Springer 大学数学图书——影印版

Information and Coding Theory

信息和编码理论

Gareth A. Jones
J. Mary Jones 著

TSINGHUA UNIVERSITY PRESS
Springer

Springer 大学数学图书——影印版

Information and Coding Theory

信息和编码理论

TSINGHUA UNIVERSITY PRESS
北京

Springer

内容提要

信息和代数编码理论是数学、计算机科学和信息科学领域的重要学科。本书介绍信息理论、有限域和线性代数的基本知识，起点较低，易于学习；书中有大量例子和习题，并附有习题解答或提示。本书适合用作数学系、计算机科学系和信息科学系本科生高年级必修或选修课程教材或参考书。

Gareth A. Jones and J. Mary Jones
Information and Coding Theory
EISBN:1-85233-622-6

图书在版编目（CIP）数据

信息和编码理论 = Information and Coding Theory: 英文 /（美）卡雷斯（Jones G. A.），（美）玛丽著（Jones J. M.）. —影印本. —北京：清华大学出版社，2009.11
（Springer 大学数学图书）

ISBN 978-7-302-21488-5

Ⅰ. ①信…　Ⅱ. ①卡…　②玛…　Ⅲ. ①信息论－高等学校－教材－英文　②信源编码－编码理论－高等学校－教材－英文　③信道编码－编码理论－高等学校－教材－英文　Ⅳ. ①TN911.2

中国版本图书馆 CIP 数据核字（2009）第 200080 号

责任编辑：陈朝晖
责任印制：孟凡玉
出版发行：清华大学出版社　　地　址：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　邮　编：100084
社 总 机：010-62770175　　邮　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 刷 者：清华大学印刷厂
装 订 者：三河市金元印装有限公司
经　销：全国新华书店
开　本：175×245　**印　张**：14.5
版　次：2009 年 11 月第 1 版　　**印　次**：2009 年 11 月第 1 次印刷
印　数：1~4000
定　价：34.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题，请与清华大学出版社出版部联系调换。联系电话：010-62770177 转 3103　　产品编号：016920-01

Springer 大学数学图书——影印版

- Information and Coding Theory（信息和编码理论）
  Jones, Gareth A., Jones, J.Mary
- Symmetries（对称）
  Johnson, D.L.
- Matrix Groups: An Introduction to Lie Group Theory（矩阵群：李群理论基础）
  Baker, Andrew
- Essential Mathematical Biology（生物数学引论）
  Britton, Nicholas F.
- Special Relativity（狭义相对论）
  Woodhouse, N.M.J.
- Second Year Calculus（高等微积分）
  Bressoud, David M.
- Linear Functional Analysis（线性泛函分析）
  Rynne, Bryan P., Youngson, Martin A.
- A First Course in Discrete Mathematics（离散数学引论）
  Anderson, Ian
- Discrete Mathematics（离散数学）
  Lovász, L., Pelikán, J., Vesztergombi, K.
- Reading, Writing, and Proving（数学的读写和证明）
  Daepp Ulrich, Gorkin Pamela
- Basic Stochastic Processes（随机过程基础）
  Brzeźniak Zdzisław, Zastawniak Tomasz
- The Fundamental Theorem of Algebra（代数基本定理）
  Fine Benjamin, Rosenberger Gerhard
- The Heritage of Thales（泰勒斯的遗产）
  Anglin W. S., Lambek J.

Springer 大学数学图书——翻译版

- ○ 信息和编码理论（Information and Coding Theory）
  Jones, Gareth A., Jones, J.Mary
- ○ 对称（Symmetries）
  Johnson, D.L.
- ○ 矩阵群：李群理论基础（Matrix Groups: An Introduction to Lie Group Theory）
  Baker, Andrew
- ○ 生物数学引论（Essential Mathematical Biology）
  Britton, Nicholas F.
- ○ 狭义相对论（Special Relativity）
  Woodhouse, N.M.J.
- ○ 高等微积分（Second Year Calculus）
  Bressoud, David M.
- ○ 线性泛函分析（Linear Functional Analysis）
  Rynne, Bryan P., Youngson, Martin A.
- ○ 离散数学引论（A First Course in Discrete Mathematics）
  Anderson, Ian
- ○ 离散数学（Discrete Mathematics）
  Lovász, L., Pelikán, J., Vesztergombi, K.
- ○ 数学的读写和证明（Reading, Writing, and Proving）
  Daepp Ulrich, Gorkin Pamela
- ○ 随机过程基础（Basic Stochastic Processes）
  Brzeźniak Zdzisław, Zastawniak Tomasz
- ○ 代数基本定理（The Fundamental Theorem of Algebra）
  Fine Benjamin, Rosenberger Gerhard
- ○ 泰勒斯的遗产（The Heritage of Thales）
  Anglin W. S., Lambek J

序　言

在学校教书多年，当学生（特别是本科生）问有什么好的参考书时，我们所能推荐的似乎除了教材还是教材，而且不同教材之间的差别并不明显、特色也不鲜明。所以多年前我们就开始酝酿，希望为本科学生引进一些好的参考书，为此清华大学数学科学系的许多教授与清华大学出版社共同付出了很多心血。

这里首批推出的十余本图书，是从 Springer 出版社的多个系列丛书中精心挑选出来的。在丛书的筹划过程中，我们挑选图书最重要的标准并不是完美，而是有特色并包容各个学派（有些书甚至有争议，比如从数学上看也许不够严格），其出发点是希望我们的学生能够吸纳百家之长；同时，在价格方面，我们也做了很多工作，以使得本系列丛书的价格能让更多学校和学生接受，使得更多学生能够从中受益。

本系列图书按其定位，大体有如下四种类型（一本书可以属于多类，但这里限于篇幅不能一一介绍）。

一、适用面比较广、有特色并可以用作教材或参考书的图书。例如：

- Lovász et al.: Discrete Mathematics. 2003

  该书是离散数学的入门类型教材。与现有的教材（包括国外的教材）相比，它涵盖了离散数学新颖而又前沿的研究课题，同时还涉及信息科学方面既基本又有趣的应用；在着力打好数学基础的同时，也强调了数学与信息科学的关联。姚期智先生倡导和主持的清华大学计算机科学试验班，已经选择该书作为离散数学课程的教材。

二、在目前国内的数学教育中，课程主要以学科的纵向发展为主线，而对数学不同学科之间的联系讨论很少。学生缺乏把不同学科视为一个数学整体的训练，这方面的读物尤其欠缺。这是本丛书一个重要的着力点。最典型的是：

- Fine/Rosenberger: The Fundamental Theorem of Algebra. 1997

  该书对数学中最重要的定理——代数基本定理给出了六种证明，方法涉及到分析、代数与拓扑；附录中还给出了 Gauss 的证明和 Cauchy 的证明。全书以一个数学问题为主线展开，纵横数学的核心领域；结构严谨、文笔流畅、浅显易懂、引人入胜，是一本少见的能够让读者入迷的好读物，用它来引导学生欣赏和领会“数学的美”绝对不会落于空谈。该书适于自学、讨论，也是极好的短学期课程教材。

- Baker: Matrix Groups. 2001

  就内容而言，本书并不超出我国大学线性代数、抽象代数和一般拓扑学课程的内容，但是本书所讲的是李群和李代数的基础理论——这是现代数学和物理学非常重要的工具。各种矩阵群和矩阵代数是李群和李代数最典型和

最重要的例子，同时也能帮助学生建立数学不同学科之间的联系。从矩阵出发，既能把握李群和李代数的实质，又能学会计算和运用，所以这是一本不可多得的好书。

三、科学与技术的发展不断为数学提出新的研究课题，因此在数学学科的发展过程中，来自其他学科的推动力是不能忽视的。本系列中第三种类型的读物就是强调数学与其他学科的联系。例如：

- Woodhouse: Special Relativity. 2003

  该书将物理与数学有机结合，体现了物理学家伽利略的名言：“大自然是一部用数学语言写成的巨著。” 不仅如此，本书作者还通过对线性代数、微积分、场论等数学的运用进一步强调并贯穿这样的观点：数学的真谛和发展存在并产生于物理或自然规律及其发现中。精读此书有助于理解物理学和数学的整体关系。

- Britton: Essential Mathematical Biology. 2003

  生命科学在本世纪一定会有很大发展，其对数学的需求和推动是可以预见的。因此生物数学在应用数学中占有日益重要的地位，数学系培养的学生至少一部分人应当对这个领域有所了解。随着生命科学的迅速发展，生物数学也发展很快。本书由浅入深，从经典的问题入手，最后走向学科前沿和近年的热点问题。该书至少可以消除学生对生物学的神秘感。

四、最后一类是适合本科学生的课外读物。这类图书对激发和引导学生学习数学的兴趣会非常有帮助，而且目前国内也急需这样的图书。例如：

- Daepp/Gorkin: Reading, Writing and Proving. 2003

  该书对初学高等数学的读者来说特别有意义。它的基本出发点是引导读者以研究的心态去学习，让读者养成独立思考的习惯，并进而成为研究型的学习者。该书将一个学习数学的过程在某种意义下程序化，努力让学习者养成一个好的学习习惯，以及学会如何应对问题。该书特色鲜明，类似的图书确实很少。

- Brzezniak/Zastawniak: Basic Stochastic Processes. 1998

  随机过程理论在数学、科学和工程中有越来越广泛的应用，本书适合国内的需要。其主要特点是：书中配有的习题是巩固和延伸学习内容的基本手段，而且有十分完整的解答，非常适合自学和作为教学参考书。这是一本难得的好书，它 1999 年出版，到 2000 年已经是第 3 次印刷，到 2003 年则第 6 次重印。

- Anglin/Lambek: The Heritage of Thales. 1995

  该书的基本内容是数学的历史和数学的哲学。数学历史是该书的线索，数学是内容的主体，引申到的是数学哲学。它不是一本史论型的著作，而是采用专题式编写方式，每个专题相对独立，所以比较易读、易懂，是本科生学习数学过程中非常好的课外读物。

本系列丛书中的大部分图书还将翻译为中文出版，以适应更多读者的需要。丛书筹划过程中，冯克勤、郑志勇、卢旭光、郑建华、王殿军、杨利军、叶俊、扈志明等很多清华大学的教授都投入了大量精力。他们之中很多人也将是后面中文版的译者。此外，我们今后还将不断努力丰富引进丛书的种类，同时也会将选书的范围在可能情况下进一步扩大到其他高水平的出版机构。

教育是科学和技术发展的基石，数学教育更是基石的基础。因为是基础所以它重要；也因为是基础所以它显示度不高，容易不被重视。只有将人才培养放到更高的地位上，中国成为创新型国家的目标才会成为可能。

本系列丛书的正式推出，圆了一个我们多年的梦，但这无疑仅仅是开始。

白峰杉

2006 年 6 月于清华园

Preface

As this Preface is being written, the twentieth century is coming to an end. Historians may perhaps come to refer to it as the century of information, just as its predecessor is associated with the process of industrialisation. Successive technological developments such as the telephone, radio, television, computers and the Internet have had profound effects on the way we live. We can see pictures of the surface of Mars or the early shape of the Universe. The contents of a whole shelf-load of library books can be compressed onto an almost weightless piece of plastic. Billions of people can watch the same football match, or can keep in instant touch with friends around the world without leaving home. In short, massive amounts of information can now be stored, transmitted and processed, with surprising speed, accuracy and economy.

Of course, these developments do not happen without some theoretical basis, and as is so often the case, much of this is provided by mathematics. Many of the first mathematical advances in this area were made in the mid-twentieth century by engineers, often relying on intuition and experience rather than a deep theoretical knowledge to lead them to their discoveries. Soon the mathematicians, delighted to see new applications for their subject, joined in and developed the engineers' practical examples into wide-ranging theories, complete with definitions, theorems and proofs. New branches of mathematics were created, and several older ones were invigorated by unexpected applications: who could have predicted that error-correcting codes might be based on algebraic curves over finite fields, or that cryptographic systems might depend on prime numbers?

Information Theory and Coding Theory are two related aspects of the problem of how to transmit information efficiently and accurately from a source, through a channel, to a receiver. This includes the problem of how to store information, where the receiver may be the same as the source (but later in time). As an example, space exploration has created a demand for accurate transmission of very weak signals through an extremely noisy channel: there is no point in sending a probe to Mars if one cannot hear and decode the messages it sends back. In its simplest form this theory uses elementary techniques from Probability Theory and Algebra, though later advances have been based on such topics as Combinatorics and Algebraic Geometry.

One important problem is how to compress information, in order to transmit it rapidly or store it economically. This can be done by reducing redundancy: a familiar example is the use of abbreviations and acronyms such as "UK", "IBM" and "radar" in place of full names, many of whose symbols are redundant from the point of view of information content. Similarly, we often shorten the names of our closest friends and relatives, so that William becomes Will or Bill.

Another important problem is how to detect and correct errors in information. Human beings and machines cannot be relied upon always to avoid mistakes, and if these are not corrected the consequences can be serious. Here the solution is to *increase* redundancy, by adding symbols which reinforce and protect the message. Thus the NATO alphabet Alpha, Bravo, Charlie, ..., used by armed forces, airlines and emergency services for spoken communication, replaces the letters A, B, C, ... with words which are chosen to sound as unlike each other as possible: for instance, B and V are often confused (they are essentially the same in some languages), but Victor is unlikely to be mistaken for Bravo, even when misheard as Bictor.

Information Theory, much of which stems from an important 1948 paper of Shannon [Sh48], uses probability distributions to quantify information (through the entropy function), and to relate it to the average word-lengths of encodings of that information. In particular, Shannon's Fundamental Theorem guarantees the existence of good error-correcting codes, and the aim of Coding Theory is to use mathematical techniques to construct them, and to provide effective algorithms with which to use them. Despite its name, Coding Theory does not involve the study of secret codes: this subject, Cryptography, is closely related to Information Theory through the concepts of entropy and redundancy, but since the mathematical techniques involved tend to be rather different, we have not included them.

This book, based on a third-year undergraduate course introduced at Southampton University in the early 1980s, is an attempt to explain the basic ideas of Information and Coding Theory. The main prerequisites are elementary Probability Theory and Linear Algebra, together with a little Calculus, as covered in a typical first-year university syllabus for mathematicians, engineers or scientists. Most textbooks in this area concentrate mainly or entirely on either Information Theory or Coding Theory. However, the two subjects are intimately related (through Shannon's Theorem), and we feel that there are strong arguments for learning them together, at least initially.

Chapters 1–5 (representing about 60% of the main text) are mainly on Information Theory. Chapter 1, which has very few prerequisites, shows how to encode information in such a way that its subsequent decoding is unambiguous and instantaneous: the main results here are the Sardinas–Patterson Theorem (proved in Appendix A), and the Kraft and MacMillan inequalities, concerning the existence of such codes. Chapter 2 introduces Huffman codes, which — rather like Morse code — minimise average word-length by systematically assigning shorter code-words to more frequent symbols; here (as in Chapters 3–5) we use some elementary Probability Theory, namely finite probability distributions. In Chapter 3 we use the entropy function, based on probabilities and their logarithms, to measure information and to relate it, through a theorem of Shannon, to the average word-lengths of encodings. Chapter 4 studies how information is transmitted through a channel, possibly subject to distortion by "noise" which may introduce errors; conditional probabilities allow us to define certain system entropies, which measure information from several points of view, such as those of the sender and the receiver. These lead to the concept of channel capacity, which is the maximum amount of information a channel can transmit. In Chapter 5 we meet Shannon's Fundamental Theorem, which states that, despite noise, information can be transmitted through a channel with arbitrarily great accuracy, at rates arbitrarily close to the channel capacity. We sketch a proof of this in the simple but important case of the binary symmetric channel; a full proof for this channel, given in Appendix C, relies on the only advanced result we need from Probability Theory, namely the Law of Large Numbers, explained in Appendix B.

The basic idea of Shannon's Theorem is that one can transmit information accurately by using code-words which are sufficiently unlike each other that, even if some of their symbols are incorrect, the receiver is unlikely to confuse them (think of Bravo and Victor). Unfortunately, neither the theorem nor its proof shows us how to find specific examples of such codes, and this is the aim of Coding Theory, the subject matter of Chapters 6 and 7. In these chapters, which are rather longer than their predecessors, we introduce a number of fairly simple examples of error-correcting codes. In Chapter 6 we use elementary, direct methods for this; the main result here is Hamming's sphere-packing bound, which uses a simple geometric idea to give an upper bound on the number of code-words which can correct a given number of errors. In Chapter 7 we construct slightly more advanced examples of error-correcting codes using Linear Algebra and Matrix Theory, specifically the concepts of vector spaces and subspaces, bases and dimensions, matrix rank, and row and column operations. We also briefly show how some ideas from Combinatorics and Geometry, such as block designs and projective geometries, are related to codes.

The usual constraints of space and time have forced us to omit several interesting topics, such as the links with Cryptography mentioned above, and only briefly to sketch a few others. In Information Theory, for instance, Markov sources (those with a "memory" of previous events) appear only as an exercise, and similarly in Coding Theory we have not discussed cyclic codes and their connections with polynomial rings. Instead, we give some suggestions for further reading at the end of the book.

The lecture course on which this book is based follows Chapters 1–7, usually omitting Sections 5.5, 6.5, 6.6 and 7.5 and the Appendices. A course on Information Theory could be based on Chapters 1–5, perhaps with a little more material on Markov sources or on connections with Cryptography. A course on Coding Theory could follow Chapters 6 and 7, with some background material from Chapter 5 and some extra material on, for instance, cyclic codes or weight enumerators.

We have tried, wherever possible, to give credit to the originators of the main ideas presented in this book, and to acknowledge the sources from which we have obtained our results, examples and exercises. No doubt we have made omissions in this respect: if so, they are unintentional, and no slight was intended.

We are grateful to Keith Lloyd and Robert Syddall, who have both taught and improved the course on which this book is based, together with the hundreds of students whose reactions to the course have been so instructive. We thank Karen Barker, Beverley Ford, David Ireland and their colleagues at Springer for their advice, encouragement and expertise during the writing of this book. We are indebted to W.S. (further symbols are surely redundant) for providing the quotations which begin each chapter, and finally we thank Peter and Elizabeth for tolerating their occasionally distracted parents with unteenagerly patience and good humour.

Notes to the Reader

Chapters 1–5 cover the basic material on Information Theory, and they should be read in that order since each depends fairly heavily on its predecessors. The Sardinas–Patterson Theorem (§1.2) and Shannon's Fundamental Theorem (§5.4) are important results with rather long proofs; we have simply outlined the proofs in the text, and the complete proofs in Appendices A and C can be omitted on first reading since their details are not required later. Other sections not required later are §5.5, §6.5, §6.6 and §7.5.

In a sense, the book starts afresh with Chapters 6 and 7, which are about Coding Theory. These two chapters could be read on their own, though it would help to look first at some of Chapter 5, in particular §5.2 for the example of repetition codes, §5.3 for the concept of Hamming distance, and §5.4 and §5.6 for the motivation provided by Shannon's Theorem.

We assume familiarity with some of the basic concepts of Probability Theory (in Chapters 1–5) and Linear Algebra (in Chapters 6 and 7), together with a few results from Calculus; there is some suggested background reading on these topics at the end of the book, in the section *Suggestions for Further Reading*, together with some comments on further reading in Information and Coding Theory.

The exercises are an important feature of the book. Those embedded in the text are designed to test and reinforce the reader's understanding of the concepts immediately preceding them. Most of these are fairly straightforward, and it is best to attempt them right away, before reading further. The supplementary exercises at the end of each chapter are often more challenging; they may require several ideas from that chapter, and possibly also from earlier chapters. Some of these supplementary exercises are designed to encourage the reader to explore the theory further, beyond the topics we have covered. Solutions of all the exercises are given at the end of the book, but it is strongly recommended to try the exercises first before reading the solutions.

Contents

Preface ix

Notes to the Reader xvii

1. Source Coding 1
- 1.1 Definitions and Examples 1
- 1.2 Uniquely Decodable Codes 4
- 1.3 Instantaneous Codes 9
- 1.4 Constructing Instantaneous Codes 11
- 1.5 Kraft's Inequality 13
- 1.6 McMillan's Inequality 14
- 1.7 Comments on Kraft's and McMillan's Inequalities 16
- 1.8 Supplementary Exercises 17

2. Optimal Codes 19
- 2.1 Optimality 19
- 2.2 Binary Huffman Codes 22
- 2.3 Average Word-length of Huffman Codes 26
- 2.4 Optimality of Binary Huffman Codes 27
- 2.5 r-ary Huffman Codes 28
- 2.6 Extensions of Sources 30
- 2.7 Supplementary Exercises 32

3. Entropy 35
- 3.1 Information and Entropy 35
- 3.2 Properties of the Entropy Function 40
- 3.3 Entropy and Average Word-length 42

3.4 Shannon–Fano Coding 45
3.5 Entropy of Extensions and Products 47
3.6 Shannon's First Theorem 48
3.7 An Example of Shannon's First Theorem 49
3.8 Supplementary Exercises 51

4. Information Channels 55
4.1 Notation and Definitions 55
4.2 The Binary Symmetric Channel 60
4.3 System Entropies 62
4.4 System Entropies for the Binary Symmetric Channel 64
4.5 Extension of Shannon's First Theorem to Information Channels 67
4.6 Mutual Information 70
4.7 Mutual Information for the Binary Symmetric Channel 72
4.8 Channel Capacity 73
4.9 Supplementary Exercises 76

5. Using an Unreliable Channel 79
5.1 Decision Rules 79
5.2 An Example of Improved Reliability 82
5.3 Hamming Distance 85
5.4 Statement and Outline Proof of Shannon's Theorem 88
5.5 The Converse of Shannon's Theorem 90
5.6 Comments on Shannon's Theorem 93
5.7 Supplementary Exercises 94

6. Error-correcting Codes 97
6.1 Introductory Concepts 97
6.2 Examples of Codes 100
6.3 Minimum Distance 104
6.4 Hamming's Sphere-packing Bound 107
6.5 The Gilbert–Varshamov Bound 111
6.6 Hadamard Matrices and Codes 114
6.7 Supplementary Exercises 118

7. Linear Codes 121
7.1 Matrix Description of Linear Codes 121
7.2 Equivalence of Linear Codes 127
7.3 Minimum Distance of Linear Codes 131
7.4 The Hamming Codes 133
7.5 The Golay Codes 136
7.6 The Standard Array 141

7.7 Syndrome Decoding 143
7.8 Supplementary Exercises 146

Suggestions for Further Reading 149

Appendix A. Proof of the Sardinas–Patterson Theorem 153

Appendix B. The Law of Large Numbers 157

Appendix C. Proof of Shannon's Fundamental Theorem 159

Solutions to Exercises 165

Bibliography 191

Index of Symbols and Abbreviations 195

Index 201

1
Source Coding

Words, words, words. (*Hamlet*)

This chapter considers how the information emanating from a source can be encoded, so that it can later be decoded unambiguously and without delay. These two requirements lead to the concepts of uniquely decodable and instantaneous codes. We shall find necessary and sufficient conditions for a code to have these properties, we shall see how to construct such codes, and we shall prove Kraft's and McMillan's inequalities, which essentially say that such codes exist if and only if they have sufficiently long code-words.

1.1 Definitions and Examples

Information theory is concerned with the transmission of information from a sender, through a channel, to a receiver. The sender and receiver could be people or machines. In most cases they are different, but when information is being stored for later retrieval, the receiver could be the sender at some future time. We will assume that the information comes from a *source* $\mathcal{S}$, which emits a sequence $\mathbf{s} = X_1X_2X_3\dots$ of symbols X_n; for instance, X_n might be the n-th symbol in some message, or the outcome of the n-th repetition of some experiment. In practice, this sequence will always be finite (nothing lasts for ever), but for theoretical purposes it is sometimes useful also to consider infinite