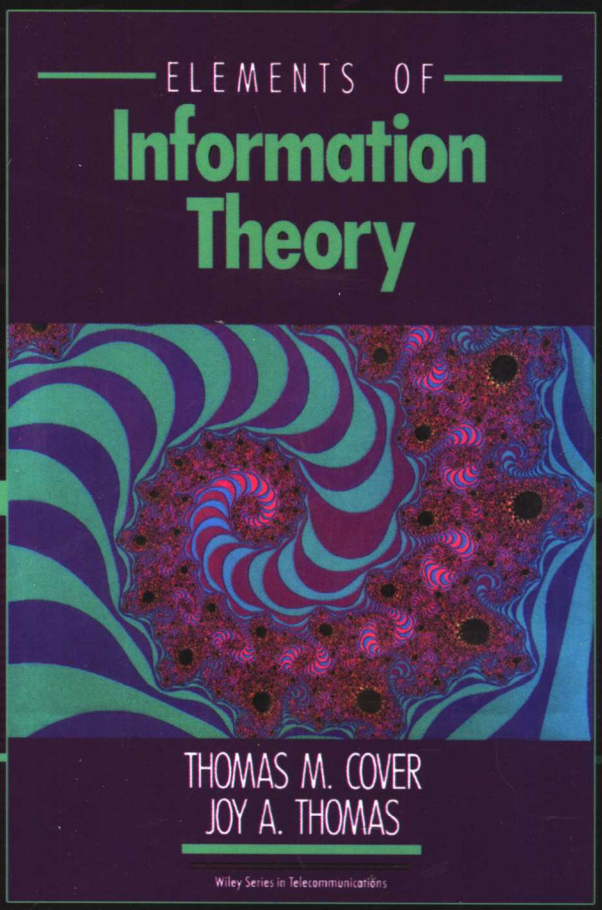
HZ BOOKS
电子与电气工程丛书

信息论基础

Elements of Information Theory

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



机械工业出版社
China Machine Press

电子与电气工程丛书

信息论基础

Elements of Information Theory

(美) Thomas M. Cover
Joy A. Thomas 著

阮吉寿 张 华 译
沈世镒 审校

机械工业出版社
China Machine Press

本书全面系统地介绍了香农信息论的基本理论以及多类应用问题，其中包括了作者的许多研究成果。本书阐述了熵、相对熵和互信息之间的基本代数关系，论述了渐近均分性（AEP）、随机过程和数据压缩的熵率、Kolmogorov复杂度、信道容量定理、微分熵以及网络信息理论等内容，并采用“使用不等式串、中间不加任何文字、最后直接加以解释”的创新表述方式，使读者学习了一定数量的证明后，在没有任何解释的情况下就能理解其中的大部分步骤，并给予必要的解释。

本书适合作为通信理论、计算机科学和数学等专业学生学习信息论的教材。章后提供的习题便于老师的教学，以及增强学生对信息论的理解。

本书版权登记号：图字：01-2003-7224

图书在版编目（CIP）数据

信息论基础 /（美）科弗（Cover, T. M.），（美）托马斯（Thomas, J. A.）著；阮吉寿等译. -北京：机械工业出版社，2005. 5

（电子与电气工程丛书）

书名原文：Elements of Information Theory

ISBN 7-111-16245-5

Ⅰ. 信…　Ⅱ. ① 科…　② 托…　③ 阮…　Ⅲ. 信息论-高等学校-教材　Ⅳ. G201

中国版本图书馆CIP数据核字（2005）第018954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：傅志红

北京诚信伟业印刷有限公司印刷 · 新华书店北京发行所发行

2005年5月第 1 版第 1 次印刷

787mm × 1092mm 1/16 · 28印张

印数：0 001-4 000册

定价：56.00元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换

本社购书热线：（010）68326294

译 者 序

本书是Thomas M. Cover 的重要著作，全面系统地介绍了香农信息论的基本理论以及多类应用问题，包括了作者自己的许多研究成果，其内容要点与特点概述如下。

本书的第 2、3、4、9章是关于信息度量问题的讨论，除了介绍香农熵的引入与性质外，还介绍了随机序列的熵率与AEP（渐近均分性）理论。随机序列的熵率与AEP理论是经典的信息论问题，在20世纪60年代的早期信息论著作中有较多的讨论，但在近期的信息论著作中却较少见，这些问题在随机过程的信息处理中有用。第9章则是关于连续型随机变量微分熵的讨论。

通信系统编码理论是信息论的主体部分，其中包括无失真信源编码问题（又称数据压缩理论）、信道编码问题、有失真信源编码问题（又称率失真理论）与网络信息论（又称多用户信息论）四部分内容，分别在第 5、6、8、10、13、14 章中讨论，其中信源与信道编码理论在其他教材或著作中虽有讨论，但本书的内容更为丰富，如有反馈高斯信道编码定理等在其他教材或著作中较为少见。有失真信源编码理论与网络信息论在国内的有关教材中虽有提及，但篇幅不大。本书对这两部分内容都做了系统论述，因此本书是全面介绍信息编码理论的著作。这部分内容的另一个重要特点是对编码理论的描述与证明采用了典型序列的方法来叙述与讨论，对主要的信道与有失真信源编码定理采用随机码的证明方法，这是20世纪 70 年代后信息论研究的主流方法。它们对信息论的主要理论实现了完全严格的数学描述与证明。阅读这些定理的证明有一定的难度，但通过这些定理证明的阅读与理解可以加深对信息编码理论的理解。

本书的第7、11、12、15章是信息论的应用部分，其中第 7 章介绍 Kolmogorov 复杂度问题，这是信息论与计算复杂度相结合的一个学科领域，在有的文献中又称之为算法信息论。该章的主要结论是证明了Kolmogorov复杂度与香农熵的等价性定理，这就得到了信息论与计算机科学的本质联系，也说明了香农熵的广泛意义。第11章和第12章是讨论信息论与统计的联系，其中最大熵原理、谱估计理论与Fisher信息矩阵理论都是信息与统计密切相关的内容，利用信息论的相对熵与典型序列理论对多种统计问题可得到它们更精确的误差估计，证明了多个统计误差可实现指数下降的结论。第15章是利用信息论的方法研究投资组合决策问题，其中许多结果由作者得到。

本书内容十分丰富，涉及信息、统计、金融、逻辑与计算机科学等多个领域，并有大量的例题与背景说明，这些例题与背景涉及信息处理与信息世界中的许多问题，内容十分丰富。在信息编码理论的论述中，所讨论的问题与处理的方法是信息论中最典型的问题与方法，学习与研究这些问题与方法可了解到国际信息论学术界在信息论研究中所使用的语言与方法。本书对所论述的主要定理、重要性质与计算公式都有总结与汇总，并在第1章和第16章中给以

介绍和说明。

国际上许多院校把本书作为信息论的经典教材。在电子、通信、计算机与数学等专业领域中，信息论方向的研究生都采用本书为教材或主要参考书，部分本科专业也讲授其中的部分内容。因此本书也适用于国内有关专业作为研究生或本科生的教材或参考书。学好本书就可为信息论打下良好的基础。有关应用部分的内容各章并不关联，因此在学习或讲授时可以选读或选讲。

20世纪末与21世纪初是信息科学与 IT 产业突飞猛进的时代，这些进展无疑得益于信息论基础理论的建立。自本书英文版出版后信息论又有许多重要发展，如调制解调码的理论与应用、数据压缩编码理论的实现以及Turbo码与LDPC码的出现与发展，这些都是信息论理论与应用的重大突破，在信息技术与信息产业发展中发挥了重大的作用。因此，对从事信息论研究与教学的工作者或研究生来讲，还应关心与了解这些发展，以适应信息科学与技术迅速发展的时代。

本书由阮吉寿、张华、沈世镒共同翻译，其中阮吉寿翻译了前言、第1、6、7、8、9、10、11、15章；张华翻译了目录、第2、3、4、5、16章、索引，并对此进行了校对；沈世镒翻译了第12、13、14章。本书由阮吉寿统一执笔，沈世镒对全书把关。同时本书在初稿的翻译过程中，目前在美国攻读博士学位的尚越和佘涛同学分别在第7、8、9、10章和第12、13、14章做了有益的工作，这里向他们表示感谢。

由于译者水平和学识有限，难以在本书涉及的方向均具有科班水平，加上时间仓促，翻译中难免有错误和不妥之处，敬请广大读者批评指正。

译　者

2005年1月

前　言

本书试图编写成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（据说最初是在餐后的幸运蛋卷中所夹的纸条上发现的），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，不仅能够使信息论的主题显得简明，而且会在处理新问题时提供重要的直觉。

本书来自已经使用了十多年的信息论讲义，原讲义是面向信息专业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，熵和互信息这样的概念是为了回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，我们后面还会提到，互信息相当于已知边信息条件下财富双倍率的增长。第二，回答信息论理论问题的答案具有自然代数的结构。例如熵具有链式法则，因而相关联的相对熵和互信息也有相应的链式法则。因为有了它们，才使得数据压缩和通信工程中的问题得到深入的解释。我们都有这样的感受，当某人研究某个问题时，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不过的例子。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，我们首先定义了熵、相对熵和互信息，然后研究它们之间的关系，再对这些关系做一些解释，由此揭示如何融会贯通地使用各式各样的方法去解决实际问题。同理，我们顺便探讨热力学第二定律的含义。熵总是增加的吗？答案既肯定也否定。这种或然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，教师往往会对教材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"，而且也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。

本书中加入大量新素材，例如，关于信息论与博弈的关系；热力学第二定律的普遍性的讨论，包括它在马尔可夫链中，在信道容量定理的联合典型性的证明过程中，在赫夫曼码的竞争最优性中，以及在关于最大熵谱密度估计的伯格（Burg）定理的证明过程中的应用。Kolmogorov复杂度这一章也是本书的独到之处。而将Fisher信息、互信息与Brunn-Minkowski不等式和熵幂不等式联系在一起，也是我们引以为自豪之处。令我们感到惊讶的是，关于行

列式不等式的许多经典结论，当利用信息论知识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相关性。虽然香农创立信息论时是受到通信理论中的问题启发的，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。

我们之所以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，是因为已经明显不可能从这些学科中获得难以理解的关于信息的概念。

由于本书中绝大多数结论是以定理和证明的形式给出的，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。

使用不等式链、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。当读者学习我们所给的证明过程达到一定数量时，希望他或她在没有任何解释下就能理解其中的大部分步骤，并得到所需的解释。这些不等式链好比是模拟测试题，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，以致于我们轻视了写作技巧中的某条重要原则。由于没有空话，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、自然风格的信息论。

本书广泛使用弱典型序列的方法，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“几乎一切事情都差不多是等可能的”。

第2章是本书正式内容的开始，阐述了熵、相对熵和互信息之间的基本代数关系，同时对热力学第二定律和充分统计量也进行了讨论。渐近均分性是第3章重中之重的内容，这也导致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，并将数据压缩的对偶性和财富的增长率推向深入。

第7章探讨Kolmogorov复杂度的基本思想，它是对信息论进行理性思考的基础。我们的目标是寻找最普遍、最简短的描述，而不是在平均意义下的次佳描述。的确存在这样的普遍性概念可以用来描述目标的复杂度。该章也论述了神奇数Ω，它可揭示数学上的不少奥秘，是图灵机（Turing machine）停止运转的概率的推广。

第8章论述信道容量定理，这是信息论的基本定理。第9章叙述微分熵的必备知识，它们是将早期容量定理推广到连续噪声信道的基础。重要的高斯信道容量问题在第10章中论述。

第12章阐述信息论和统计学之间的关系。早在20世纪50年代，库尔贝克（Kullback）就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，因而将其放置在本书较后的第13章。

网络信息理论是个大的主题，安排在第14章，主要研究的是在噪声和干扰存在的情形下同时可达的信息流。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15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16章讲述信息论中的有关不等式，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，并且再加上一些关于随机抽取子集熵率的有趣的新不等式。集合求和

的容量的Brunn-Minkowski不等式、独立随机变量之和的有效方差的熵幂不等式以及Fisher信息不等式之间的美妙关系也在这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成绩优秀，大致为本科高年级或研究生水平。尽管如此，我们还是努力避免使用测度论，因为了解它仅仅对第15章中遍历过程的AEP证明起到简化作用。这符合我们的观点，那就是信息论基础与信息论技巧不同，后者才需要将所有推广都写进去。

本书每一章均以总结各章关键结论的方式结束。所提出的要点以方程形式表述，没有写出其限制条件。总结要点之后，我们列出一系列习题，接着以简短的历史回顾方式讲述了主要结论的来龙去脉。本书末尾的参考文献包括该领域中的许多关键性论文，以及参阅其他书目和相关主题的概述性文章的索引。

本书的主体是第2、3、4、5、8、9、10、12、13、14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7章的Kolmogorov复杂度是深入理解信息论的必备知识，余下的几章包括博弈和不等式，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。

致谢

我们真诚感谢所有参与完成本书的人们，尤其是Toby Berger、Masoud Salehi、Alon Orlitsky、Jim Mazo和Andrew Barron对本书的各版草稿给予时细致评述，这对我们最终的内容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，而且很高兴在本书中引用了他的12个问题。也感谢Aaron Wyner和Ziv赠送了关于Lempel-Ziv算法收敛性的新证明。还要感谢Norman Abramson、Ed van der Meulen、Jack Salz和Raymond Yeung给予我们许多建议。

一些重要的访问学者和专家同事也给予了许多帮助，他们有Amir Dembo、Paul Algoet、Hirosuke Yamamoto、Ben Kawabata、Makoto 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 El 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。

Tom Cover

Joy Thomas

1991年6月于Palo Alto

目　录

第1章 绪论与概览

本章是开场白，通过介绍信息论及其相关思想的来龙去脉，提纲挈领地给出本书的整体布局。所涉及的术语和内容，将从第2章开始给予详细的叙述和讨论。

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非如此狭窄。其实，信息论在统计物理（热力学）、计算机科学（Kolmogorov复杂度或算法复杂度）、统计推断（奥卡姆剃刀："最简单的解释是最佳的"）以及概率统计（关于最优化假设检验的错误概率以及估计的误差概率）等学科中都具有奠基性的贡献。

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）都有交叉部分。我们接下来对这些交叉的领域做更详细的说明：

电气工程（通信理论）。20世纪40年代早期，人们普遍认为：在任何通信信道中，信息传输速率越高，出错的概率越大。然而香农从理论上证明了只要通信速率低于信道容量，这种观点是错误的，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。

当今，信息论关注的是所有可能的通信方案集合的两个临界值，如图1-2所示。数据压缩达到最低程度的方案对应的是通信方案集合的左临界值 $I(X;\hat{X})$。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$I(X;Y)$所对应的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。

信息论也提供了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美妙，但从计算的角度看，它们往往是不切实际的。唯一的原因是，我们只有使用简单的调制与解调方案时才具有计算可行性，而在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提出的一些结论。比如，纠错码在光盘中的使用就是信息论思想的一个很好应用实例。

信息论中关于通信方面的最新研究集中在网络信息论：在通信网络中，从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，已有协定也还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（Kolmogorov复杂度）。Kolmogorov、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。Kolmogorov复杂度的定义为复杂度的理论奠定了基础。更令人惊奇的是，如果序列服从熵为H的分布，那么该序列的Kolmogorov复杂度K近似等于H。所以信息论与Kolmogorov复杂度二者有着非常紧密的联系。实际上，我们认为Kolmogorov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。

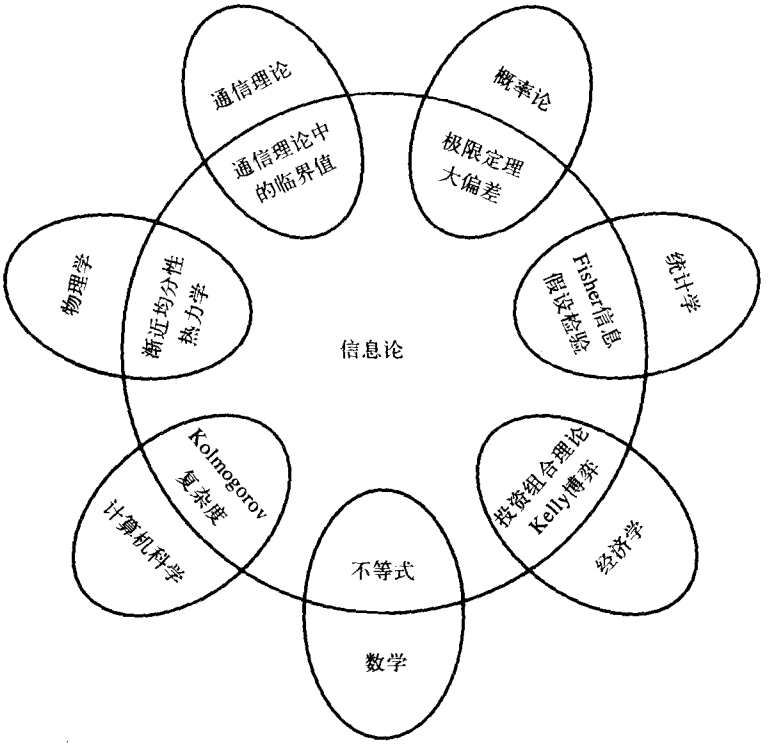


图1-1 信息论与其他学科的关系

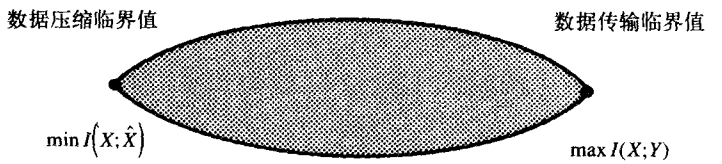


图1-2 通信理论的信息论临界值

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Kolmogorov复杂度（也就是程序长度或者描述复杂度）可以看成是对应于程序运行时间与程序长度的两条轴。Kolmogorov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。

3 **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就是促使我们抛弃了存在永动机的幻想。我们将在第2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，被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），并且能够在假设检验中找到最佳的误差指数。

科学的哲学观（奥卡姆剃刀）。奥卡姆居士威廉曾说过“因不宜超出果之所需”。大概意思是“最简单的解释是最佳的”。Solomonoff以及Chaitin有说服力地讨论了这样的推理：谁能得到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；将程序应用到硬币的正面出现概率为0.7的抛硬币问题中，它也能得出推断；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”并做出最优化的外推；特别是，这样的程序能够从理论上保证我们推导出物理学中的牛顿三大定律。当然，这样的推理是极度不切实际的，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。打个比方，如果我们按照这种推理来预测明天将要发生的事情，那么首先需要花上一百年时间来排除那些不合适的方法。

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（称为双倍率）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，通信理论中所有以信息论为基础开发的成果，都会对计算理论造成直接的影响。

4

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，它们是通信随机过程的概率分布函数的基础。先给出一些定义，这会有助于将要开始的讨论，而在第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：

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

使用以2为底的对数函数，此时熵的量纲为比特。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此随机变量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(1/2, 1/4, 1/8 1/16, 1/64, 1/64, 1/64, 1/64)。可以计算出该场赛马比赛的熵为:

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

■

假定要发送一则消息给某人，告诉他哪匹马会赢得本场比赛，一个等价的策略是发送胜出马的编号。这样，对任何一匹马，描述只需要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
5 长的描述。这样做，会获得一个更短的平均描述长度。例如使用以下的一组二进制字符串来表示8匹马: 0, 10, 110, 1110, 111100, 111101, 111110, 111111。此时，平均描述长度仅为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下界。另外，在“二十问题”的游戏中，将所需要的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。

信息论中熵的概念与统计力学中熵的概念有着紧密的联系。如果我们抽出一个包含n个独立同分布（i.i.d.）的随机变量的样本序列，那么将证明该样本序列是“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（就是著名的渐近均分性，简记为AEP）是信息论中许多证明的基础。随后我们将介绍用熵来自然解答的一些问题（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Kolmogorov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。我们将会推出如果字符串确实是随机的，那么该字符串的Kolmogorov复杂度接近于它的熵。对于考虑统计推断和建模问题，Kolmogorov复杂度是一个自然的框架。它也使我们对奥卡姆剃刀“最简单的解释是最佳的”有更加透彻的理解。我们将在第7章中叙述Kolmogorov复杂度的一些简单性质。

单个随机变量的熵是该随机变量的不确定度。我们还可以定义涉及两个随机变量的条件熵，即一个随机变量在给定另外一个随机变量的条件下的熵。由于一个随机变量的不确定度可以因为另一个随机变量而缩减，不确定度的缩减量称为互信息。具体地讲，设X和Y是两个随机变量，那么这个缩减量为:

$$I(X;Y) = H(X) - H(X|Y) = \sum_{x,y} p(x, y)\log\frac{p(x, y)}{p(x)p(y)} \tag{1-4}$$

互信息$I(X; Y)$是两个随机变量相互之间的独立程度的度量，它关于两个随机变量X和Y是对称的，并且永远是非负值。

通信信道是一个随机系统，系统的输出信号按概率依赖于输入信号。该系统特征由一个转移概率矩阵决定，该矩阵决定了在给定输入情况下输出的条件概率分布。对于输入信号为X
6 和输出信号为Y的通号信道，我们定义它的信道容量C为:

$$C = \max_{p(x)} I(X; Y) \tag{1-5}$$

后面我们将证明信道容量是信息传输的最大速率，若以此速率发送信息，我们不仅可以使用该信道，而且可以在接收端以极低的误差概率恢复出该信息。下面用一些例子来说明这点。

例1.1.3（无噪声二元信道） 对于无噪声二元信道，二元输入信号在输出端精确地恢复出来，如图1-3所示。在此信道中，任何传输的信号都会毫无误差地被接收。因此，在每一次传输过程中，我们就可以将1比特的信息可靠地发送给接收端，从而信道容量为1比特。当然，我们也可以通过计算得出信道容量为$C = \max I(X; Y) = 1$比特。

例1.1.4（有噪声四字符信道） 观察如图1-4所示的信道。在该信道中，传输每个字符时，能够正确地收到该字符的概率为1/2，而被误判为它的下一个字符的概率也为1/2。如果将4个输入字符全部考虑进去，那么在接收端，仅凭输出结果根本不能确切地判定原来传输的是哪个字符。但是，如果仅使用2个输入字符（比如1和3），我们立即可以根据输出结果知道传输的是哪个输入字符。这种信道的作用相当于前面例子中的无噪声信道，因此，我们在该信道上每传输一次可以毫无误差地发送1比特信息。此时，我们可以计算信道容量$C = \max I(X; Y)$，它也等于1比特/传输，这符合上述分析。

图1-3 无噪声二元信道

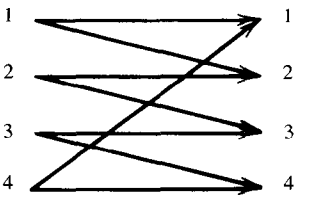


图1-4 有噪声信道

一般地，通信信道的结构不会像我们所举的例子这样简单，所以不是总能准确无误地识别出所发送的信息的某个子集。但是，如果我们考虑一系列传输，那么任何信道看起来都会像图1-4所示例子那样，并且均可以识别出输入序列集（码字集）的一个子集，其中的码字是按如下方式来传输信息的：对应于每个码字的所有可能输出序列构成的集合近似不相交。此时我们可以观察输出序列，并且能够以极低的误差概率识别出相应的输入码字。

7

例1.1.5（二元对称信道） 二元对称信道是有噪声通信系统的一个基本例子，这种信道的示意图如图1-5所示。

此信道有2个输入字符，输出字符与输入字符具有相同的概率为$1-p$。另一方面，0被接收为1的概率为p，1被接收为0的概率也是p。

此时，通过计算可以得到信道容量为$C = 1 + p \log p + (1-p) \log (1-p)$比特/传输。然而，如何达到该信道容量已经不再明显了。如果我们多次利用该信道，那么该信道就会开始类似于前面例子中所示的有噪声四字符信道，从而能以C比特/传输的速率发送信息而几乎不发生误差。 ■

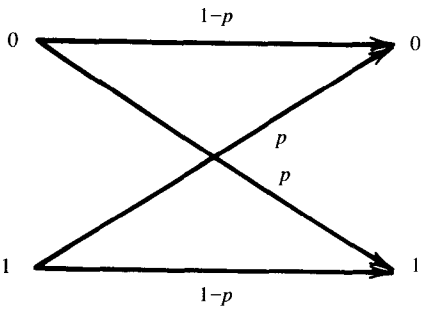


图1-5 二元对称信道

信道上的信息通信速率的临界值由信道容量决定。信道编码定理证明了该值可利用长的分组长度编码达到。在实际的通信系统中，由于能够使用的编码的复杂度是有限制的，因此一般无法达到该信道容量。

互信息实际上是相对熵$D(p\|q)$的特殊情形，相对熵是两个概率分布函数p和q之间的“距离”的度量。它的定义为：

$$D(p\|q)=\sum_x p(x)\log\frac{p(x)}{q(x)} \tag{1-6}$$

尽管相对熵不是真正的度量，但它有着度量的某些性质。例如，相对熵总是非负的，且它为0
8 的充分必要条件为$p=q$。在两个分布p和q之间的假设检验中，相对熵就是误差概率的指数。它也可以用来定义概率分布的几何结构，使得我们能够解释大偏差理论中的许多结论。

信息论和股票市场的投资理论有许多相似之处。可将股票市场定义为一个随机向量$\boldsymbol{X}$，其分量是非负的数值，等于某只股票当天的收盘价与当天的开盘价的比值。若股票市场X的分布为$F(\boldsymbol{x})$，那么我们定义双倍率W为：

$$W=\max_{\mathbf{b}:b_i\geq 0,\ \Sigma b_i=1}\int\log\boldsymbol{b}^t\boldsymbol{x}\mathrm{d}F(\boldsymbol{x}) \tag{1-7}$$

双倍率是财富增长的最大渐近指数。双倍率有一系列性质与熵的对应性质类似。我们将在第15章中探讨这些性质。

H、I、C、D、K和W这些量自然出现在以下领域中：

- **数据压缩**。随机变量的熵H是最短描述此随机变量所需的平均长度的下界。我们可以构造一个平均长度不超出熵1比特的描述。

  如果放宽完全恢复信源信息的限制，那么我们此时可以问：如果不计较失真D的话，需要多大的码率来描述信源？另外，需要多大的信道容量，才足够能让信源信息在信道上传输，并且在失真不超过D的情况下重构信源？这是率失真理论的研究课题。

  当我们试图对非随机性目标的最短描述的概念进行严格定义时，Kolmogorov复杂度K的概念就应运而生了。稍后，我们将证明Kolmogorov复杂度的通用性以及符合最短描述理论的许多直觉要求。

- **数据传输**。我们所要考虑的信息传输问题是希望接收器能够以小误差概率恢复消息。从本质上讲，我们希望找到的码字（信道的输入字符序列）彼此之间离得很远，目的是当它们在信道中被噪声污染后依然能够区分开来。这等价于高维空间中的填球问题。对任何码字集，要计算出接收器可能出错（换言之，将传送过来的码字做了错误的判断）的
9 概率是可以办到的。然而，在绝大多数情形下，这种计算很繁琐。

  假定使用随机生成编码方案，香农证明了如果码率不超过信道容量C，那么我们能够以任意小的误差概率发送信息。随机码的思想非同寻常，它为简化难解问题打下了基础。香农在该证明过程中所使用的关键思想之一是所谓的典型序列概念。

- **网络信息理论**。前面所提到的每一个主题涉及的均是单一信源或单一信道。但如果我们希望同时压缩大量的信源信息，然后将压缩后的描述进入信源联合重构的环节，情况将