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

ABCDEFGHIJ-HA-8987654321

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



Manufactured in the United States of America

GIAN-CARLO ROTA, *Editor*

**ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS**

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

# 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 |
|---|---|---|
| $(\Omega, \mathfrak{F}, P)$ | Probability space | 1.1 |
| $(\Omega_\xi, \mathfrak{F}_\xi, P_\xi)$ | Factor space of $\xi$ | 1.2 |
| $(S, \mathbb{S}, u_f)$ | Discrete probability space with distribution $f$ | 1.2 |
| $(I, \mathfrak{L}, \lambda)$ | 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 |
| $\mathfrak{Z}$ or $\mathfrak{Z}(\mathfrak{U})$ | Collection of all measurable partitions | 1.3; 4.4 |
| $\mathfrak{Z}_k(\Omega)$ | Collection of all measurable partitions with no more than $k$ atoms | 4.4 |
| $(\Omega, \mathfrak{F}, P, \mathbf{T})$ | Dynamical system | 1.7 |
| $(\mathbf{T}, \xi)$ | Stationary stochastic process determined by $\xi$ | 1.7 |
| $(\mathbf{B}; p_1, \ldots, p_k)$ | Bernoulli shift with distribution $(p_1, \ldots, p_k)$ | 4.3 |
| Tail $(\mathbf{B}, \xi_0)$ | Tail of the process $(\mathbf{B}, \xi_0)$ | 4.3 |
| $\Omega_\Lambda$ | Configuration space of a lattice system in $\Lambda$ | 6.3 |
| $[\Sigma(S), \mu]$ | Information source | 3.2 |
| $[\Sigma(S), P(\omega,\ ), \Sigma(B)]$ | Channel | 3.4 |
| $\hat{\xi}$ | The $\sigma$-field of $\xi$-sets | 1.3 |
| $\xi, \eta, \zeta, \alpha, \beta$ | Measurable partitions | 1.2 |
| $\nu$ | Trivial partition | 1.2 |
| $\varepsilon$ | Point partition | 1.2 |
| $\pi(\mathbf{T})$ or $\pi$ | Pinsker partition of $\mathbf{T}$ | 2.9 |
| $\mathfrak{A}, \mathfrak{B}$ | Open covers of a topological space | 5.2 |
| $\xi \leqslant \eta$ | $\xi$ is refined by $\eta$ | 1.3 |
| $\xi \overset{c}{\leqslant} \eta$ | $\xi$ is $c$-refined by $\eta$ | 4.4 |
| $\mathfrak{A} < \mathfrak{B}$ | Open cover $\mathfrak{B}$ refines $\mathfrak{A}$ | 5.2 |

| Symbol | Description | Section |
|---|---|---|
| $\xi \overset{c}{\copyright} \eta$ | $\xi$ is $c$-independent of $\eta$ | 4.3 |
| $\xi \vee \eta$ | Supremum or common refinement of $\xi$ and $\eta$ | 1.3 |
| $\bigvee_\alpha \xi_\alpha$ | Supremum or common refinement of the family $\{\xi_\alpha\}$ | 1.3 |
| $\mathcal{A} \vee \mathcal{B}$ | 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 |
| $\xi^n$ | Common refinement of $\{\mathbf{T}^j\xi : 0 \leqslant j \leqslant n-1\}$ | 4.3 |
| $\xi^+$ | Common refinement of $\{\mathbf{T}^j\xi : 0 \leqslant j < \infty\}$ | 4.3 |
| $\xi^{-n}$ | Common refinement of $\{\mathbf{T}^{-j}\xi : 1 \leqslant j \leqslant n\}$ | 4.3 |
| ${}^1\xi^{-n}$ | Common refinement of $\{\mathbf{T}^{-j}\xi : 0 \leqslant j \leqslant n-1\}$ | 4.5 |
| $\xi^-$ | Common refinement of $\{\mathbf{T}^{-j}\xi : 1 \leqslant j < \infty\}$ | 4.3 |
| $\xi^\infty$ | Common refinement of $\{\mathbf{T}^j\xi : -\infty < j < \infty\}$ | 4.3 |
| $\|d(\xi) - d(\eta)\|$ | Distribution distance between $\xi$ and $\eta$ | 4.4 |
| $\|\xi - \eta\|$ | Partition distance between $\xi$ and $\eta$ | 4.4 |
| $R(\xi, \eta)$ | Rohlin distance between $\xi$ and $\eta$ | 4.4 |
| $\bar{d}$ | $\bar{d}$-metric | 4.5 |
| Ham | Hamming metric | 4.5 |
| $\mathbf{N}_\xi$ | Projection onto the factor space of $\xi$ | 1.3 |
| $\mathbf{N}_{\zeta, \xi}$ | Projection of factor space of $\zeta$ onto factor space of $\xi$ | 1.3 |
| $\mathbf{M}_{\xi^{-n}}(l)$ | $\xi$ $n$-name of $l$ | 4.5 |
| $p_\Lambda$ | Restriction of a configuration to $\Lambda$ | 6.4 |
| $p_{\Lambda_1 \Lambda_2}$ | Restriction of a configuration $\Omega_{\Lambda_2}$ to $\Lambda_1$ | 6.4 |
| $E(x)$ | Expected value of the random variable $x$ | 1.4 |

| Symbol | Description | Section |
|---|---|---|
| $P(\ \vert A)$ | Conditional probability given the event $A$ | 1.5 |
| $P^{\xi}(\ \vert c)$ or $P^{\xi}(\omega,\ )$ | Canonical family of measures for $\xi$ | 1.5 |
| $E^{\xi}(x\vert c)$ or $E^{\xi}(x)$ | Conditional expectation of random variable $x$ given $\xi$ | 1.6 |
| $d(\xi)$ | Discrete probability vector associated with an ordered partition | 4.4 |
| $\bar{I}(\xi)$ | Information function of $\xi$ | 2.2 |
| $H(\xi)$ | Entropy of $\xi$ | 2.2 |
| $I(\xi/\eta)$ | Conditional information of $\xi$ given $\eta$ | 2.4; 2.6 |
| $H(\xi/\eta)$ | Conditional entropy of $\xi$ given $\eta$ | 2.4; 2.6 |
| $I(\xi;\eta)$ | Mutual information between $\xi$ and $\eta$ | 2.5 |
| $h(\mathbf{T},\xi)$ | Entropy of $\mathbf{T}$ given $\xi$ 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 invariation state $\mu$ | 6.5 |
| $P(\mathbf{T},\ )$ | Pressure of a continuous map $\mathbf{T}$ | 5.4 |
| $P(\phi)$ | Pressure of a translation invariant interaction $\phi$ | 6.5 |
| $\mu(\phi)$ | Energy of the interaction $\phi$ for the state $\mu$ | 6.5 |
| $U_{\Lambda}$ | Energy function | 6.4 |
| $W_{\Lambda_1\Lambda_2}$ | Interaction between $\Lambda_1$ and $\Lambda_2$ | 6.4 |
| $\mathfrak{Z}_{\Lambda}(\phi)$ | Partition function | 6.4 |
| $C(P)$ | Capacity of a channel | 3.4 |
| $R(\mu,P)$ | Rate of transmission of a channel | 3.4 |

# CONTENTS

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 $(\Omega, \mathfrak{F}, P)$ where $\Omega$ is a set, $\mathfrak{F}$ is a collection of subsets of $\Omega$, and $P$ is a nonnegative real valued function defined on $\mathfrak{F}$ 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