Springer 国外信息科学与技术优秀图书系列

信息论基础

A First Course in Information Theory

（英文版）

〔加〕 Raymond W. Yeung 著

科 学 出 版 社

国外信息科学与技术优秀图书系列

信息论基础

（英文版）

A First Course in Information Theory

〔加〕　Raymond W. Yeung　著

科学出版社
北京

图字：01-2012-1768

内容简介

本书作者现为香港中文大学网络编码研究所主任，是网络编码理论的提出者之一。本书原版自 2002 年出版以来，被哥伦比亚大学、康奈尔大学、麻省理工学院、斯坦福大学等美国著名学府所采用，是信息理论方面的重要教材。本书首先介绍了信息论的经典内容，然后全面详细地论述了 *I*-度量、网络编码、Shannon 型与非 Shannon 型信息不等式等理论，以及熵函数与群论之间的关系。书中配有大量的实例、插图和习题，适合作为通信、电子信息、计算机等专业的高年级本科生和研究生的教材，也可供相关领域的科研人员参考。

图书在版编目(CIP)数据

信息论基础= A First Course in Information Theory : 英文 / (加) 杨伟豪著. —北京 : 科学出版社, 2012

(国外信息科学与技术优秀图书系列)

ISBN 978-7-03-034456-4

Ⅰ. ①信… Ⅱ. ①杨… Ⅲ. ①信息论-英文 Ⅳ.①G201

中国版本图书馆 CIP 数据核字(2012)第 108624 号

责任编辑：朱雪玲 / 责任印制：张　倩 / 封面设计：刘可红

科学出版社出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
双青印刷厂 印刷

科学出版社发行　各地新华书店经销

*

2012 年 7 月第　一　版　　开本：B5(720×1000)
2012 年 7 月第一次印刷　　印张：27
字数：430 000

定价：89.00 元

（如有印装质量问题，我社负责调换）

推 荐 序

信息论是信息科学的主要理论基础，它利用概率统计方法研究信息的表征、发送、传递、交换、接收和存储中的一般规律和本质属性，在通信、计算机网络、数字音像、信息处理等工程实践中得到广泛应用。特别地，信息论在提高现代通信系统的可靠性、有效性和保密性等方面提供了理论基础，是推动现代无线和有线通信技术发展的动力与源泉。作为一本入门性著作，本书通过大量实例，通俗易懂地介绍了香农创立的信息论的核心内容，同时也给出了一些前沿研究成果。

本书作者杨伟豪（Raymond W. Yeung）先生是国际知名的信息论学者、IEEE 会士（Fellow）、香港中文大学讲座教授。杨伟豪先生于 1988 年获美国康奈尔大学哲学博士（电机工程），曾供职于美国 AT&T Bell 实验室，之后长期从事信息理论、网络编码和性能分析的研究工作。作为网络编码研究领域的创始人之一，他与合作者的两篇网络编码奠基性研究论文至今已被引用近 6 000 次（Google Scholar 统计）。其中《线性网络编码》一文获得 IEEE 信息论学会 2005 年度最佳论文奖，这是三十多年来亚洲学者首次获此荣誉。此外，他还担任多个重要国际学术期刊的编委、副主编或主编，十多次在国际著名学术会议如 IEEE ISIT、IEEE ICC、ISITA 等做主旨演讲。

本书内容丰富翔实，自成体系地介绍了信息论的基本原理，覆盖了信息传输、信息处理和信息存储的重要内容。全书共 16 章，其中第 1~5 章和第 8~10 章介绍了信息论的经典内容，包括信息理论的基本概念、无失真信源编码定理、限失真信源编码定理和信道编码定理等基础理论；第 6、7、12、13 和 14 章引入了一些研究信息论的基本工具，包括 I-度量（I-Measure）、马尔可夫结构、信息不等式等；第 11 章和第 15 章分别讨论了单源和多源网络编码理论；第 16 章探讨了信息论与群论之间的内在关系。本书各章结尾均附有习题，并对信息论相关概念和重要结果给出了历史评述。

本书的最大特色在于，对基本概念和基础理论的阐述清晰明了，同时也反映了相关领域的进展。因此，本书适合作为高等院校信息与通信工程专业高年级本科生、研究生的教材或参考书。对于从事通信、计算机、信息处理等IT领域的工程技术人员，本书也是一本值得推荐的信息论经典读物。

范平志
（西南交通大学副校长、教授）
2012年6月于成都

Foreword

The first course usually is an appetizer. In the case of Raymond Yeung's *A First Course in Information Theory*, however, another delectable dish gets served up in each of the sixteen chapters. Chapters 1 through 7 deal with the basic concepts of entropy and information with applications to lossless source coding. This is the traditional early fare of an information theory text, but Yeung flavors it uniquely. No one since Shannon has had a better appreciation for the mathematical structure of information quantities than Prof. Yeung. In the early chapters this manifests itself in a careful treatment of information measures via both Yeung's analytical theory of I-Measure and his geometrically intuitive information diagrams. (This material, never before presented in a textbook, is rooted in works by G. D. Hu, by H. Dyckman, and by R. Yeung *et al.*) Fundamental interrelations among information measures and Markovianness are developed with precision and unity. New slants are provided on staples like the divergence inequality, the data processing theorem, and Fano's inequality. There is also a clever, Kraft-inequality-free way of proving that the average length of the words in a lossless prefix source code must exceed the source's entropy. An easily digestible treatment of the redundancy of lossless prefix source codes also is served up, an important topic in practice that usually is slighted in textbooks.

The concept of weakly typical sequences is introduced and then used to anchor Yeung's proof of the lossless block source coding theorem. The concept of strongly typical sequences is introduced next. Later extended to joint typicality, this provides a foundation for proving the channel coding theorem in Chapter 8, the lossy source coding (rate-distortion) theorem in Chapter 9, and selected multi-source network coding theorems in Chapter 15. Although the proof of the channel coding theorem follows standard lines, Yeung's tasteful development of the interplay between information quantities and Markovianness readies one's palate for a rigorous proof that feedback around a discrete memoryless channel does not increase its capacity. In most information theory books this basic result of Shannon either does not appear or is relegated to a problem in which the several steps are outlined in order to guide the reader toward the goal. Rate-distortion theory and Shannon's lossy source coding theorem are treated in familiar ways. When proving the latter, one confronts lack of independence of the events $\{(\mathbf{X}, \hat{\mathbf{X}}(i)) \in T^n\}$, where $\mathbf{X}$ is a random source word, $\hat{\mathbf{X}}(i)$ is the ith word in a randomly chosen source code, and T^n is the set of jointly typical vector pairs. In those instances in which this widely unappreciated stumbling block is not overlooked entirely, it usually is addressed via either a non-selfcontained reference or a mammoth problem at the end of the chapter However, Yeung's thorough earlier development of strong joint typicality concepts allows him to tackle it head-on.

Chapter 10 dishes up a careful treatment of the iterative algorithms for computation of channel capacity and rate-distortion functions pioneered by R. E. Blahut and S. Arimoto, which is generally accepted as today's preferred approach to computational information theory. Moreover, it has the extra advantage that iterative optimization algorithms are finding widespread application to areas as diverse as decoding of turbo and low-density parity-check codes and belief propagation in artificial intelligence and in real and artificial neural nets.

Chapters 11 through 16 are a unique tour de force. In as digestible a fashion as could possibly be expected, Yeung unveils a smorgasbord of topics in modern information theory that heretofore have been available only in research papers generated principally by Yeung and his research collaborators. Chapter 11 is a strong treatment of single-source network coding which develops carefully the relationships between information multicasting and the max-flow min-cut theory. Yeung makes an iron-clad case for how nodes must in general perform coding, not just storing and forwarding. Chapters 12, 13 and 14 on information inequalities of both Shannon and non-Shannon type constitute a definitive presentation of these topics by the master chef himself. Connections with linear programming are exploited, culminating in explication of Information Theory Inequality Prover (ITIP) of R. Yeung and Y.-O. Yan for inequalities of Shannon-type which comes with this book (also WWW-available). This leads, in turn, to the fascinating area of non-Shannon-type information inequalities, pioneered by R. Yeung and Z. Zhang. This material has been found to possess profound implications for the general area of information structures being studied by mathematical logicians and may also contribute to thermodynamics and statistical mechanics wherein the concept of entropy originated and which continue to be heavily concerned with various families of general inequalities The theory of I-Measure introduced in Chapter 6 provides the essential insight into those of the non-Shannon type inequalities that are discussed here. Multi-source network coding in Chapter 15 is a confounding area in which Yeung and others have made considerable progress but a comprehensive theory remains elusive. Nonetheless, the geometrical framework for information inequalities developed in Chapters 12 and 13 renders a unifying tool for attacking this class of problems. The closing chapter linking entropy to the theory of groups is mouthwateringly provocative, having the potential to become a major contribution of information theory to this renowned branch of mathematics and mathematical physics.

Savor this book; I think you will agree the proof is in the pudding.

Toby Berger
Irwin and Joan Jacobs Professor of Engineering
Cornell University, Ithaca, New York

Preface

Cover and Thomas wrote a book on information theory [52] ten years ago which covers most of the major topics with considerable depth. Their book has since become the standard textbook in the field, and it was no doubt a remarkable success. Instead of writing another comprehensive textbook on the subject, which has become more difficult as new results keep emerging, my goal is to write a book on the fundamentals of the subject in a unified and coherent manner

During the last ten years, significant progress has been made in understanding the entropy function and information inequalities of discrete random variables. The results along this direction not only are of core interest in information theory, but also have applications in network coding theory, probability theory, group theory, Kolmogorov complexity, and possibly physics. This book is an up-to-date treatment of information theory for discrete random variables, which forms the foundation of the theory at large. There are eight chapters on classical topics (Chapters 1, 2, 3, 4, 5, 8, 9, and 10), five chapters on fundamental tools (Chapters 6, 7, 12, 13, and 14), and three chapters on selected topics (Chapters 11, 15, and 16). The chapters are arranged according to the logical order instead of the chronological order of the results in the literature.

What is in this book

Out of the sixteen chapters in this book, the first thirteen chapters are basic topics, while the last three chapters are advanced topics for the more enthusiastic reader. A brief rundown of the chapters will give a better idea of what is in this book

Chapter 1 is a very high level introduction to the nature of information theory and the main results in Shannon's original paper in 1948 which founded the field. There are also pointers to Shannon's biographies and his works.

Chapter 2 introduces Shannon's information measures and their basic properties. Useful identities and inequalities in information theory are derived and explained. Extra care is taken in handling joint distributions with zero probability masses. The chapter ends with a section on the entropy rate of a stationary information source.

Chapter 3 is a discussion of zero-error data compression by uniquely decodable codes, with prefix codes as a special case. A proof of the entropy bound for prefix codes which involves neither the Kraft inequality nor the fundamental inequality is given. This proof facilitates the discussion of the redundancy of prefix codes.

Chapter 4 is a thorough treatment of weak typicality. The weak asymptotic equipartition property and the source coding theorem are discussed. An explanation of the fact that a good data compression scheme produces almost i.i.d. bits is given. There is also a brief discussion of the Shannon-McMillan-Breiman theorem.

Chapter 5 introduces a new definition of strong typicality which does not involve the cardinalities of the alphabet sets. The treatment of strong typicality here is more detailed than Berger [21] but less abstract than Csiszár and Körner [55]. A new exponential convergence result is proved in Theorem 5.3.

Chapter 6 is an introduction to the theory of I-Measure which establishes a one-to-one correspondence between Shannon's information measures and set theory. A number of examples are given to show how the use of information diagrams can simplify the proofs of many results in information theory. Most of these examples are previously unpublished. In particular, Example 6.15 is a generalization of Shannon's perfect secrecy theorem.

Chapter 7 explores the structure of the I-Measure for Markov structures. Set-theoretic characterizations of full conditional independence and Markov random field are discussed. The treatment of Markov random field here is perhaps too specialized for the average reader, but the structure of the I-Measure and the simplicity of the information diagram for a Markov chain is best explained as a special case of a Markov random field.

Chapter 8 consists of a new treatment of the channel coding theorem. Specifically, a graphical model approach is employed to explain the conditional independence of random variables. Great care is taken in discussing feedback.

Chapter 9 is an introduction to rate-distortion theory. The version of the rate-distortion theorem here, proved by using strong typicality, is a stronger version of the original theorem obtained by Shannon.

In Chapter 10, the Blahut-Arimoto algorithms for computing channel capacity and the rate-distortion function are discussed, and a simplified proof for convergence is given. Great care is taken in handling distributions with zero probability masses.

Chapter 11 is an introduction to network coding theory. The surprising fact that coding at the intermediate nodes can improve the throughput when an information source is multicast in a point-to-point network is explained. The max-flow bound for network coding with a single information source is explained in detail. Multi-source network coding will be discussed in Chapter 15 after the necessary tools are developed in the next three chapters.

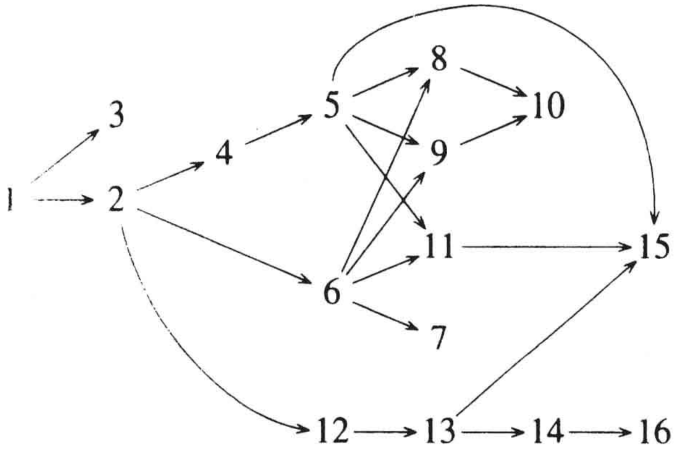
Information inequalities are sometimes called the laws of information theory because they govern the impossibilities in information theory. In Chapter 12, the geometrical meaning of information inequalities and the relation between information inequalities and conditional independence are explained in depth. The framework for information inequalities discussed here is the basis of the next two chapters.

Chapter 13 explains how the problem of proving information inequalities can be formulated as a linear programming problem. This leads to a complete characterization of all information inequalities which can be proved by conventional techniques. These are called Shannon-type inequalities, which can now be proved by the software ITIP which comes with this book. It is also shown how Shannon-type inequalities can be used to tackle the implication problem of conditional independence in probability theory.

All information inequalities we used to know were Shannon-type inequalities. Recently, a few non-Shannon-type inequalities have been discovered. This means that there exist laws in information theory beyond those laid down by Shannon. These inequalities and their applications are explained in depth in Chapter 14.

Network coding theory is further developed in Chapter 15. The situation when more than one information source are multicast in a point-to-point network is discussed. The surprising fact that a multi-source problem is not equivalent to a few single-source problems even when the information sources are mutually independent is clearly explained. Implicit and explicit bounds on the achievable coding rate region are discussed. These characterizations on the achievable coding rate region involve almost all the tools that have been developed earlier in the book, in particular, the framework for information inequalities.

Chapter 16 explains an intriguing relation between information theory and group theory. Specifically, for every information inequality satisfied by any joint distribution, there is a corresponding group inequality satisfied by any finite group and its subgroups, and vice versa. Inequalities of the latter type govern the orders of any finite group and their subgroups. Group-theoretic proofs of Shannon-type information inequalities are given. At the end of this chapter, a group inequality is obtained from a non-Shannon-type inequality discussed in Chapter 14. The meaning and the implication of this inequality are yet to be understood.



How to use this book

You are recommended to read the chapters according to the above chart. However, you will not have too much difficulty jumping around in the book because there should be sufficient references to the previous relevant sections.

As a relatively slow thinker, I feel uncomfortable whenever I do not reason in the most explicit way. This probably has helped in writing this book, in which all the derivations are from the first principle. In the book, I try to explain all the subtle mathematical details without sacrificing the big picture. Interpretations of the results are usually given before the proofs are presented. The book also contains a large number of examples. Unlike the examples in most books which are supplementary, the examples in this book are essential.

This book can be used as a reference book or a textbook. For a two-semester course on information theory, this would be a suitable textbook for the first semester. This would also be a suitable textbook for a one-semester course if only information theory for discrete random variables is covered. If the instructor also wants to include topics on continuous random variables, this book can be used as a textbook or a reference book in conjunction with another suitable textbook. The instructor will find this book a good source for homework problems because many problems here do not appear in any other textbook. A comprehensive instructor's manual is available upon request. Please contact the author at whyeung@ie.cuhk.edu.hk for information and access.

Just like any other lengthy document, this book for sure contains errors and omissions. To alleviate the problem, an errata will be maintained at the book homepage http://www.ie.cuhk.edu.hk/IT_book/.

RAYMOND W. YEUNG

Acknowledgments

Writing a book is always a major undertaking. This is especially so for a book on information theory because it is extremely easy to make mistakes in the writing if one has not done research work on that particular topic. Besides, the time needed for writing a book cannot be under estimated. Thanks to the generous support of a fellowship from the Croucher Foundation, I have had the luxury of taking a one-year leave in 2000-01 to write this book. I also thank my department for the support which made this arrangement possible.

There are many individuals who have directly or indirectly contributed to this book. First, I am indebted to Toby Berger who taught me information theory and writing. Venkat Anantharam, Dick Blahut, Dave Delchamps, Terry Fine, and Chris Heegard all had much influence on me when I was a graduate student at Cornell.

I am most thankful to Zhen Zhang for his friendship and inspiration. Without the results obtained through our collaboration, this book would not be complete. I would also like to thank Julia Abrahams, Vijay Bhargava, Agnes and Vincent Chan, Tom Cover, Imre Csiszár, Bob Gallager, Bruce Hajek, Te Sun Han, Jim Massey, Alon Orlitsky, Shlomo Shamai, Sergio Verdú, Victor Wei, Frans Willems, and Jack Wolf for their support and encouragement throughout the years. In particular, Imre Csiszár has closely followed my work and has given invaluable feedback. I would also like to thank all the collaborators of my work for their contribution and all the anonymous reviewers for their useful comments.

During the process of writing this book, I received tremendous help from Ning Cai and Fangwei Fu in proofreading the manuscript. Lihua Song and Terence Chan helped compose the figures and gave useful suggestions. Without their help, it would have been quite impossible to finish the book within the set time frame. I also thank Yu He, Andries Hekstra, Siu-Wai Ho, Kingo Kobayashi, Frank Kschischang, Amos Lapidoth, Li Ping, Prakash Narayan, Igal Sason, Wojtek Szpankowski, Emre Telatar, Chunxuan Ye, and Ken Zeger for their valuable inputs. The code for ITIP was written by Ying-On Yan, and Jack Lee helped modify the code and adapt it for the PC. Gordon Yeung helped on numerous Unix and LaTeX problems. The two-dimensional Venn diagram representation for four sets which is repeatedly used in the book was taught to me by Yong Nan Yeh.

On the domestic side, I am most grateful to my wife Rebecca for her love and support. My daughter Shannon has grown from an infant into a toddler during this time. She has added much joy to our family. I also thank my mother-in-law Mrs. Tsang and my sister-in-law Ophelia for coming over from time to time to take care of Shannon. Life would have been a lot more hectic without their generous help.

Contents