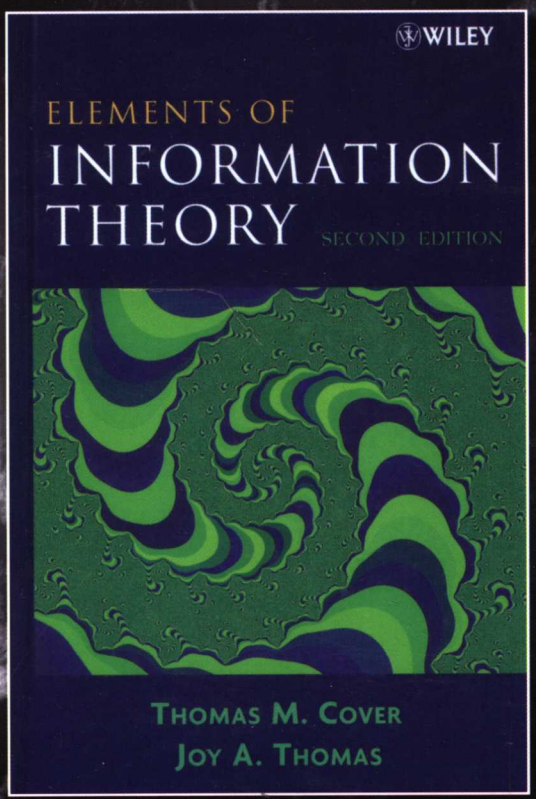
HZ BOOKS
WILEY

计 算 机 科 学 丛 书

原书第2版

信息论基础

(美) Thomas M. Cover Joy A. Thomas 著 阮吉寿 张华 译 沈世镒 审校



Elements of Information Theory

Second Edition

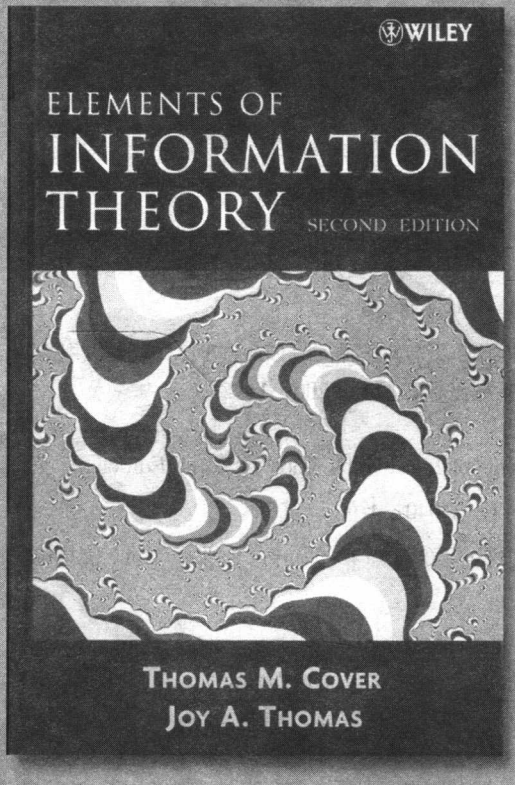
机械工业出版社
China Machine Press

计 算 机 科 学 丛 书

原书第2版

信息论基础

(美) Thomas M. Cover Joy A. Thomas 著 阮吉寿 张华 译 沈世镒 审校



Elements of Information Theory
Second Edition

机械工业出版社
China Machine Press

本书是信息论领域中一本简明易懂的教材。主要内容包括:熵、信源、信道容量、率失真、数据压缩与编码理论和复杂度理论等方面的介绍。本书还对网络信息论和假设检验等进行了介绍,并且以赛马模型为出发点,将对证券市场的研究纳入了信息论的框架,从新的视角给投资组合的研究带来了全新的投资理念和研究技巧。

本书适合作为电子工程、统计学以及电信方面的高年级本科生和研究生的信息论基础教程教材,也可供研究人员和专业人士参考。

Thomas M. Cover, Joy A. Thomas: Elements of Information Theory, Second Edition (ISBN-13 978-0-471-24195-9, ISBN-10 0-471-24195-4)

本书版权登记号:图字:01-2005-5619

图书在版编目(CIP)数据

信息论基础(原书第2版)/(美)科弗(Cover, T. M.),(美)托马斯(Thomas, J. A.)著;阮吉寿,张华译. -北京:机械工业出版社,2007.11
(计算机科学丛书)
书名原文:Elements of Information Theory, Second Edition
ISBN 978-7-111-22040-4

Ⅰ.信… Ⅱ.①科… ②托… ③阮… ④张… Ⅲ.信息论-高等学校-教材 Ⅳ.G201

中国版本图书馆CIP数据核字(2007)第116475号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)
责任编辑:王 玉
北京京北制版厂印刷·新华书店北京发行所发行
2008年1月第1版第1次印刷
184mm×260mm·28.25印张
定价:58.00元

www.hzbook.com

填写读者调查表　加入华章书友会
获赠精彩技术书　参与活动和抽奖

尊敬的读者：

感谢您选择华章图书。为了聆听您的意见，以便我们能够为您提供更优秀的图书产品，敬请您抽出宝贵的时间填写本表，并按底部的地址邮寄给我们（您也可通过www.hzbook.com填写本表）。您将加入我们的“华章书友会”，及时获得新书资讯，免费参加书友会活动。我们将定期选出若干名热心读者，免费赠送我们出版的图书。请一定填写书名书号并留全您的联系信息，以便我们联络您，谢谢！

书名：　　书号：7-111-(　　)

姓名：	性别：☐男　☐女	年龄：	职业：
通信地址：		E-mail：	
电话：	手机：	邮编：	

1. 您是如何获知本书的：

☐朋友推荐　☐书店　☐图书目录　☐杂志、报纸、网络等　☐其他

2. 您从哪里购买本书：

☐新华书店　☐计算机专业书店　☐网上书店　☐其他

3. 您对本书的评价是：

技术内容	☐很好	☐一般	☐较差	☐理由_____
文字质量	☐很好	☐一般	☐较差	☐理由_____
版式封面	☐很好	☐一般	☐较差	☐理由_____
印装质量	☐很好	☐一般	☐较差	☐理由_____
图书定价	☐太高	☐合适	☐较低	☐理由_____

4. 您希望我们的图书在哪些方面进行改进？

_____

_____

5. 您最希望我们出版哪方面的图书？如果有英文版请写出书名。

_____

_____

6. 您有没有写作或翻译技术图书的想法？

☐是，我的计划是_____☐否

7. 您希望获取图书信息的形式：

☐邮件　☐信函　☐短信　☐其他_____

请寄：北京市西城区百万庄南街1号　机械工业出版社　华章公司　计算机图书策划部收
邮编：100037　电话：（010）88379512　传真：（010）68311602　E-mail: hzjsj@hzbook.com

译 者 序

Cover M. Thomas 与 Joy A. Thomas 的信息论基础可谓跨世纪的一本好书，其读者人数在信息论领域名列榜首。说本书是信息论领域中的 Bible（圣经），也不算过分。本书涉及的相关知识领域广泛，我们第一次接到翻译此书的任务时，多少有些惶恐，担心无法准确地将 Cover 的精神和深刻的内涵活灵活现地呈现给读者。1985 年 Cover 曾经是沈世镒教授的老师。沈先生回国后在南开大学带出了许许多多的优秀学生。他们在国内乃至国际上都是信息论的骨干和学术带头人（比如，杨恩辉，孙凤文，张箴，符方伟，叶中行，岳殿武，陈鲁生等，他们曾以南开大学的信息论为荣，南开大学的信息论现在又以他们为荣）。为报 Cover 之师恩，也为更多不曾在南开大学学习的广大信息论学子能够领略 Cover 的大师风范，我们欣然接受了此项翻译任务，并且力争不辱使命。

本书可谓信息量巨大的好书。在熵、信道、信源、数据压缩与编码理论，复杂度理论等方面独具特色，网络信息论更是一个新的亮点。本书还以赛马模型为出发点，将证券市场的研究纳入信息论的框架内研究，给证券市场研究以一个新的视角。更难得的是，作者利用自己深厚的研究功力，将这三部分有机地结合在一起，不仅增加了信息论的内涵，也增加了读者群。特别是研究投资组合者，在适当学习第 2 章与第 11 章的基础上，读懂第 6 章与第 16 章，将会带来全新的投资理念和证券研究的新技巧。

本书的写作风格独特，横跨信息论、信号学、计算机逻辑、概率论、图论以及金融等若干领域。因此，为了使得本书的翻译风格尽可能完整，并保持其在各领域的特色，我们在翻译中颇费心思，字斟句酌，反复思考，同时，虚心地请教南开大学从事相应领域的同事，在此，对他们表示感谢。我们的许多研究生在第 1 版和第 2 版的翻译和校对的过程中也做出了贡献。而且，在第 2 版翻译时，我们虚心听取了第 1 版的读者的反馈意见，特在此向他们表示衷心感谢。最后，我们要对机械工业出版社华章分社表示感谢，编辑们的认真、仔细和热情合作提高了本书的翻译质量。

译 者

2007 年 7 月

第 2 版前言

自从本书第 1 版出版以来，我们希望书中的许多方面能得到改进、重新编排或者扩充，但是需再版的限制并不允许我们在已经出版的书中实现这样的愿望。而今在出新版之际，我们终于有机会对原书做些改变，增加一些习题，同时，讨论一些在第 1 版中忽略的专题。

本书主要的变化包括：各章的重新编排，使得本书更易于教学；还增加了 200 多个新习题。在某些专题中，我们也增加了一些素材，如在普适性投资组合理论、通用信源编码、高斯反馈信道容量、网络信息论等方面，并且阐述了数据压缩和信道容量的对偶性。另外，本书还新增加了一章，同时对原书中大量的证明过程进行简化，而且更新了参考文献和历史回顾点评。

本书可以分成两个学期学习。建议第一学期学习第 1～9 章，包括渐近均分性、数据压缩和信道容量，结束于高斯信道容量。第二学期学习余下的几章，包括率失真理论、型方法、科尔莫戈罗夫复杂度、网络信息论、通用信源编码和投资组合理论。如果只开一个学期的课，建议将率失真、科尔莫戈罗夫复杂度和网络信息论加入第一学期的教学中，其中后两者只需各上一节课。

自第 1 版以来，信息论迎来了它的 50 岁生日（香农的领域开创性文章 50 周年纪念），源自信息论的许多思想已经广泛应用于科学技术的众多问题，如生物信息学、网络搜索、无线通信、视频压缩以及其他等。信息论的应用是无止境的，然而其完美的数学理论始终是该领域最引人注目的地方。我们希望借此书给大家带来某些共识，使得大家坚信在涉及数学、物理学、统计学和工程学的交叉领域中，信息论是最有趣的领域之一。

TOM COVER
JOY THOMAS

Palo Alto, California
2006 年 1 月

第 1 版前言

本书是一本简明易懂的信息论教材。正如爱因斯坦所说："凡事应该尽可能使其简单到不能再简单为止。"虽然我们没有深入考证过该引语的来源（据说最初是在幸运蛋卷中发现的），但我们自始至终都将这种观点贯穿到本书的写作中。信息论中的确有这样一些关键的思想和技巧，一旦掌握了它们，不仅使信息论的主题简明，而且在处理新问题时提供重要的直觉。

本书来自使用了十多年的信息论讲义，原讲义是信息论课程的高年级本科生和一年级研究生两学期用的教材。本书打算作为通信理论、计算机科学和统计学专业学生学习信息论的教材。

信息论中有两个简明要点。第一，熵与互信息这样的特殊量是为了解答基本问题而产生的。例如，熵是随机变量的最小描述复杂度，互信息是度量在噪声背景下的通信速率。另外，我们在以后还会提到，互信息相当于已知边信息条件下财富双倍的增长。第二，回答信息理论问题的答案具有自然的代数结构。例如，熵具有链式法则，因而，熵和互信息也是相关的。因此，数据压缩和通信中的问题得到广泛的解释。我们都有这样的感受，当研究某个问题时，往往历经大量的代数运算推理得到了结果，但此时没有真正了解问题的全貌，最终是通过反复观察结果，才对整个问题有完整、明确的认识。所以，对一个问题的全面理解，不是靠推理，而是靠对结果的观察。要更具体地说明这一点，物理学中的牛顿三大定律和薛定谔波动方程也许是最合适的例子。谁曾预见过薛定谔波动方程后来会有如此令人敬畏的哲学解释呢？

在本书中，我们常会在着眼于问题之前，先了解一下答案的性质。比如第 2 章中，我们定义熵、相对熵和互信息，研究它们之间的关系，再对这些关系作一点解释，由此揭示如何融会贯通地使用各式各样的方法解决实际问题。同理，我们顺便探讨热力学第二定律的含义。熵总是增加吗？答案既肯定也否定。这种结果会令专家感兴趣，但初学者或许认为这是必然的而不会深入考虑。

在实际教学中，教师往往会加入一些自己的见解。事实上，寻找无人知道的证明或者有所创新的结果是一件很愉快的事情。如果有人将新的思想和已经证明的内容在课堂上讲解给学生，那么不仅学生会积极反馈"对，对，对"，而且会大大地提升教授该课程的乐趣。我们正是这样从研究本教材的许多新想法中获得乐趣的。

本书加入的新素材实例包括信息论与博弈之间的关系，马尔可夫链背景下热力学第二定律的普遍性问题，信道容量定理的联合典型性证明，赫夫曼码的竞争最优性，以及关于最大熵谱密度估计的伯格（Burg）定理的证明。科尔莫戈罗夫复杂度这一章也是本书的独到之处。而将费希尔信息，互信息、中心极限定理以及布伦－闵可夫斯基不等式与熵幂不等式联系在一起，也是我们引以为豪之处。令我们感到惊讶的是，关于行列式不等式的许多经典结论，当利用信息论不等式后会很容易得到证明。

自从香农的奠基性论文面世以来，尽管信息论已有了相当大的发展，但我们还是要努力强调它的连贯性。虽然香农创立信息论时受到通信理论中的问题启发，然而我们认为信息论是一门独立的学科，可应用于通信理论和统计学中。我们将信息论作为一个学科领域从通信理论、概率论和统计学的背景中独立出来，因为明显不可能从这些学科中获得难以理解的信息概念。

由于本书中绝大多数结论以定理和证明的形式给出，所以，我们期望通过对这些定理的巧妙证明能说明这些结论的完美性。一般来讲，我们在介绍问题之前先描述问题的解的性质，而这些很有趣的性质会使接下来的证明顺理成章。

使用不等式串、中间不加任何文字、最后直接加以解释，是我们在表述方式上的一项创新。希望读者学习我们所给的证明过程达到一定数量时，在没有任何解释的情况下就能理解其中的大部分步骤，并自己给出所需的解释。这些不等式串好比模拟测试题，读者可以通过它们确认自己是否已掌握证明那些重要定理的必备知识。这些证明过程的自然流程是如此引人注目，以至于导致我们轻视了写作技巧中的某条重要原则。由于没有多余的话，因而突出了思路的逻辑性与主题思想。我们希望当读者阅读完本书后，能够与我们共同分享我们所推崇的，具有优美、简洁和自然风格的信息论。

本书广泛使用弱的典型序列的方法，此概念可以追溯到香农 1948 年的创造性工作，而它真正得到发展是在 20 世纪 70 年代初期。其中的主要思想就是所谓的渐近均分性（AEP），或许可以粗略地说成“几乎一切事情都是等可能的”。

第 2 章阐述了熵、相对熵和互信息之间的基本代数关系。渐近均分性是第 3 章重中之重的内容，这也使我们将随机过程和数据压缩的熵率分别放在第 4 章和第 5 章中论述。第 6 章介绍博弈，研究了数据压缩的对偶性和财富的增长率。

可作为对信息论进行理性思考基础的科尔莫戈罗夫复杂度，拥有着巨大的成果，放在第 14 章中论述。我们的目标是寻找一个通用的最短描述，而不是平均意义下的次佳描述。的确存在这样的普遍性概念用来刻画一个对象的复杂度。该章也论述了神奇数 Ω，揭示数学上的不少奥秘，是图灵机停止运转概率的推广。

第 7 章论述信道容量定理。第 8 章叙述微分熵的必需知识，它们是将早期容量定理推广到连续噪声信道的基础。基本的高斯信道容量问题在第 9 章中论述。

第 11 章阐述信息论和统计学之间的关系，20 世纪 50 年代初期库尔贝克（Kullback）首次对此进行了研究，此后相对被忽视。由于率失真理论比无噪声数据压缩理论需要更多的背景知识，因而将其放置在正文中比较靠后的第 10 章。

网络信息理论是个大的主题，安排在第 15 章，主要研究的是噪声和干扰存在情形下的同时可达的信息流。有许多新的思想在网络信息理论中开始活跃起来，其主要新要素有干扰和反馈。第 16 章讲述股票市场，这是第 6 章所讨论的博弈的推广，也再次表明了信息论和博弈之间的紧密联系。

第 17 章讲述信息论中的不等式，我们借此一隅把散布于全书中的有趣不等式重新收拢在一个新的框架中，再加上一些关于随机抽取子集熵率的有趣新不等式。集合和的体积的布伦－闵可夫斯基不等式，独立随机变量之和的有效方差的熵幂不等式以及费希尔信息不等式之间的美妙关系也将在此章中得到详尽的阐述。

本书力求推理严密，因此对数学的要求相当高，要求读者至少学过一学期的概率论课程且有扎实的数学背景，大致为本科高年级或研究生一年级水平。尽管如此，我们还是努力避免使用测度论。因为了解它只对第 16 章中的遍历过程的 AEP 的证明过程起到简化作用。这符合我们的观点，那就是信息论基础与技巧不同，后者才需要将所有推广都写进去。

本书的主体是第 2，3，4，5，7，8，9，10，11 和 15 章，它们自成体系，读懂了它们就可以对信息论有很好的理解。但在我们看来，第 14 章的科尔莫戈罗夫复杂度是深入理解信息论所需的必备知识。余下的几章，从博弈到不等式，目的是使主题更加连贯和完美。

任何教程都有它的第一讲，目的是给出其主要思想的简短预览和概述。本书的第 1 章就是为这个目的而设置的。

TOM COVER
JOY THOMAS

Palo Alto, California
1990 年 6 月

第 2 版致谢

自从第 1 版面世以后，我们荣幸地收到了许多读者的反馈意见和修改建议。然而，向每一位曾经帮助过我们的读者致谢，这对我们来讲心有余而力不足，但我们仍然想道出其中的一些名字以表谢意。我们特别要感谢所有使用本书讲授和学习这门课的老师和学生们，正是通过他们，才使我们能从不同视角重新审视所选择的内容。

我们特别要感谢 Andrew Barron、Alon Orlitsky、T. S. Han、Raymond Yeung、Nam Phamdo、Franz Willems 和 Marty Cohn，他们给出了许多宝贵的评论和建议。这些年来，斯坦福大学的学生为本书的修改给了我们许多的思想和启发，他们是 George Gemelos、Navid Hassanpour、Young-Han Kim、Charles Mathis、Styrmir Sigurjonsson、Jon Yard、Michael Baer、Mung Chiang、Suhas Diggavi、Elza Erkip、Paul Fahn、Garud Iyengar、David Julian、Yiannis Kontoyiannis、Amos Lapidoth、Erik Ordentlich、Sandeep Pombra、Jim Roche、Arak Sutivong、Joshua Sweetkind-Singer 和 Assaf Zeevi。在第 2 版准备期间，Denise Murphy 给我们提供了许多支持和帮助。

Joy Thomas 要感谢在 IBM 和 Stratify 的同事的支持和提出的有价值的意见和建议。特别感谢 Peter Franaszek、C. S. Chang、Randy Nelson、Ramesh Gopinath、Pandurang Nayak、John Lamping、Vineet Gupta 和 Ramana Venkata。特别是与 Brandon Roy 长达几个小时的讨论有助于将书中的某些论述写得更加精练简洁。最为重要地，Joy 感谢妻子 Priya 的全力支持，如果没有她的支持和鼓励，第 2 版的完成是不可能的。

Tom Cover 感谢他的学生们和妻子 Karen 给予的帮助。

第 1 版致谢

我们真诚感谢所有参与完成本书的人们，尤其是 Aaron Wyner、Toby Berger、Masoud Salehi、Alon Orlitsky、Jim Mazo 和 Andrew Barron 对本书的各版草稿给予了细致评述，这对我们最终内容的取舍起了指导性的作用。还要感谢我们手写稿的第一位读者 Bob Gallager，以及他对出版本书的支持。感谢 Aaron Wyner 和 Ziv 赠送了关于 Lempel-Ziv 算法收敛性的新证明。还要感谢 Normam Abramson、Ed van der Meulen、Jack Salz 和 Raymond Yeung 给予我们很多修订建议。

一些重要的访问学者和专家同事也给予了很多帮助，他们是 Amir Dembo、Paul Algoet、Hirosuke Yamamoto、Ben Kawabata、M. Shimizu 和 Yoichiro Watanabe。John Gill 在教学中使用了本书，从他的建议中我们获益匪浅。当我们计划编写一本面向广泛读者的信息论专著时，Abbas EI Gamal 在几年前就已经开始帮助写作此书，其贡献是不可估量的。还要感谢在本书成形阶段研究信息论方向的博士生们，他们是 Laura Ekroot、Will Equitz、Don Kimber、Mitchell Trott、Andrew Nobel、Jim Roche、Erik Ordentlich、Elza Erkip 和 Vittorio Castelli。Mitchell Oslick、Chien-Wen Tseng 和 Michael Morrell 是其中提出问题和建议最为主动的学生。Marc Goldberg 和 Anil Kaul 帮助我们制作了其中的一些图形。最后，我们还要感谢 Kirsten Goodell 和 Kathy Adams 在原稿准备过程中提供的支持和帮助。

Joy Thomas 也要感谢 Peter Franaszek、Steve Lavenberg、Fred Jelinek、David Nahamoo 和 Lalit Bahl 在完成本书的最后阶段给予的鼓励和支持。

目　录

第 1 章　绪论与概览

信息论解答了通信理论中的两个基本问题：临界数据压缩的值(答案：熵 H)和临界通信传输速率的值(答案：信道容量 C)。因此，有人认为信息论是通信理论的一个组成部分，但我们将竭力阐明信息论远不止于此。其实，信息论在统计物理(热力学)、计算机科学(科尔莫戈罗夫(Kolmogorov)复杂度或算法复杂度)、统计推断(奥克姆剃刀(Occam Razor)："最简洁的解释最佳")以及概率和统计(关于最优化假设检验与估计的误差指数)等学科中都具有奠基性的贡献。

本章是"开场白"，通过介绍信息论及其关联的思想的来龙去脉，提纲挈领地给出该书的整体布局。所涉及的术语和内容，将从第 2 章开始逐步给予详细叙述和讨论。图 1-1 揭示了信息论与其他学科之间的关系。如图中所示，信息论与物理学(统计力学)、数学(概率论)、电子工程(通信理论)以及计算机科学(算法复杂度)都有交叉。我们接下来对这些交叉的领域作更详细的说明。

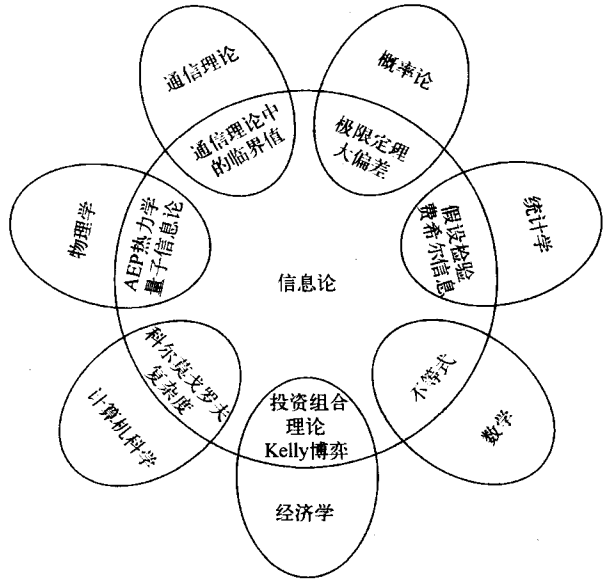


图 1-1　信息论与其他学科的关系

电子工程(通信理论)。20 世纪 40 年代早期，人们普遍认为，以正速率发送信息，而忽略误差概率是不可能做到的。然而，香农(Shannon)证明了只要通信速率低于信道容量，总可以使误差概率接近于零，这个结论震惊了通信理论界。信道容量可以根据信道的噪声特征简单地计算出来。香农还进一步讨论了诸如音乐和语音等随机信号都有一个不可再降低的复杂度，当低于该值时，信号就不可能被压缩。遵从热力学的习惯，他将这个临界复杂度命名为熵，并且讨论了当信源的熵小于信道容量时，可以实现渐近无误差通信。

如果将所有可能的通信方案看成一个集合，那么今天的信息论描绘了这个集合的两个临界值，如图 1-2 所示。数据压缩达到最低程度的方案对应的是该集合的左临界值 $I(X;\hat{X})$。所有数据压缩方案所需的描述速率不得低于该临界值。右临界值 $I(X;Y)$ 所对应方案的数据传输速率最大，临界值 $I(X;Y)$就是信道容量。因此，所有调制方案和数据压缩方案都必须介于这两个临界值之间。

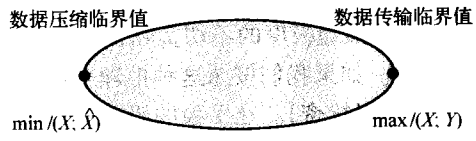


图 1-2　通信理论的信息论临界点

信息论也提供能够达到这些临界值的通信方案。从理论上讲，最佳通信方案固然很好，但从计算的角度看，它们往往是不切实际的。惟一的原因是，只有使用简单的调制与解调方案时才具

有计算可行性，而香农信道容量定理的证明过程中所提出的随机编码和最邻近译码规则却不然。集成电路与编码设计方面的进展使得我们能获得香农理论所蕴涵的一些硕果。随着 Turbo 码的诞生，最终实现了计算的实用性。比如，纠错码在光盘和 DVD 中的应用就是信息论的一个绝好实例。

信息论中关于通信方面的近期研究集中在网络信息论：存在干扰和噪声的情况下，大量发送器到大量接收器之间的通信同步率理论。目前，多个发送器与多个接收器之间的一些速率协定还无法预料，已有协定也有待于从数学上得到一定程度的简化。因而，一套统一的理论尚待发掘。

计算机科学(科尔莫戈罗夫复杂度)。科尔莫戈罗夫、Chaitin 和 Solomonoff 指出，一组数据串的复杂度可以定义为计算该数据串所需的最短二进制程序的长度。因此，复杂度就是最小描述长度。利用这种方式定义的复杂度是通用的，即与具体的计算机无关，因此该定义具有相当重要的意义。科尔莫戈罗夫复杂度的定义为描述复杂度的理论奠定了基础。更令人愉快的是，如果序列服从熵为 H 的分布，那么该序列的科尔莫戈罗夫复杂度 K 近似等于香农熵 H。所以信息论与科尔莫戈罗夫复杂度二者有着非常紧密的联系。实际上，科尔莫戈罗夫复杂度比香农熵更为基础。它不仅是数据压缩的临界值，而且也可以导出逻辑上一致的推理过程。

算法复杂度与计算复杂度二者之间存在着微妙的互补关系。计算复杂度(也就是时间复杂度)与科尔莫戈罗夫复杂度(也就是程序长度或描述复杂度)可以看成是对应于程序运行时间与程
3 序长度的两条轴。科尔莫戈罗夫复杂度是沿第二条轴的最小化问题，而计算复杂度是沿第一条轴的最小化问题。沿两条轴同时进行最小化的工作几乎没有。

物理学(热力学)。熵与热力学第二定律都诞生于统计力学。对于孤立系统，熵永远增加。热力学第二定律的贡献之一是促使我们抛弃了存在永动机的幻想。我们将在第 4 章中简述该定律。

数学(概率论和统计学)。信息论中的基本量——熵、相对熵与互信息，定义成概率分布的泛函数。它们中的任何一个量都能刻画随机变量长序列的行为特征，使得我们能够估计稀有事件的概率(大偏差理论)，并且在假设检验中找到最佳的误差指数。

科学的哲学观(奥克姆剃刀)。奥克姆居士威廉说过“因不宜超出果之所需。”其意思是“最简单的解释是最佳的”。Solomonoff 和 Chaitin 很有说服力地讨论了这样的推理：谁能获得适合处理数据的所有程序的加权组合，并能观察到下一步的输出值，谁就能得到万能的预测程序。如果是这样，这个推理可以用来解决许多使用统计方法不能处理的问题。例如，这样的程序能够最终预测圆周率 π 的小数点后面遥远位置上的数值。将这个程序应用到硬币的正面出现概率为 0.7 的硬币抛掷问题中，也能得出推断。不仅如此，如果应用到股票市场，程序能从根本上抓住市场的“规律”并做出最优化的推断。这样的程序能够从理论上保证推出物理学中的牛顿三大定律。当然，这样的推理极度的不切实际，因为清除所有不适合生成现有数据的程序需要花费的时间是不可接受的。如果我们按照这种推理来预测明天将要发生的事情，那么需要花一百年的时间。

经济学(投资)。在平稳的股票市场中重复投资会使财富以指数增长。财富的增长率与股票市场的熵率有对偶关系。股票市场中的优化投资理论与信息论的相似性是非常显著的。我们将通过探索这种对偶性来丰富投资理论。

计算与通信。当将一些较小型的计算机组装成较大型的计算机时，会受到计算和通信的双重限制。计算受制于通信速度，而通信又受制于计算速度，它们相互影响、相互制约。因此，通
4 信理论中所有以信息论为基础所开发的成果，都会对计算理论造成直接的影响。

本书概览

信息论最初所处理的问题是数据压缩与传输领域中的问题，其处理方法利用了熵和互信息等基本量，它们是通信过程的概率分布的函数。先给出一些定义，这会有助于开始讨论，在第2章中我们会重述这些定义。

如果随机变量 X 的概率密度函数为 $p(x)$，那么 X 的熵定义为

$$H(X) = -\sum_x p(x)\log_2 p(x) \tag{1-1}$$

使用以2为底的对数函数，熵的量纲为比特。熵可以看作是随机变量的平均不确定度的度量。在平均意义下，它是为了描述该随机变量所需的比特数。

例 1.1.1 考虑一个服从均匀分布且有32种可能结果的随机变量。为确定一个结果，需要一个能够容纳32个不同值的标识。因此，用5比特的字符串足以描述这些标识。

该随机变量的熵为

$$H(X) = -\sum_{i=1}^{32} p(i)\log p(i) = -\sum_{i=1}^{32} \frac{1}{32}\log\frac{1}{32} = \log 32 = 5 \text{ 比特} \tag{1-2}$$

这个值恰好等于描述该随机变量 X 所需要的比特数。在此情形中，所有结果都有相同长度的表示。

下面考虑一个非均匀分布的例子。

例 1.1.2 假定有8匹马参加的一场赛马比赛。设8匹马的获胜概率分布为 $\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{16},\frac{1}{64},\frac{1}{64},\frac{1}{64},\frac{1}{64}\right)$。我们可以计算出该场赛马的熵为

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{16}\log\frac{1}{16} - 4\,\frac{1}{64}\log\frac{1}{64} = 2 \text{ 比特} \tag{1-3}$$

5

假定我们要把哪匹马会获胜的消息发送出去，其中一个策略是发送胜出马的编号。这样，对任何一匹马，描述需要3比特。但由于获胜的概率不是均等的，因此，明智的方法是对获胜可能性较大的马使用较短的描述，而对获胜可能性较小的马使用较长的描述。这样做，我们会获得一个更短的平均描述长度。例如，使用以下的一组二元字符串来表示8匹马：0，10，110，1110，111100，111101，111110，111111。此时，平均描述长度为2比特，比使用等长编码时所用的3比特小。注意，此时的平均描述长度2正好等于熵。在第5章中，我们将证明任何随机变量的熵必为表示这个随机变量所需要的平均比特数的一个下界。另外，在“20问题”的游戏中，将所需问题的数目看成随机变量，那么它的熵也是所需问题数目的平均值的下界。我们也将说明如何构造一些表示法使其平均长度与熵相比较不超过1比特。

信息论中的熵与统计力学中的熵概念有着紧密的联系。如果抽出一个包含 n 个独立同分布(i.i.d.)的随机变量的序列，我们将证明该序列是“典型”序列的概率大约为 $2^{-nH(X)}$，而且大约只能抽出 $2^{nH(X)}$ 个典型序列。这个性质(著名的渐近均分性，AEP)是信息论中许多证明的基础。随后我们将介绍利用熵自然地解答的一些问题(例如，生成一个随机变量所需的抛掷均匀硬币的次数)。

随机变量的描述复杂度的概念可以推广到定义单个字符串的描述复杂度。二元字符串的科尔莫戈罗夫复杂度定义为输出该字符串所需的最短计算机程序的长度。如果字符串确实是随机的，那么其科尔莫戈罗夫复杂度接近于它的熵。从统计推断和建模问题的角度考虑，科尔莫戈罗夫复杂度是一个自然的框架，使我们对奥克姆剃刀“最简洁的解释最佳”有更加透彻的理解。我们将在第14章中叙述科尔莫戈罗夫复杂度的一些简单性质。