184**
- Discrete entropy, *see* Entropy
- Discrete memoryless channel, *see* Channel, discrete memoryless
- Discrete random variable, 13
- Discrete time, 249, 378
- Discrimination, 49
- Distance:
 - Euclidean, 296–298, 364, 368, 379
 - Hamming, 339, 369
 - \mathcal{L}_1 , 299
 - variational, 300
- Distortion, ix, 9, 279, 336–372, 376, 377, 439–444, 452, 458, 508. *See also* Rate distortion theory
- Distortion function, **339**
- Distortion measure, 336, 337, **339**, 340–342, 349, 351, 352, 367–369, 373
 - bounded, 342, 354
 - Hamming, 339
 - Itakura–Saito, 340
 - squared error, 339
- Distortion rate function, **341**
- Distortion typical, **352**, 352–356, 361
- Distributed source code, **408**
- Distributed source coding, 374, 377, 407. *See also* Slepian–Wolf coding
- Divergence, 49
- DMC (discrete memoryless channel), 193, 208
- Dobrushin, R. L., 515
- Dog, 75
- Doubling rate, 9, 10, **126**, 126–131, 139, **460**, 462–474
- Doubly stochastic matrix, 35, 72
- Duality, x, 4, 5
 - data compression and data transmission, 184
 - gambling and data compression, 125, 128, 137
 - growth rate and entropy rate, 459, 470
 - multiple access channel and Slepian–Wolf coding, 416–418
 - rate distortion and channel capacity, 357
 - source coding and generation of random variables, 110
- Dueck, G., 457, 458, 515
- Durbin algorithm, 276
- Dutch, 419
- Dutch book, 129
- Dyadic, 103, **108**, 110, 113–116, 123
- Ebert, P. M., 265, 515
- Economics, 4
- Effectively computable, 146
- Efficient estimator, 327, 330
- Efficient frontier, 460
- Eggleston, H. G., 398, 515
- Eigenvalue, 77, 255, 258, 262, 273, 349, 367
- Einstein, A., vii
- Ekroot, L., xi
- El Gamal, A., xi, 432, 449, 457, 458, 513, 515
- Elias, P., 124, 515, 518. *See also* Shannon–Fano–Elias code
- Empirical, 49, 133, 151, 279
- Empirical distribution, 49, 106, 208, 266, 279, 296, 402, 443, 485
- Empirical entropy, 195
- Empirical frequency, 139, 155
- Encoder, 104, 137, 184, **192**
- Encoding, 79, 193
- Encoding function, **193**
- Encrypt, 136
- Energy, 239, 243, 249, 266, 270
- England, 34
- English, 80, 125, 133–136, 138, 139, 143, 151, 215, 291
 - entropy rate, 138, 139
 - models of, 133–136
- Entropy, vii–x, 1, 3–6, **5**, 9–13, **13**, 14. *See also* Conditional entropy; Differential entropy; Joint entropy; Relative entropy
 - and Fisher information, 494–496
 - and mutual information, 19, 20
 - and relative entropy, 27, 30
 - Renyi, **499**
- Entropy of English, 133, 135, 138, 139, 143
- Entropy power, 499
- Entropy power inequality, viii, x, 263, 482, 494, **496**, 496–501, 505, 509
- Entropy rate, **63**, 64–78, 88–89, 104, 131–139, 215–218
 - differential, 273
 - Gaussian process, 273–276
 - hidden Markov models, 69
 - Markov chain, 66
 - subsets, 490–493
- Epimenides liar paradox, 162
- Epitaph, 49
- Equitz, W., xi
- Erasure, 187–189, 214, 370, 391, 392, 449, 450, 452
- Ergodic, x, 10, 59, 65–67, 133, 215–217, 319–326, 332, 457, 471, 473, **474**, 475–478
- Erkip, E., xi

- Erlang distribution, 486
- Error correcting codes, 3, 211
- Error exponent, 4, 291, 305–316, 332
- Estimation, 1, 326, 506
 - spectrum, *see* Spectrum estimation
- Estimator, **326**, 326–329, 332, 334, 335, 488, 494, 506
 - efficient, **330**
- Euclidean distance, 296–298, 364, 368, 379
- Expectation, 13, 16, 25
- Exponential distribution, 270, 304, 486
- Extension of channel, 193
- Extension of code, **80**

- Face vase illusion, 182
- Factorial, 282, 284
 - function, 486
- Fair odds, 129, 131, 132, 139–141, 166, 167, 473
- Fair randomization, 472, 473
- Fan, Ky, 237, 501, 509
- Fano, R. M., 49, 87, 124, 223, 455, 515. *See also* Shannon–Fano–Elias code
- Fano code, 97, 124
- Fano's inequality, 38–40, **39**, 42, 48, 49, 204–206, 213, 223, 246, 400–402, 413–415, 435, 446, 447, 455
- F* distribution, 486
- Feedback:
 - discrete memoryless channels, 189, 193, 194, 212–214, 219, 223
 - Gaussian channels, 256–264
 - networks, 374, 383, 432, 448, 450, 457, 458
- Feinstein, A., 222, 515, 516
- Feller, W., 143, 516
- Fermat's last theorem, 165
- Finite alphabet, 59, 154, 319
- Finitely often, 479
- Finitely refutable, 164, 165
- First order in the exponent, **55**, 281, 285
- Fisher, R. A., 49, 516
- Fisher information, *x*, 228, 279, **327**, 328–332, 482, 494, 496, 497
- Fixed rate block code, 288
- Flag, 53
- Flow of information, 445, 446, 448
- Flow of time, 72
- Flow of water, 377
- Football, 317, 318
- Ford, L. R., 376, 377, 516
- Forney, G. D., 516
- Fourier transform, 248, 272
- Fractal, 152
- Franaszek, P. A., *xi*, 124, 516
- Frequency, 247, 248, 250, 256, 349, 406
 - empirical, 133–135, 282
- Frequency division multiplexing, 406
- Fulkerson, D. R., 376, 377, 516
- Functional, 4, 13, 127, 252, 266, 294, 347, 362

- Gaarder, T., 450, 457, 516
- Gadsby, 133
- Gallager, R. G., *xi*, 222, 232, 457, 516, 523
- Galois fields, 212
- Gambling, *viii–x*, 11, 12, 125–132, 136–138, 141, 143, 473
 - universal, 166, 167
- Game:
 - baseball, 316, 317
 - football, 317, 318
 - Hi-lo, 120, 121
 - mutual information, 263
 - red and black, 132
 - St. Petersburg, 142, 143
 - Shannon guessing, 138
 - stock market, 473
 - twenty questions, 6, 94, 95
- Game, theoretic optimality, 107, 108, 465
- Gamma distribution, 486
- Gas, 31, 266, 268, 270
- Gaussian channel, 239–265. *See also*
 - Broadcast channel; Interference channel; Multiple access channel; Relay channel
 additive white Gaussian noise (AWGN), 239–247, 378
 - achievability, 244–245
 - capacity, 242
 - converse, 245–247
 - definitions, 241
 - power constraint, 239
 band-limited, 247–250
 - capacity, 250
 colored noise, 253–256
 - feedback, 256–262
 parallel Gaussian channels, 250–253
- Gaussian distribution, *see* Normal distribution
- Gaussian source:
 - quantization, 337, 338
 - rate distortion function, 344–346
- Gauss–Markov process, 274–277
- Generalized normal distribution, 487
- General multiterminal network, 445
- Generation of random variables, 110–117
- Geodesic, 309
- Geometric distribution, 322
- Geometry:
 - Euclidean, 297, 357
 - relative entropy, 9, 297, 308
- Geophysical applications, 273

- Gilbert, E. W., 124, 516
 Gill, J. T., xi
 Goldbach's conjecture, 165
 Goldberg, M., xi
 Goldman, S., 516
 Gödel's incompleteness theorem, 162–164
 Goodell, K., xi
 Gopinath, B., 514
 Gotham, 151, 409
 Gradient search, 191
 Grammar, 136
 Graph:
 binary entropy function, 15
 cumulative distribution function, 101
 Kolmogorov structure function, 177
 random walk on, 66–69
 state transition, 62, 76
 Gravestone, 49
 Gravitation, 169
 Gray, R. M., 182, 458, 514, 516, 519
 Grenander, U., 516
 Grouping rule, 43
 Growth rate optimal, 459
 Guisasu, S., 516

 Hadamard inequality, 233, 502
 Halting problem, 162–163
 Hamming, R. V., 209, 516
 Hamming code, 209–212
 Hamming distance, 209–212, 339
 Hamming distortion, 339, 342, 368, 369
 Han, T. S., 449, 457, 458, 509, 510, 517
 Hartley, R. V., 49, 517
 Hash functions, 410
 Hassner, M., 124, 510
 HDTV, 419
 Hekstra, A. P., 457, 524
 Hidden Markov models, 69–71
 High probability sets, 55, 56
 Histogram, 139
 Historical notes, 49, 59, 77, 124, 143, 182,
 222, 238, 265, 278, 335, 372, 457, 481, 509
 Hocquenghem, P. A., 212, 517
 Holsinger, J. L., 265, 517
 Hopcroft, J. E., 517
 Horibe, Y., 517
 Horse race, 5, 125–132, 140, 141, 473
 Huffman, D. A., 92, 124, 517
 Huffman code, 78, 87, 92–110, 114, 119,
 121–124, 171, 288, 291
 Hypothesis testing, 1, 4, 10, 287, 304–315
 Bayesian, 312–315
 optimal, *see* Neyman–Pearson lemma

 iff (if and only if), 86

 i.i.d. (independent and identically
 distributed), 6
 i.i.d. source, 106, 288, 342, 373, 474
 Images, 106
 distortion measure, 339
 entropy of, 136
 Kolmogorov complexity, 152, 178, 180, 181
 Incompressible sequences, 110, **157**,
 156–158, 165, 179
 Independence bound on entropy, 28
 Indicator function, 49, 165, 176, 193, 216
 Induction, 25, 26, 77, 97, 100, 497
 Inequalities, 482–509
 arithmetic mean geometric mean, 492
 Brunn–Minkowski, viii, x, 482, 497, **498**,
 500, 501, 509
 Cauchy–Schwarz, 327, 329
 Chebyshev's, 57
 determinant, 501–508
 entropy power, viii, x, 263, 482, 494, **496**,
 496–501, 505, 509
 Fano's, 38–40, **39**, 42, 48, 49, 204–206, 213,
 223, 246, 400–402, 413–415, 435, 446, 447,
 455, 516
 Hadamard, 233, 502
 information, **26**, 267, 484, 508
 Jensen's, 24, **25**, 26, 27, 29, 41, 47, 155,
 232, 247, 323, 351, 441, 464, 468, 482
 Kraft, 78, **82**, 83–92, 110–124, 153, 154, 163,
 171, 519
 log sum, **29**, 30, 31, 41, 483
 Markov's, 47, **57**, 318, 466, 471, 478
 Minkowski, 505
 subset, 490–493, 509
 triangle, 18, 299
 Young's, **498**, 499
 Ziv's, 323
 Inference, 1, 4, 6, 10, 145, 163
 Infinitely often (i.o.), 467
 Information, *see* Fisher information; Mutual
 information; Self information
 Information capacity, 7, **184**, 185–190, 204,
 206, 218
 Gaussian channel, **241**, 251, 253
 Information channel capacity, *see* Information
 capacity
 Information inequality, 27, 267, 484, 508
 Information rate distortion function, **341**,
 342, 346, 349, 362
 Innovations, 258
 Input alphabet, **184**
 Input distribution, 187, 188
 Instantaneous code, 78, **81**, 82, 85, 90–92,
 96, 97, 107, 119–123. *See also* Prefix code
 Integrability, 229

- Interference, ix, 10, 76, 250, 374, 388, 390, 406, 444, 458
- Interference channel, 376, 382, 383, 458
- Gaussian, 382–383
- Intersymbol interference, 76
- Intrinsic complexity, 144, 145
- Investment, 4, 9, 11
- horse race, 125–132
- stock market, 459–474
- Investor, 465, 466, 468, 471–473
- Irreducible Markov chain, 61, 62, 66, 216
- ISDN, 215
- Itakura-Saito distance, 340
- Jacobs, I. M., 524
- Jaynes, E. T., 49, 273, 278, 517
- Jefferson, the Virginian, 140
- Jelinek, F., xi, 124, 517
- Jensen's inequality, 24, **25**, 26, 27, 29, 41, 47, 155, 232, 247, 323, 351, 441, 464, 468, 482
- Johnson, R. W., 523
- Joint AEP, **195**, 196–202, 217, 218, 245, 297, 352, 361, **384–388**
- Joint distribution, 15
- Joint entropy, **15**, 46
- Jointly typical, **195**, 196–202, 217, 218, 297, 334, 378, 384–387, 417, 418, 432, 437, 442, 443
- distortion typical, 352–356
- Gaussian, 244, 245
- strongly, 358–361, 370–372
- Joint source channel coding theorem, 215–218, **216**
- Joint type, 177, 371
- Justesen, J., 212, 517
- Kailath, T., 517
- Karush, J., 124, 518
- Kaul, A., xi
- Kawabata, T., xi
- Kelly, J., 143, 481, 518
- Kemperman, J. H. B., 335, 518
- Kendall, M., 518
- Keyboard, 160, 162
- Khinchin, A. Ya., 518
- Kieffer, J. C., 59, 518
- Kimber, D., xi
- King, R., 143, 514
- Knuth, D. E., 124, 518
- Kobayashi, K., 458, 517
- Kolmogorov, A. N., 3, 144, 147, 179, 181, 182, 238, 274, 373, 518
- Kolmogorov complexity, viii, ix, 1, 3, 4, 6, 10, 11, **147**, 144–182, 203, 276, 508
- of integers, **155**
- and universal probability, 169–175
- Kolmogorov minimal sufficient statistic, **176**
- Kolmogorov structure function, **175**
- Kolmogorov sufficient statistic, 175–179, 182
- Körner, J., 42, 279, 288, 335, 358, 371, 454, 457, 458, 510, 514, 515, 518
- Kotel'nikov, V. A., 518
- Kraft, L. G., 124, 518
- Kraft inequality, 78, **82**, 83–92, 110–124, 153, 154, 163, 173
- Kuhn-Tucker conditions, 141, 191, 252, 255, 348, 349, 364, **462–466**, 468, 470–472
- Kullback, S., ix, 49, 335, 518, 519
- Kullback Leibler distance, **18**, 49, 231, 484
- Lagrange multipliers, 85, 127, 252, 277, 294, 308, 347, 362, 366, 367
- Landau, H. J., 249, 519
- Landauer, R., 49, 511
- Langdon, G. G., 107, 124, 519
- Language, entropy of, 133–136, 138–140
- Laplace, 168, 169
- Laplace distribution, 237, 486
- Large deviations, 4, 9, 11, 279, 287, 292–318
- Lavenberg, S., xi
- Law of large numbers:
- for incompressible sequences, 157, 158, 179, 181
- method of types, 286–288
- strong, 288, 310, 359, 436, 442, 461, 474
- weak, 50, 51, 57, 126, 178, 180, 195, 198, 226, 245, 292, 352, 384, 385
- Lehmann, E. L., 49, 519
- Leibler, R. A., 49, 519
- Lempel, A., 319, 335, 519, 525
- Lempel-Ziv algorithm, 320
- Lempel-Ziv coding, xi, 107, 291, 319–326, 332, 335
- Leningrad, 142
- Letter, 7, 80, 122, 133–135, 138, 139
- Leung, C. S. K., 450, 457, 458, 514, 520
- Levin, L. A., 182, 519
- Levinson algorithm, 276
- Levy's martingale convergence theorem, 477
- Lexicographic order, 53, 83, **104–105**, 137, 145, 152, 360
- Liao, H., 10, 457, 519
- Liar paradox, 162, 163
- Lieb, E. J., 512
- Likelihood, 18, 161, 182, 295, 306
- Likelihood ratio test, 161, 306–308, 312, 316
- Lin, S., 519

- Linde, Y., 519
 Linear algebra, 210
 Linear code, 210, 211
 Linear predictive coding, 273
 List decoding, 382, 430–432
 Lloyd, S. P., 338, 519
 Logarithm, base of, 13
 Logistic distribution, 486
 Log likelihood, 18, 58, 307
 Log-normal distribution, 487
 Log optimal, **127**, 130, 137, 140, 143, 367, 461–473, 478–481
 Log sum inequality, **29**, 30, 31, 41, 483
 Longo, G., 511, 514
 Lovasz, L., 222, 223, 519
 Lucky, R. W., 136, 519

 Macroscopic, 49
 Macrostate, 266, 268, 269
 Magnetic recording, 76, 80, 124
 Malone, D., 140
 Mandelbrot set, 152
 Marcus, B., 124, 519
 Marginal distribution, 17, 304
 Markov chain, **32**, 33–38, 41, 47, 49, **61**, 62, 66–77, 119, 178, 204, 215, 218, 435–437, 441, 484
 Markov field, 32
 Markov lemma, 436, 443
 Markov process, 36, 61, 120, 274–277. *See also* Gauss–Markov process
 Markov's inequality, 47, 57, 318, 466, 471, 478
 Markowitz, H., 460
 Marshall, A., 519
 Martian, 118
 Martingale, 477
 Marton, K., 457, 458, 518, 520
 Matrix:
 channel transition, 7, **184**, 190, 388
 covariance, 238, 254–262, 330
 determinant inequalities, 501–508
 doubly stochastic, 35, 72
 parity check, 210
 permutation, 72
 probability transition, **35**, 61, 62, 66, 72, 77, 121
 state transition, 76
 Toeplitz, 255, 273, 504, 505
 Maximal probability of error, 193
 Maximum entropy, viii, 10, 27, 35, 48, 75, 78, 258, 262
 conditional limit theorem, 269
 discrete random variable, 27
 distribution, 266–272, **267**
 process, 75, 274–278
 property of normal, 234
 Maximum likelihood, 199, 220
 Maximum *posteriori* decision rule, 314
 Maxwell–Boltzmann distribution, 266, 487
 Maxwell's demon, 182
 Mazo, J., xi
 McDonald, R. A., 373, 520
 McEliece, R. J., 514, 520
 McMillan, B., 59, 124, 520. *See also* Shannon–McMillan–Breiman theorem
 McMillan's inequality, 90–92, 117, 124
 MDL (minimum description length), 182
 Measure theory, x
 Median, 238
 Medical testing, 305
 Memory, channels with, 221, 253
 Memoryless, 57, 75, **184**. *See also* Channel, discrete memoryless
 Mercury, 169
 Merges, 122
 Merton, R. C., 520
 Message, 6, 10, 184
 Method of types, 279–286, 490
 Microprocessor, 148
 Microstates, 34, 35, 266, 268
 Midpoint, 102
 Minimal sufficient statistic, **38**, 49, 176
 Minimum description length (MDL), 182
 Minimum distance, 210–212, 358, 378
 between convex sets, 364
 relative entropy, 297
 Minimum variance, 330
 Minimum weight, 210
 Minkowski, H., 520
 Minkowski inequality, 505
 Mirsky, L., 520
 Mixture of probability distributions, 30
 Models of computation, 146
 Modem, 250
 Modulation, 3, 241
 Modulo 2 arithmetic, 210, 342, 452, 458
 Molecules, 266, 270
 Moments, 234, 267, 271, 345, 403, 460
 Money, 126–142, 166, 167, 471. *See also* Wealth
 Monkeys, 160–162, 181
 Moore, E. F., 124, 516
 Morrell, M., xi
 Morse code, 78, 80
 Moy, S. C., 59, 520
 Multiparameter Fisher information, 330
 Multiple access channel, 10, 374, 377, 379, 382, 387–407, 416–418, 446–454, 457

- Multiple access channel (*Continued*)
 achievability, 393
 capacity region, 389
 converse, 399
 with correlated sources, 448
 definitions, 388
 duality with Slepian-Wolf coding, 416–418
 examples, 390–392
 with feedback, 450
 Gaussian, 378–379, 403–407
- Multiplexing, 250
 frequency division, 407
 time division, 406
- Multiuser information theory, *see* Network information theory
- Multivariate distributions, **229**, 268
- Multivariate normal, 230
- Music, 3, 215
- Mutual information, vii, viii, 4–6, 9–12, **18**,
 19–33, 40–49, 130, 131, 183–222, **231**,
 232–265, 341–457, 484–508
 chain rule, 22
 conditional, 22
- Myers, D. L., 524
- Nahamoo, D., xi
- Nats, 13
- Nearest neighbor, 3, 337
- Neal, R. M., 124, 524
- Neighborhood, 292
- Network, 3, 215, 247, 374, 376–378, 384, 445,
 447, 448, 450, 458
- Network information theory, ix, 3, 10, 374–458
- Newtonian physics, 169
- Neyman, J., 520
- Neyman-Pearson lemma, 305, 306, 332
- Nobel, A., xi
- Noise, 1, 10, 183, 215, 220, 238–264, 357, 374,
 376, 378–384, 388, 391, 396, 404–407, 444,
 450, 508
 additive noise channel, 220
- Noiseless channel, 7, 416
- Nonlinear optimization, 191
- Non-negativity:
 discrete entropy, 483
 mutual information, 484
 relative entropy, 484
- Nonsense, 162, 181
- Non-singular code, **80**, 80–82, 90
- Norm:
 \mathcal{L}_1 , 299, 488
 \mathcal{L}_∞ , 498
- Normal distribution, **38**, 225, 230, 238–265,
 487. *See also* Gaussian channels, Gaussian
 source
 entropy of, 225, 230, 487
 entropy power inequality, 497
 generalized, 487
 maximum entropy property, 234
 multivariate, 230, 270, 274, 349, 368,
 501–506
- Nyquist, H., 247, 248, 520
- Nyquist-Shannon sampling theorem, 247, 248
- Occam's Razor, 1, 4, 6, 145, 161, 168, 169
- Odds, 11, 58, **125**, 126–130, 132, 136, 141,
 142, 467, 473
- Olkin, I., 519
- Ω , **164**, 165–167, 179, 181
- Omura, J. K., 520, 524
- Oppenheim, A., 520
- Optimal decoding, 199, 379
- Optimal doubling rate, **127**
- Optimal portfolio, **459**, 474
- Oracle, 165
- Ordentlich, E., xi
- Orey, S., 59, 520
- Orlitsky, A., xi
- Ornstein, D. S., 520
- Orthogonal, 419
- Orthonormal, 249
- Oscillate, 64
- Oslick, M., xi
- Output alphabet, **184**
- Ozarow, L. H., 450, 457, 458, 520
- Pagels, H., 182, 520
- Paradox, 142–143, 162, 163
- Parallel channels, 253, 264
- Parallel Gaussian channels, 251–253
- Parallel Gaussian source, 347–349
- Pareto distribution, 487
- Parity, 209, 211
- Parity check code, 209
- Parity check matrix, 210
- Parsing, 81, **319**, 320, 322, 323, 325, 332, 335
- Pasco, R., 124, 521
- Patterson, G. W., 82, 122, 123, 522
- Pearson, E. S., 520
- Perez, A., 59, 521
- Perihelion, 169
- Periodic, 248
- Periodogram, 272
- Permutation, 36, 189, 190, 235, 236, 369, 370
 matrix, 72
- Perpendicular bisector, 308
- Perturbation, 497
- Philosophy of science, 4
- Phrase, 135, 319–325, 332
- Physics, viii, 1, 4, 33, 49, 145, 161, 266

- Pinkston, J. T., 369, 521
 Pinsker, M. S., 238, 265, 457, 521
 Pitfalls, 162
 Pixels, 152
 Pollak, H. O., 249, 519, 523
 Polya urn model, 73
 Polynomial number of types, 280, 281, 286, 288, 297, 302
 Pombra, S., 265, 514
 Portfolio, 127, 143, 367, 459, 460, **461**, 461–473, 478–481
 competitively optimal, 471–474
 log optimal, **127**, 130, 137, 139, 143, 367, 461–473, 478–481
 rebalanced, 461
 Positive definite matrices, 255, 260, 501–507
 Posner, E., 514
 Power, **239**, 239–264, 272, 353, 357, 378–382, 384, 403–407, 453, 458
 Power spectral density, 249, 262, 272
 Precession, 169
 Prefix, 53, 81, 82, 84, 92, 98, 99, 102–104, 121–124, 146, 154, 164, 167, 172, 319, 320
 Prefix code, **81**, 82, 84, 85, 92, 122, 124
 Principal minor, 502
 Prior:
 Bayesian, 312–316
 universal, 168
 Probability density, **224**, 225, 230, 276
 Probability mass function, 5, **13**
 Probability of error, **193**, 305
 average, **194**, 199–202, 389
 maximal, 193, 202
 Probability transition matrix, 6, **35**, 61, 62, 66, 72, 77, 121
 Projection, 249, 333
 Prolate spheroidal functions, 249
 Proportional gambling, **127**, 129, 130, 137, 139, 143, 475
 Punctuation, 133, 138
 Pythagorean theorem, 297, 301, 332

 Quantization, 241, 294, 336–338, 346, 368
 Quantized random variable, 228
 Quantizer, 337

 Rabiner, L. R., 521
 Race, *see* Horse race
 Radio, 240, 247, 407, 418
 Radium, 238
 Random coding, 3, **198**, 203, 222, 357, 416, 432
 Randomization, 472, 473
 Random number generator, 134

 Rank, 210
 Rao, C. R., 330, 334, 521
 Rate, 1, 3, 4, 7, 9–11
 achievable, *see* Achievable rate
 channel coding, 194
 entropy, *see* Entropy rate
 source coding, 55
 Rate distortion code, **340**, 351–370
 with side information, 441
 Rate distortion function, **341**, 342–369
 binary source, 342
 convexity of, 349
 Gaussian source, 344
 parallel Gaussian source, 346
 Rate distortion function with side information, **439**, 440–442
 Rate distortion region, **341**
 Rate distortion theory, ix, 10, 287, 336–373
 achievability, 351–358
 converse, 349
 theorem, 342
 Rate region, **393**, 408, 411, 427, 448, 454, 456
 Rathie, P. N., 485, 524
 Ray-Chaudhuri, D. K., 212, 512
 Rayleigh distribution, 487
 Rebalanced portfolio, 461
 Receiver, 3, 7, 10, **183**
 Recurrence, 73, 75
 Recursion, 73, 77, 97, 106, 147, 150
 Redistribution, 34
 Redundancy, 122, 136, 184, 209, 216
 Reed, I. S., 212
 Reed–Solomon codes, 212
 Reinvest, 126, 128, 142, 460
 Relative entropy, **18**, **231**
 Relative entropy “distance”, 11, 18, 34, 287, 364
 Relay channel, 376–378, 380–382, 430–434, 448, 451, 458
 achievability, 430
 capacity, 430
 converse, 430
 definitions, 428
 degraded, 428, 430, 432
 with feedback, 432
 Gaussian, 380
 physically degraded, 430
 reversely degraded, 432
 Renyi entropy, **499**
 Renyi entropy power, **499**
 Reverse water-filling, 349, 367
 Reza, F. M., 521
 Rice, S. O., 521
 Rissanen, J. J., 124, 182, 276, 519, 521

- Roche, J., xi
 Rook, 68, 76
 Run length, 47

 Saddlepoint, 263
 St. Petersburg paradox, 142–143
 Salehi, M., xi, 449, 457, 513
 Salz, J., xi
 Sampling theorem, 248
 Samuelson, P. A., 481, 520–522
 Sandwich argument, 59, 471, 474, 475, 477, 481
 Sanov, I. N., 522
 Sanov's theorem, 286, **292**, 291–296, 308, 313, 318
 Sardinas, A. A., 122, 522
 Satellite, 240, 374, 379, 388, 423
 Sato, H., 458, 522
 Scaling of differential entropy, 238
 Schafer, R. W., 521
 Schalkwijk, J. P. M., 457, 517, 522, 525
 Schnorr, C. P., 182, 522
 Schultheiss, P. M., 373, 520
 Schwarz, G., 522
 Score function, **327**, 328
 Second law of thermodynamics, vi, viii, 3, 10, 12, 23, 33–36, 42, 47, 49, 182. *See also* Statistical mechanics
 Self-information, 12, 20
 Self-punctuating, 80, 149
 Self-reference, 162
 Sender, 10, 189
 Set sum, **498**
 sgn function, 108–110
 Shakespeare, 162
 Shannon, C. E., viii, 1, 3, 4, 10, 43, 49, 59, 77, 80, 87, 95, 124, 134, 138–140, 143, 198, 222, 223, 248, 249, 265, 369, 372, 383, 457, 458, 499, 509, 522, 523. *See also* Shannon code; Shannon–Fano–Elias code; Shannon–McMillan–Breiman theorem
 Shannon code, 89, 95, 96, 107, 108, 113, 117, 118, 121, 124, 144, 151, 168, 459
 Shannon–Fano code, 124. *See also* Shannon code
 Shannon–Fano–Elias code, 101–104, 124
 Shannon–McMillan–Breiman theorem, 59, 474–478
 Shannon's first theorem (source coding theorem), 88
 Shannon's second theorem (channel coding theorem), 198
 Shannon's third theorem (rate distortion theorem), 342
 Sharpe, W. F., 460, 481, 523
 Shimizu, M., xi
 Shore, J. E., 523
 Shuffles, 36, 48
 Side information, vii, 11
 and doubling rate, 125, 130, 131, 467–470, 479, 480
 and source coding, 377, 433–442, 451, 452, 458
 Sigma algebra, 474
 Signal, 3
 Signal to noise ratio, 250, 380, 405
 Simplex, 308, 309, 312, 463
 Sinc function, 248
 Singular code, 81
 Sink, 376
 Slepian, D., 249, 407, 457, 523
 Slepian–Wolf coding, 10, 377, 407–418, 433, 444, 449, 451, 454, 457
 achievability, 410–413
 achievable rate region, 408, 409
 converse, 413–415
 distributed source code, 408
 duality with multiple access channels, 416–418
 examples, 409–410, 454
 interpretation, 415
 many sources, 415
 Slice code, 96
 SNR (signal to noise ratio), 250, 380, 405
 Software, 149
 Solomon, G., 212
 Solomonoff, R. J., 3, 4, 182, 524
 Source, 2, 9, 10
 binary, 342
 Gaussian, 344
 Source channel separation, 215–218, 377, 448, 449
 Source code, **79**
 universal, 288
 Source coding theorem, 88
 Spanish, 419
 Spectral representation theorem, 255, 349
 Spectrum, 248, 255, 272, 274, 275–278, 349
 Spectrum estimation, 272–278
 Speech, 1, 135, 136, 215, 273, 340
 Sphere, 243, 357, 498
 Sphere packing, 9, 243, 357, 358
 Squared error, 274, 329, 334, 494, 503–505
 distortion, 337, 338, 340, 345, 346, 367–369
 Stam, A., 497, 509, 523

- State diagram, 76
- Stationary:
 - ergodic process, 59, 319–326, 474–478
 - Gaussian channels, 255
 - Markov chain, 35, 36, 41, 47, 61–74, 485
 - process, 4, 47, 59, 60–78, **60**, 89
 - stock market, 469–471
- Statistic, **36**, 36–38, 176, 333
 - Kolmogorov sufficient, 176
 - minimal sufficient, 38
 - sufficient, 37
- Statistical mechanics, 1, 3, 6, 10, 33, 49, 278
- Statistics, viii–x, 1, 4, 11, 12, 18, 36–38, 134, 176, 279, 304
 - of English, 136
- Stein's lemma, 279, 286, 305, **309**, 310, 312, 332, 333
- Stirling's approximation, 151, 181, 269, 282, 284
- Stochastic approximation to English, 133–136
- Stochastic process, **60**, 61–78, 88, 89, 272–277, 469, 471, 476
 - ergodic, *see* Ergodic
 - Gaussian, 255, 272–278
 - stationary, *see* Stationary
- Stock, 9, 459, 460, 465, 466, 471, 473, 480
- Stock market, x, 4, 9, 11, 125, 131, 367, **459**, 459–481, 508
- Strategy, 45
 - football, 317, 318
 - gambling, 126, 129–131, 166
 - portfolio, 465–471
- Strong converse, 207, 223
- Strongly typical, 358, 359, 370–372, 436, 442
- Strongly typical set, **287**, **358**
- Stuart, A., 518
- Student-*t* distribution, 487
- Subfair odds, 130, 141
- Submatrix, 502
- Subset inequalities, 490–493, 509
- Subspace, 210
- Subtree, 105, 106
- Sufficient statistic, **37**, 38, 49, 175
- Suffix, 120, 122, 123
- Summary, 40, 56, 71, 117, 140, 178, 218, 236, 262, 275, 276, 331, 367, 450, 479, 508
- Superfair odds, 129, 141
- Superposition coding, 422, 424, 430–432, 457
- Support set, **224**
- Surface, 407
- Surface area, 228, 331
- Symbol, 7–9, 63
- Symmetric channel, **190**, 220
- Synchronization, 76
- Systematic code, 211
- Szasz's lemma, 502, 503
- Szego, G., 516
- Table of differential entropies, 486, 487
- Tallin, 182
- Tanabe, M., 458, 522
- Tang, D. L., 523
- Tangent, 460
- Telephone, 247, 250
 - channel capacity, 250
- Temperature, 266, 268
- Ternary, 93, 118–120, 181, 280, 391
- Test channel, 345
- Thermodynamics, 1, 3, 13, 34, 35, 49. *See also* Second law of thermodynamics
- Thomas, J. A., x, xi, 509, 513
- Thomasian, A. J., 512
- Time, flow of, 72
- Time division multiplexing, 379, 406, 407
- Time invariant Markov chain, **61**
- Time reversible, 69
- Timesharing, 391, 395, 396, 406, 407, 420, 435, 447, 453
- Toeplitz matrix, 255, 273, 504, 505
- Tornay, S. C., 524
- Trace, 254, 255
- Transmitter, 183
- Treasury bonds, 460
- Tree, 72–74, 82, 84, 93, 96, 97, 98, 104–106, 111–116, 124, 171–174
- Triangle inequality, 18, 299
- Triangular distribution, 487
- Trigram model, 136
- Trott, M., xi
- Tseng, C. W., xi
- Turing, A., 146, 147
- Turing machine, ix, 146, 147, 160
- Tusnády, G., 364–367, 373, 514
- Tuttle, D. L., 519
- TV, 374, 418
- Two-way channel, 375, 383, 384, 450, 456, 457
- Type, 155, 169, 178, **279**, 280–318, 490
- Type class, 154, **280**, 281–286, 490
- Typewriter, 63, 161, 185, 191, 208, 220
- Typical, 10, 11, **50**, 51–54, 65, **226**
- Typical sequence, 386, 407, 437
- Typical set, 50–56, **51**, 65, 159, **226**, 227, 228
 - jointly typical set, *see* Jointly typical
 - strongly typical set, 370, 371
- Typical set decoding, 200

- Ullman, J. D., 517
 Unbiased, **326**, 326–329, 332, 334, 494
 Uncertainty, 5, 6, 10, 12, 14, 18–20, 22, 28, 35, 47, 48, 72, 382, 450
 Unfair odds, 141
 Uniform distribution:
 continuous, 229, 487
 discrete, 5, 27, 30, 35, 41, 72, 73
 Uniform fair odds, 129
 Uniquely decodable, 79, **80**, 81, 82, 85, 90–92, 101, 108, 117, 118, 120–124
 Universal, ix, 3, 9, 144–148, 155, 160–162, 170, 174, 178–182
 Universal gambling, 166, 167
 Universal probability, **160**
 Universal source coding, 107, 124, 287–291, **288**, 319, 326
 UNIX, 326
 Urn, 45, 73

 V'yugin, V. V., 182, 524
 Van Campenhout, J. M., 523
 Van der Meulen, E. C., 457, 458, 515, 523, 524
 Variance, **36**, 57, 234–264, 270, 274, 316, 317, 327, 330. *See also* Covariance matrix
 Variational distance, **300**
 Vase, 181
 Venn diagram, 21, 45, 46
 Verdugo Lazo, A. C. G., 485, 524
 Vitamins, x
 Viterbi, A. J., 523
 Vocabulary, 419
 Volume, 58, 225, **226**, 226–229, 243, 331, 357, 498–501
 Voronoi region, 338

 Wald, A., 523
 Wallmeier, H. M., 454, 512

 Watanabe, Y., xi
 Waterfilling, 264, 349
 Waveform, 247, 340, 383
 Weakly typical, *see* Typical
 Wealth, x, 4, 9, 10, 34, 125–142, 166, 167, 459–480
 Wealth relative, **126**
 Weather, 151, 409
 Weaver, W. W., 523
 Weibull distribution, 487
 Welch, T. A., 320, 523
 White noise, 247, 249, 256, 405. *See also* Gaussian channel
 Wiener, N., 523
 Wilcox, H. J., 524
 Willems, F. M. J., 238, 450, 457, 524
 William of Occam, 4, 161
 Witten, I. H., 124, 320, 335, 511, 524
 Wolf, J. K., 407, 450, 457, 516, 523
 Wolfowitz, J., 223, 335, 524
 Woodward, P. M., 524
 World Series, 44
 Wozencraft, J. M., 524
 Wyner, A. D., xi, 320, 335, 443, 458, 516, 524, 525

 Yamamoto, H., xi
 Yao, A. C., 124, 518
 Yeung, R. W., xi, 458, 511
 Young's inequality, **498**, 499

 Zero-error capacity, 222, 223
 Zhang, Z., 457, 525
 Ziv, J., xi, 107, 124, 319, 335, 443, 458, 519, 524, 525. *See also* Lempel–Ziv coding
 Ziv's inequality, 323
 Zurek, W. H., 182, 525
 Zvonkin, A. K., 519