

GIAN-CARLO ROTA, *Editor*
ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS
Volume 12

Section: Real Variables
James K. Brooks, *Section Editor*

**Mathematical Theory of
Entropy**

Nathaniel F. G. Martin

James W. England

GIAN-CARLO ROTA, *Editor*
ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS
Volume 12

Section: Real Variables
James K. Brooks, *Section Editor*

Mathematical Theory of Entropy

Nathaniel F. G. Martin

University of Virginia
Charlottesville, Virginia

James W. England

Swarthmore College
Swarthmore, Pennsylvania

Foreword by
James K. Brooks
University of Florida, Gainesville



1981

Addison-Wesley Publishing Company

Advanced Book Program
Reading, Massachusetts

London • Amsterdam • Don Mills, Ontario • Sydney • Tokyo

Library of Congress Cataloging in Publication Data

Martin, Nathaniel F. G.
Mathematical theory of entropy.

(Encyclopedia of mathematics and its applications; v. 12)

Bibliography: p.

Includes index.

1. Entropy (Information theory) 2. Ergodic theory.
3. Statistical mechanics. 4. Topological dynamics.

I. England, James W. II. Title. III. Series.

Q360.M316 519.2 81-834

ISBN 0-201-13511-6 AACR2

ABCEFGHIJ-HA-8987654321

American Mathematical Society (MOS) Subject Classification Scheme (1980): 28-02, 28A6;
28D20, 94A17, 28D05, 54H20, 58F11, 60G10, 82A05, 94A15.

Copyright © 1981 by Addison-Wesley Publishing Company, Inc.
Published simultaneously in Canada.

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, mechanical, photocopying, recording
or otherwise, without the prior written permission of the publisher, Addison-Wesley Publishing
Company, Inc., Advanced Book Program, Reading, Massachusetts 01867, U.S.A.

Manufactured in the United States of America

GIAN-CARLO ROTA, *Editor*
ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

Volume		Section
1	LUIS A. SANTALÓ Integral Geometry and Geometric Probability , 1976	Probability
2	GEORGE E. ANDREWS The Theory of Partitions , 1976	Number Theory
3	ROBERT J. McELIECE The Theory of Information and Coding A Mathematical Framework for Communication, 1977	Probability
4	WILLARD MILLER, Jr. Symmetry and Separation of Variables , 1977	Special Functions
5	DAVID RUELLE Thermodynamic Formalism The Mathematical Structures of Classical Equilibrium Statistical Mechanics, 1978	Statistical Mechanics
6	HENRYK MINC Permanents , 1978	Linear Algebra
7	FRED S. ROBERTS Measurement Theory with Applications to Decisionmaking, Utility, and the Social Sciences, 1979	Mathematics and the Social Sciences
8	L. C. BIEDENHARN and J. D. LOUCK Angular Momentum in Quantum Physics: Theory and Application, 1981	Mathematics of Physics
9	L. C. BIEDENHARN and J. D. LOUCK The Racah-Wigner Algebra in Quantum Theory , 1981	Mathematics of Physics
10	JOHN D. DOLLARD and CHARLES N. FRIEDMAN Product Integration with Application to Differential Equations, 1979	Analysis

GIAN-CARLO ROTA, *Editor*
ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

Volume		Section
11	WILLIAM B. JONES and W. J. THRON Continued Fractions: Analytic Theory and Applications, 1980	Analysis
12	NATHANIEL F. G. MARTIN and JAMES W. ENGLAND Mathematical Theory of Entropy, 1981	Real Variables
13	GEORGE A. BAKER, JR. and PETER GRAVES-MORRIS Padé Approximants, Vol. I: Basic Theory, 1981	Mathematics of Physics
14	GEORGE A. BAKER, JR. and PETER GRAVES-MORRIS Padé Approximants, Vol. II: Extensions and Applications, 1981	Mathematics of Physics

Other volumes in preparation

ENCYCLOPEDIA OF MATHEMATICS and Its Applications

GIAN-CARLO ROTA, Editor
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Massachusetts

Editorial Board

- | | |
|--|--|
| Janos D. Aczel, <i>Waterloo</i> | Shizuo Kakutani, <i>Yale</i> |
| Richard Askey, <i>Madison</i> | Samuel Karlin, <i>Stanford</i> |
| Michael F. Atiyah, <i>Oxford</i> | J. F. C. Kingman, <i>Oxford</i> |
| Donald Babbitt, <i>U.C.L.A.</i> | Donald E. Knuth, <i>Stanford</i> |
| Edwin F. Beckenbach, <i>U.C.L.A.</i> | Joshua Lederberg, <i>Rockefeller</i> |
| Lipman Bers, <i>Columbia</i> | André Lichnerowicz, <i>College de France</i> |
| Garrett Birkhoff, <i>Harvard</i> | M. J. Lighthill, <i>Cambridge</i> |
| Salomon Bochner, <i>Rice</i> | Chia-Chiao Lin, <i>M.I.T.</i> |
| Raoul Bott, <i>Harvard</i> | Jacques-Louis Lions, <i>Paris</i> |
| James K. Brooks, <i>Gainesville</i> | G. G. Lorentz, <i>Austin</i> |
| Felix E. Browder, <i>Chicago</i> | Roger Lyndon, <i>Ann Arbor</i> |
| A. P. Calderón, <i>Buenos Aires</i> | Marvin Marcus, <i>Santa Barbara</i> |
| Peter A. Carruthers, <i>Los Alamos</i> | N. Metropolis, <i>Los Alamos</i> |
| S. Chandrasekhar, <i>Chicago</i> | Jan Mycielski, <i>Boulder</i> |
| S. S. Chern, <i>Berkeley</i> | Steven A. Orszag, <i>M.I.T.</i> |
| Hermann Chernoff, <i>M.I.T.</i> | Alexander Ostrowski, <i>Basle</i> |
| P. M. Cohn, <i>Bedford College, London</i> | Roger Penrose, <i>Oxford</i> |
| H. S. MacDonald Coxeter, <i>Toronto</i> | Carlo Pucci, <i>Florence</i> |
| Nelson Dunford, <i>Sarasota, Florida</i> | C. R. Rao, <i>Indian Statistical Institute</i> |
| F. J. Dyson, <i>Inst. for Advanced Study</i> | Fred S. Roberts, <i>Rutgers</i> |
| Harold M. Edwards, <i>Courant</i> | Abdus Salam, <i>Trieste</i> |
| Harvey Friedman, <i>Ohio State</i> | M. P. Schützenberger, <i>Paris</i> |
| Giovanni Gallavotti, <i>Rome</i> | Jacob T. Schwartz, <i>Courant</i> |
| Andrew M. Gleason, <i>Harvard</i> | Irving Segal, <i>M.I.T.</i> |
| James Glimm, <i>Rockefeller</i> | Olga Taussky, <i>Caltech</i> |
| A. González Domínguez, <i>Buenos Aires</i> | René Thom, <i>Bures-sur-Yvette</i> |
| M. Gordon, <i>Essex</i> | John Todd, <i>Caltech</i> |
| Peter Henrici, <i>ETH, Zurich</i> | John W. Tukey, <i>Princeton</i> |
| Nathan Jacobson, <i>Yale</i> | Stanislaw Ulam, <i>Colorado</i> |
| Mark Kac, <i>Rockefeller</i> | Veeravalli S. Varadarajan, <i>U.C.L.A.</i> |
| Antoni Zygmund, <i>Chicago</i> | |

Editor's Statement

A large body of mathematics consists of facts that can be presented and described much like any other natural phenomenon. These facts, at times explicitly brought out as theorems, at other times concealed within a proof, make up most of the applications of mathematics, and are the most likely to survive changes of style and of interest.

This *ENCYCLOPEDIA* will attempt to present the factual body of all mathematics. Clarity of exposition, accessibility to the non-specialist, and a thorough bibliography are required of each author. Volumes will appear in no particular order, but will be organized into sections, each one comprising a recognizable branch of present-day mathematics. Numbers of volumes and sections will be reconsidered as times and needs change.

It is hoped that this enterprise will make mathematics more widely used where it is needed, and more accessible in fields in which it can be applied but where it has not yet penetrated because of insufficient information.

GIAN-CARLO ROTA

Foreword

Entropy is a subject which has played a central role in a number of areas such as statistical mechanics and information theory. The connections between the various applications of entropy have become clearer in recent years by the introduction of probability theory into its foundations. It is now possible to see a number of what were previously isolated results in various disciplines as part of a more general mathematical theory of entropy.

This volume presents a self-contained exposition of the mathematical theory of entropy. Those parts of probability theory which are necessary for an understanding of the central topics concerning entropy have been included. In addition, carefully chosen examples are given in order that the reader may omit proofs of some of the theorems and yet by studying these examples and discussion obtain insight into the theorems.

The last four chapters give a description of those parts of information theory, ergodic theory, statistical mechanics, and topological dynamics which are most affected by entropy. These chapters may be read independently of each other. The examples show how ideas originating in one area have influenced other areas. Chapter III contains a brief description of how entropy as a measure of information flow has affected information theory and complements the first part of *The Theory of Information and Coding* by R. J. McEliece (volume 3 of this *ENCYCLOPEDIA*). Recent applications of entropy to statistical mechanics and topological dynamics are given in chapters V and VI. These two chapters provide a good introduction to *Thermodynamic Formalism* by D. Ruelle (volume 5 of this *ENCYCLOPEDIA*). The chapter on ergodic theory describes the development of Kolmogorov's adoption of Shannon entropy to the study of automorphisms on a finite measure space. It contains the culmination of this work in the proof of the Isomorphism Theorem of Kolmogorov and Ornstein. The mathematical treatment presented here of the major properties of entropy and the various applications to other fields make this volume a valuable addition to the *ENCYCLOPEDIA*.

JAMES K. BROOKS
General Editor, Section on Real Variables

Preface

Thirty years ago, Claude Shannon published a paper with the title “A mathematical theory of communication”. In this paper, he defined a quantity, which he called entropy, that measures the uncertainty associated with random phenomena. The effects of this paper on communications in both theory and practice are still being felt, and his entropy function has been applied very successfully to several areas of mathematics. In particular, an extension of it to dynamic situations by A. N. Kolmogorov and Ja. G. Sinai led to a complete solution of a long-unsolved problem in ergodic theory, to a new invariant for differentiable dynamic systems, and to more precision in certain concepts in classical statistical mechanics.

Our intent in this book is to give a rather complete and self-contained development of the entropy function and its extension that is understandable to a reader with a knowledge of abstract measure theory as it is taught in most first-year graduate courses and to indicate how it has been applied to the subjects of information theory, ergodic theory, and topological dynamics. We have made no attempt to give a comprehensive treatment of these subjects; rather we have restricted ourselves to just those parts of the subject which have been influenced by Shannon’s entropy and the Kolmogorov-Sinai extension of it. Thus, our purpose is twofold: first, to give a self-contained treatment of all the major properties of entropy and its extension, with rather detailed proofs, and second, to give an exposition of its uses in those areas of mathematics where it has been applied with some success. Our most extensive treatment is given to ergodic theory, since this is where the most spectacular results have been obtained.

The word entropy was first used in 1864 by Rudolph Clausius, in his book *Abhandlungen über die Wärmetheorie*, to describe a quantity accompanying a change from thermal to mechanical energy, and it has continued to have this meaning in thermodynamics. The connection between entropy as a measure of uncertainty and thermodynamic entropy was unclear for a number of years. With the introduction of measures, called Gibbs states, on infinite systems, this connection has been made clear. In the last chapter, we discuss this connection in the context of classical lattice systems.

In this connection we cannot resist repeating a remark made by Claude Shannon to Myron Tribus that Tribus reports in his and Edward McIrvine’s article “Energy and information” (*Scientific American*, 1971). Tribus was speaking to Shannon about his measure of uncertainty and

Shannon said, "My greatest concern was what to call it. I thought of calling it 'information,' but the word was overly used, so I decided to call it 'uncertainty.' When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, You should call it entropy, for two reasons. In the first place, your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage." We hope our reader will also have the advantage after reading this book.

The preparation of our manuscript would have been much more difficult without the generous support of the Mathematics Departments at the University of Virginia and Swarthmore College, and the careful and accurate typing of Beverley Watson, whose care and patience in typing the bulk of the manuscript and whose facility for accurately translating the first author's tiny, sometimes illegible, scrawl are most gratefully acknowledged. Our thanks also go to Janis Babbitt, Barbara Smith, and Jo Fields, who typed portions of the first chapter, and to Marie Brown, who typed the revisions. Finally, our thanks go to Alan Saleski for his careful reading of the first three chapters.

NATHANIEL F. G. MARTIN
JAMES W. ENGLAND

Special Symbols

Symbol	Description	Section
(Ω, \mathcal{F}, P)	Probability space	1.1
$(\Omega_\xi, \mathcal{F}_\xi, P_\xi)$	Factor space of ξ	1.2
(S, \mathcal{S}, u_f)	Discrete probability space with distribution f	1.2
$(I, \mathcal{L}, \lambda)$	Unit interval with Lebesgue measure	1.2
$\Sigma(S)$	Set of doubly infinite sequences of elements from S	1.2
$\Sigma'(S)$	Set of (one-sided) infinite sequences of elements from S	4.8
\mathcal{Z} or $\mathcal{Z}(\mathcal{U})$	Collection of all measurable partitions	1.3; 4.4
$\mathcal{Z}_k(\Omega)$	Collection of all measurable partitions with no more than k atoms	4.4
$(\Omega, \mathcal{F}, P, T)$	Dynamical system	1.7
(T, ξ)	Stationary stochastic process determined by ξ	1.7
$(B; p_1, \dots, p_k)$	Bernoulli shift with distribution (p_1, \dots, p_k)	4.3
Tail (B, ξ_0)	Tail of the process (B, ξ_0)	4.3
Ω_Λ	Configuration space of a lattice system in Λ	6.3
$[\Sigma(S), \mu]$	Information source	3.2
$[\Sigma(S), P(\omega, \cdot), \Sigma(B)]$	Channel	3.4
$\hat{\xi}$	The σ -field of ξ -sets	1.3
$\xi, \eta, \zeta, \alpha, \beta$	Measurable partitions	1.2
ν	Trivial partition	1.2
ε	Point partition	1.2
$\pi(T)$ or π	Pinsker partition of T	2.9
\mathcal{Q}, \mathcal{B}	Open covers of a topological space	5.2
$\xi \leq \eta$	ξ is refined by η	1.3
$\xi \leq_c \eta$	ξ is c -refined by η	4.4
$\mathcal{Q} < \mathcal{B}$	Open cover \mathcal{B} refines \mathcal{Q}	5.2

Symbol	Description	Section
$\xi \overset{c}{\odot} \eta$	ξ is c -independent of η	4.3
$\xi \vee \eta$	Supremum or common refinement of ξ and η	1.3
$\bigvee_{\alpha} \xi_{\alpha}$	Supremum or common refinement of the family $\{\xi_{\alpha}\}$	1.3
$\mathcal{Q} \vee \mathcal{B}$	Common refinement of open cover	5.2
$\xi \wedge \eta$	Infimum of partitions	1.3
$\bigwedge_{\alpha} \xi_{\alpha}$	Infimum of the family of partitions $\{\xi_{\alpha}\}$	1.3
ξ^n	Common refinement of $\{\mathbf{T}^j \xi: 0 \leq j \leq n-1\}$	4.3
ξ^+	Common refinement of $\{\mathbf{T}^j \xi: 0 \leq j < \infty\}$	4.3
ξ^{-n}	Common refinement of $\{\mathbf{T}^{-j} \xi: 1 \leq j \leq n\}$	4.3
${}^1\xi^{-n}$	Common refinement of $\{\mathbf{T}^{-j} \xi: 0 \leq j \leq n-1\}$	4.5
ξ^{-}	Common refinement of $\{\mathbf{T}^{-j} \xi: 1 \leq j < \infty\}$	4.3
ξ^{∞}	Common refinement of $\{\mathbf{T}^j \xi: -\infty < j < \infty\}$	4.3
$ d(\xi) - d(\eta) $	Distribution distance between ξ and η	4.4
$ \xi - \eta $	Partition distance between ξ and η	4.4
$R(\xi, \eta)$	Rohlin distance between ξ and η	4.4
\bar{d}	\bar{d} -metric	4.5
Ham	Hamming metric	4.5
\mathbf{N}_{ξ}	Projection onto the factor space of ξ	1.3
$\mathbf{N}_{\xi, \xi}$	Projection of factor space of ξ onto factor space of ξ	1.3
$\mathbf{M}_{\xi^{-n}}(l)$	ξ n -name of l	4.5
p_{Λ}	Restriction of a configuration to Λ	6.4
$p_{\Lambda_1 \Lambda_2}$	Restriction of a configuration Ω_{Λ_2} to Λ_1	6.4
$E(x)$	Expected value of the random variable x	1.4

Symbol	Description	Section
$P(\cdot A)$	Conditional probability given the event A	1.5
$P^\xi(\cdot c)$ or $P^\xi(\omega, \cdot)$	Canonical family of measures for ξ	1.5
$E^\xi(x c)$ or $E^\xi(x)$	Conditional expectation of random variable x given ξ	1.6
$d(\xi)$	Discrete probability vector associated with an ordered partition	4.4
$\bar{I}(\xi)$	Information function of ξ	2.2
$H(\xi)$	Entropy of ξ	2.2
$I(\xi/\eta)$	Conditional information of ξ given η	2.4; 2.6
$H(\xi/\eta)$	Conditional entropy of ξ given η	2.4; 2.6
$I(\xi; \eta)$	Mutual information between ξ and η	2.5
$h(\mathbf{T}, \xi)$	Entropy of \mathbf{T} given ξ or rate of information generation	2.7
$h(\mathbf{T})$ or $h_\mu(\mathbf{T})$	Entropy of \mathbf{T}	2.8
$H(\mathcal{Q})$	Entropy of open cover \mathcal{Q}	5.2
$h(\mathbf{T}, \mathcal{Q})$	Topological entropy of \mathbf{T} given \mathcal{Q}	5.2
$h_d(\mathbf{T}, K)$	Bowen topological entropy of \mathbf{T} given a compact set K	5.4
$h_d(\mathbf{T})$	Bowen topological entropy of \mathbf{T}	5.4
$S(P)$	Entropy of the state P	6.3
$S(\mu)$	Mean entropy of a translation invariance state μ	6.5
$P(\mathbf{T}, \cdot)$	Pressure of a continuous map \mathbf{T}	5.4
$P(\phi)$	Pressure of a translation invariant interaction ϕ	6.5
$\mu(\phi)$	Energy of the interaction ϕ for the state μ	6.5
U_Λ	Energy function	6.4
$W_{\Lambda_1 \Lambda_2}$	Interaction between Λ_1 and Λ_2	6.4
$\mathcal{Z}_\Lambda(\phi)$	Partition function	6.4
$C(P)$	Capacity of a channel	3.4
$R(\mu, P)$	Rate of transmission of a channel	3.4

CONTENTS

Editor's Statement	xiii
Section Editor's Foreword	xv
Preface	xvii
Special Symbols	xix
Chapter 1 Topics from Probability Theory	1
1.1 Probability Spaces	1
1.2 Measurable Partitions and Lebesgue Spaces	3
1.3 The Lattice of Measurable Partitions	10
1.4 Random Variables	13
1.5 Conditional Probability and Independence	15
1.6 Conditional Expectation of Random Variables	27
1.7 Stochastic Processes and Dynamical Systems	33
1.8 The Ergodic Theorem and the Martingale Convergence Theorem	38
Chapter 2 Entropy and Information	51
2.1 Information and Uncertainty of Events	51
2.2 The Information Function of an Experiment and Entropy	53
2.3 An Example	55
2.4 Conditional Information and Conditional Entropy	58
2.5 Properties of Entropy and Conditional Entropy	60
2.6 Entropy of Arbitrary Measurable Partitions and Limit Theorems	66
2.7 Rate of Information Generation	78
2.8 Entropy of Dynamical Systems	84
2.9 Factor Automorphisms and Factor Systems	90
2.10 Shannon's Theorem and the Equipartition Property	95
2.11 Entropy as a Function of Distributions	101
2.12 Examples	104
2.12.1 Direct Products	105
2.12.2 Skew Products	105
2.12.3 Powers of Endomorphisms	105

2.12.4	Flows	106
2.12.5	Induced Automorphisms	106
2.12.6	Periodic Automorphisms	107
2.12.7	Rotations of the Circle	107
2.12.8	Ergodic Automorphisms of Compact Abelian Groups	107
2.12.9	Bernoulli Shifts	109
2.12.10	Markov Shifts	109
2.12.11	S -Automorphisms	110
2.12.12	Unilateral Shifts	111
2.12.13	Continued Fraction Transformations	111
2.12.14	f -Transformations	112
2.13	Sequence Entropy and r -Entropy	113
Chapter 3	Information Theory	117
3.1	A Model of an Information System	117
3.2	The Source	119
3.3	Coding	121
3.4	The Channel	122
3.5	The Noisy-Channel Coding Theorem	128
3.6	Source Coding	130
Chapter 4	Ergodic Theory	147
4.1	Introduction	147
4.2	Unitary Operator of a System and Bernoulli Shifts	148
4.3	K -Systems and K -Automorphisms	150
4.4	Spaces of Ordered Partitions, Weak Independence, and Weak Dependence	165
4.5	Coding and Ornstein's Fundamental Lemma	173
4.6	The Isomorphism Theorem for Bernoulli Systems	195
4.7	Characterization of Bernoulli Systems	199
4.8	Relative Isomorphism	202
4.9	Special Flows and Equivalence Theory	206
Chapter 5	Topological Dynamics	213
5.1	Introduction	213
5.2	Definition and Basic Properties of Topological Entropy	215
5.3	Connection between Topological and Measure Theoretic Entropy	219
5.4	An Alternative Definition of Topological Entropy	222

Chapter 6	Statistical Mechanics	229
6.1	Introduction	229
6.2	Classical Continuous Systems	230
6.3	Classical Lattice Systems	234
6.4	Gibbs States for Lattice Systems	236
6.5	Equilibrium States and the Concepts of Entropy and Pressure	240
Bibliography		245
Index		253

CHAPTER 1

Topics from Probability Theory

In this preliminary chapter we shall give an exposition of certain topics in probability theory which are necessary to understand and interpret the definition and properties of entropy. We have tried to write the chapter in such a way that a reader with a knowledge of measure theory as given in Ash [15], Halmos [55], or any other basic measure theory text can follow the arguments and understand the examples. We introduce just those parts of probability theory which are necessary for the subsequent chapters and attempt to make them meaningful by use of very simple examples. We also restrict the discussion to “nice” probability spaces, so that conditional expectation and conditional probability are more intuitive and hopefully easier to understand. These “nice” spaces also make it possible to use partitions as models for random experiments, even those experiments which are limits of sequences of experiments.

1.1 Probability Spaces

Entropy is a quantitative measurement of uncertainty associated with random phenomena. In order to define this quantity precisely, it is necessary to have a mathematical model for random phenomena which is general enough to include many different physical situations and which has enough structure to allow us to use mathematical reasoning to answer questions about the phenomena.

Such a model is given by a mathematical structure called a probability space, which is nothing more than a measure space in which the measure of the universe set is 1. Thus, a probability space is a triple (Ω, \mathcal{F}, P) where Ω is a set, \mathcal{F} is a collection of subsets of Ω , and P is a nonnegative real valued function defined on \mathcal{F} such that

ENCYCLOPEDIA OF MATHEMATICS and Its Applications, Gian-Carlo Rota (ed.).
Vol. 12: Nathaniel F. Martin and James W. England, Mathematical Theory of Entropy.
ISBN 0-201-13511-6

Copyright © 1981 by Addison-Wesley Publishing Company, Inc., Advanced Book Program.
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, mechanical, photocopying, recording, or otherwise, without the prior permission of the publisher.