国际知名大学原版教材

—— 信息技术学科与电气工程学科系列 21

Elements of Information Theory

信息论基础

Thomas M. Cover
Joy A. Thomas 著

WILEY
清华大学出版社

Elements of Information Theory

THOMAS M. COVER

Stanford University
Stanford, California

JOY A. THOMAS

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

Tsinghua University Press
Beijing

Thomas M. Cover, Joy A. Thomas
Elements of Information Theory
EISBN: 0-471-06259-6

北京市版权局著作权合同登记号　图字：01-2003-5603

本书封面贴有 John Wiley & Sons, Inc.防伪标签，无标签者不得销售。

图书在版编目(CIP)数据

信息论基础 = Elements of Information Theory /（美）科沃(Cover, T. M.)，（美）托马斯(Thomas J. A.)著. —影印本. —北京：清华大学出版社，2003
（国际知名大学原版教材. 信息技术学科与电气工程学科系列）
ISBN 7-302-07285-X

Ⅰ. 信…　Ⅱ. ①科…　②托…　Ⅲ. 信息论－高等学校－教材－英文　Ⅳ. G201

中国版本图书馆 CIP 数据核字（2003）第 084997 号

出 版 者：清华大学出版社　　**地　　址**：北京清华大学学研大厦
http://www.tup.com.cn　　**邮　　编**：100084
社总机：(010) 6277 0175　　**客户服务**：(010) 6277 6969
责任编辑：王一玲
印 刷 者：清华大学印刷厂
装 订 者：三河市新茂装订有限责任公司
发 行 者：新华书店总店北京发行所
开　　本：185×230　**印张**：35.75
版　　次：2003 年 11 月第 1 版　2003 年 11 月第 1 次印刷
书　　号：ISBN 7-302-07285-X/TP・5289
印　　数：1~3000
定　　价：55.00 元

国际知名大学原版教材

——信息技术学科与电气工程学科系列

出 版 说 明

郑大钟
清华大学信息科学与技术学院

当前，在我国的高等学校中，教学内容和课程体系的改革已经成为教学改革中的一个非常突出的问题，而为数不少的课程教材中普遍存在的“课程体系老化，内容落伍时代，本研层次不清”的现象又是其中的急需改变的一个重要方面。同时，随着科教兴国方针的贯彻落实，要求我们进一步转变观念扩大视野，使教学过程适应以信息技术为先导的技术革命和我国社会主义市场经济体制的需要，加快教学过程的国际化进程。在这方面，系统地研究和借鉴国外知名大学的相关教材，将会对推进我们的课程改革和推进我国大学教学的国际化进程，乃至对我们一些重点大学建设国际一流大学的努力，都将具有重要的借鉴推动作用。正是基于这种背景，我们决定在国内推出信息技术学科和电气工程学科国外知名大学原版系列教材。

本系列教材的组编将遵循如下的几点基本原则。(1)书目的范围限于信息技术学科和电气工程学科所属专业的技术基础课和主要的专业课。(2)教材的范围选自于具有较大影响且为国外知名大学所采用的教材。(3)教材属于在近5年内所出版的新书或新版书。(4)教材适合于作为我国大学相应课程的教材或主要教学参考书。(5)每本列选的教材都须经过国内相应领域的资深专家审看和推荐。(6)教材的形式直接以英文原版形式印刷出版。

本系列教材将按分期分批的方式组织出版。为了便于使用本系列教材的相关教师和学生从学科和教学的角度对其在体系和内容上的特点和特色有所了解，在每本教材中都附有我们所约请的相关领域资深教授撰写的影印版序言。此外，出于多样化的考虑，对于某些基本类型的课程，我们还同时列选了多于一本的不同体系、不同风格和不同层次的教材，以供不同要求和不同学时的同类课程的选用。

本系列教材的读者对象为信息技术学科和电气工程学科所属各专业的本科生，同时兼顾其他工程学科专业的本科生或研究生。本系列教材，既可采用作为相应课程的教材或教学参考书，也可提供作为工作于各个技术领域的工程师和技术人员的自学读物。

组编这套国外知名大学原版系列教材是一个尝试。不管是书目确定的合理性，教材选择的恰当性，还是评论看法的确切性，都有待于通过使用和实践来检验。感谢使用本系列教材的广大教师和学生的支持。期望广大读者提出意见和建议。

Elements of Information Theory

影 印 版 序

信息理论是当代国内外大学电子工程系、计算机系和统计系等为研究生和高年级本科生开设的一门基础专业课程。自 20 世纪 80 年代以来，在国内外广为流传的教材有 30 多种。Thomas M. Cover 所著的“Elements of Information Theory”一书在时间上虽不是最新的，但在美国是获得最广泛应用的一本教科书，如麻省理工学院（MIT）、斯坦福大学（Stanford University）、加州大学伯克利分校（University of California at Berkeley）等美国一流大学均采用本书作为该课程的教材或主要参考书。因此本书很值得向国内推荐。

本书的主要优点是：

（1）概念清晰。信息论涉及很多数学问题，其概念很多淹没在数学推导中，而本书将讲清概念放在第一位，且能深入浅出，使读者很快得其要领。

（2）数学工具的使用和数学推导过程的介绍恰到好处，既没有过于简化又没有拘泥于数学细节。

（3）理论与应用并重，既保证理论的完整性和系统性，又突出理论研究面向应用的性质，使读者能带着问题学，具有启发性。

（4）虽然此书出版较早（此后在美国又有三本教材问世），但从内容的覆盖面来讲此书仍有优势，且迄今仍有一定的先进性。

（5）与美国的其他教材类似，本书也不可避免地带有作者爱好的印记，书中的某些内容是作者偏爱而放入的，在一般的信息论课程中大都不介绍这些内容。作为参考书，扩大学生视野，这也是很有益的。选用本书的教员可以选择其中的若干章讲授。

本书讲述严谨，符号统一，各章有小结、有注解、有习题，是一本相当理想的英文教材。

朱雪龙　教授
清华大学电子工程系
2002 年 6 月

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

Schrödinger'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

Palo Alto, June 1991

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