高等学校教材

信息论与编码

第二版

仇佩亮　张朝阳　谢　磊　余官定　编著

高等教育出版社
HIGHER EDUCATION PRESS

高等学校教材

信息论与编码

Xinxilun yu Bianma

第二版

仇佩亮　张朝阳　谢　磊　余官定　编著

高等教育出版社·北京
HIGHER EDUCATION PRESS　BEIJING

内容提要

信息论与编码是研究信息传输和信息处理过程中一般规律和具体实现的一门应用科学，是现代信息科学和工程技术的基础理论。本书在吸取了国内外经典教材的优点，结合作者们长期教学和科研实践经验的基础上编写而成。本书深入浅出，既保持理论的完整性、系统性和严谨性，又概念清晰、易读好懂，同时还介绍了信息论与编码技术的新发展。本书主要介绍Shannon信息论和相关的编码技术。内容包括如下11章：绪论，熵和互信息，离散无记忆信源的无损编码，信道、信道容量及信道编码定理，率失真理论和保真度准则下的信源编码，受限系统和受限系统编码，线性分组纠错编码，循环码，卷积码，先进的信道编码技术，多用户信息论。

本书适合作为高等院校电子信息类专业的高年级本科生和研究生教材，对于从事信息科学和技术领域工作和研究的人员也极具参考价值。

图书在版编目(CIP)数据

信息论与编码/仇佩亮等编著．—2版．—北京：高等教育出版社，2011.4

ISBN 978-7-04-031706-0

Ⅰ.①信… Ⅱ.①仇… Ⅲ.①信息论-高等学校-教材②信源编码-编码理论-高等学校-教材③信道编码-编码理论-高等学校-教材 Ⅳ.①TN911.2

中国版本图书馆CIP数据核字(2011)第018080号

策划编辑 吴陈滨	责任编辑 许海平	封面设计 张 楠	责任绘图 尹文军
版式设计 余 杨	责任校对 王效珍	责任印制 张泽业	

出版发行	高等教育出版社	网 址	http://www.hep.edu.cn
社 址	北京市西城区德外大街4号		http://www.hep.com.cn
邮政编码	100120	网上订购	http://www.landraco.com
印 刷	中国农业出版社印刷厂印刷		http://www.landraco.com.cn
开 本	787×1092 1/16		
印 张	29.75	版 次	2003年12月第1版
字 数	730 000		2011年4月第2版
购书热线	010-58581118	印 次	2011年4月第1次印刷
咨询电话	400-810-0598	定 价	46.00元

物料号 31706-00

第二版前言

本书自第一版出版以来，已经过了8年的教学、使用实践。根据使用本教材的教师和学生的反映和建议，以及根据这几年作者对信息论与编码的教学、研究的体会，我们觉得很有必要对第一版做一些增补和修订。

本书新版改写了原来第一版中的第10章，同时对第11章做了较大的完善。新版中的第10章除了原来介绍的迭代译码算法和Turbo码外，还介绍了一些其他的先进信道编码技术，如LDPC码和喷泉码等。LDPC码在性能上已经非常接近于Shannon信道容量的极限，并已在实际通信传输系统中获得广泛应用；而数字喷泉码是一种无码率编码，它的性能突破了除删信道上已有的编码记录。数字喷泉码特别适合应用于因特网，它将极大地改善下一代因特网的传输性能。无论是Turbo码，还是LDPC码、喷泉码，它们都采用迭代的概率译码算法，因此我们在第10章也介绍了相关的几种迭代译码技术。

本书在第11章以较大的篇幅介绍多用户信息论。这不仅是因为多用户信息论是Shannon经典的单用户信息论的自然推广和发展，更是因为多用户信息论已成为当前信息论研究的前沿和主流方向。近年来，随着网络通信特别是无线网络的飞速发展，其面临的容量、效能等方面的需求矛盾日趋突出，多用户信息论无疑为解决这些问题提供了重要的理论和方法上的指导。无线网络时代的新3C技术，即竞争(Competition)、合作(Cooperation)和认知(Cognition)技术的理论基础就是多用户信息论。因此我们认为对于已经从事或即将从事无线网络通信新技术研究的人员来说，了解并掌握多用户信息论的基本概念、思想和方法是非常必要的。

本书新版对于原书其余各章均有不同程度的增补和修改，使得概念叙述更加明确，分析更为清晰。新版《信息论与编码》将更好地体现内容的基础性、理论性、应用性和可读可教性特点。我们希望本书新版能更好地满足读者的需求，更适合信息论与编码课程的教学需要。

我们对于向本书第一版中的错误和不足之处提出过宝贵意见和修改建议的老师、同学、读者们表示衷心的感谢！

作　者

2010年11月于浙江大学求是村

第一版前言

信息论与编码理论是50多年前由美国科学家C. E. Shannon、R. W. Hamming等人创立的。它以Shannon的不朽名著《通信的数学理论》为里程碑。几十年来许多优秀的学者、工程师共同努力推动了信息论与编码的理论和实践的发展，现在信息论与编码理论已成为信息科学的基础理论，也成为20世纪后半叶数字化革命的主要理论和技术支柱。

国外的一流学校在20世纪50年代末就开始设立信息论与编码课程，目前国内各高等院校的电子信息类专业本科生、研究生也都已把信息论与编码作为一门重要的专业基础理论课。由于信息论与编码的许多思想和方法已广泛地渗透到许多领域，它的许多研究成果也具有普遍意义，因此信息论与编码在许多领域，如在计算机、系统科学、统计学、物理学、生物学、经济学甚至社会学中都获得了成功的应用。信息论与编码理论对于从事这些相关领域工作和学习的人员来说也极具参考价值。

本书在作者多年教学经验和研究实践的基础上编写而成，由于信息论与编码本身既是一门工程科学，同时又是一门应用数学，因此作为教材必须既保持论述的科学性、严谨性，又要求深入浅出、通俗易懂，能为工程师所理解。本书中对于抽象的概念辅以必要的例子予以说明，对于复杂的证明则着重讲清证明思路而忽略繁琐的细节。本书所要求的数学基础是初等的，只要具有概率论、随机过程、线性代数和离散数学中的初等知识就足够了，但也要求读者具有一定的抽象思维能力。对于像Shannon理论中的精华——典型列理论和随机编码方法必须要慢慢地咀嚼、细细地品味才能体会它的真谛。

信息论与编码是一门应用科学，它最基本的应用背景是通信。著名的通信理论家Viterbi说过，如果把现代通信技术比喻成飞船，则晶体管是它的引擎，而信息论是它的方向盘。在本书中突出了信息论与编码的应用，特别强调在通信中的应用。这样理论与应用结合，有助于读者了解产生理论和解决问题的实际背景，也提高了工科学生的学习兴趣。

信息论与编码是一门不断发展的学科，虽然信息论与编码中许多新理论、新概念可以追溯到Shannon原著，但是对它们深刻的理解、生动的应用和美妙的理论化则是后来许多学者发展的，许多成果也是Shannon本人始料不及的。因此本书除了对于信息论与编码的基本内容进行全面的介绍外，还对目前信息论与编码中某些研究热点进行了介绍，如关于受限系统和受限系统编码、多用户信息论与多用户编码、Turbo码与迭代译码算法等。

本书分为11章，除第1章绪论外，第2章介绍信息量的定义和性质，第3章介绍离散信源的无损压缩编码，第4章介绍信道、信道容量和信道编码定理，第5章介绍失真受到限制的信源压缩编码问题，第6章介绍受限系统和受限系统编码，第7章介绍线性分组纠错编码，第8章介绍循环码，第9章介绍卷积码，第10章介绍Turbo码与迭代译码算法，第11章介绍多用户信息论与多用户编码。除了第6、第10、第11章以外的绝大多数内容适合于本科教学，其中带有*号的章

节适合于研究生或具有一定基础的读者进一步深入学习。

几十年来国内外有不少信息论与编码方面的优秀教科书和专著，作者在编写本书过程中得益于以前对于这些著作的学习。此外，在编写本书过程中还参阅了许多文献、资料，在此作者对于这些著作的作者深表谢意。

作者要特别感谢清华大学朱雪龙教授，他非常仔细地审阅了本书全部内容并提出许多宝贵意见，对提高本书的质量起了重要的作用。作者也要感谢浙江大学朱华飞博士、张朝阳博士和谢磊博士，他们在信息论与编码课程教学中试用了本书的部分内容，为本书提出很好的改进意见。最后要感谢作者的妻子陈邦媛教授，她不仅关心作者的生活，承担起全部家务，为作者创造了一个安宁的写作环境，而且对于本书内容的组织安排、深浅分寸的掌握提出许多建设性意见，没有她的支持本书是不可能完成的。

限于作者的水平，本书中不妥和谬误之处难免，恳请读者批评指正。

仇佩亮

2003 年 7 月于杭州浙江大学求是村

目　录

第1章 绪 论

信息论是应用近代概率统计方法来研究信息传输、交换、存储和处理的一门学科，也是源于通信实践发展起来的一门新兴应用科学。

信息是系统传输、交换、存储和处理的对象，信息载荷在语言、文字、数据、图像等消息之中。在信息论中，信息和消息是紧密相连的两个不同概念。同样一个消息，比如一张当日的报纸，对于不同的人从中可获得的信息是不一样的；同样的天气预报“明天有雨”，对于干旱地区和雨量充沛地区来说其信息含量也不一样。一张纸写上几个字成为一封家信，对于收信者是家书抵万金，但对旁人可能是废纸一张。因此信息是一种奇妙的东西，它是有别于物质和能量的一种存在。信息的本质和它的科学定义是当前科学界，乃至哲学界热衷研究的课题。信息的重要性是毋庸置疑的。控制论创始人维纳说过：“要有效地生活，就要有足够多的信息”。目前社会上流行一些提法，如“信息、材料、能源是现代科学的三大支柱”、“信息、物质、能量是构成一切系统的三大要素”这些提法充分说明了人们对信息重要性的认识。

信息的度量是信息论研究的基本问题。从目前的研究来看，要对通常意义下的信息给出一个统一的度量是困难的。存在许多种关于信息度量的定义，但至今最为成功，也是最为普及的信息度量是由信息论创始人香农(Shannon)在他的光辉著作《通信的数学理论》[17]中提出的，是建立在概率模型上的信息度量。他把信息定义为“用来消除不确定性的东西”。既然信息与不确定性相联系，因此用概率的某种函数来描述不确定性是自然的，所以香农用

$$I(A) = -\log P(A)$$

来度量事件 A 的发生所提供的信息，其中 $P(A)$ 为事件 A 的概率。这个定义与人们的直觉经验相吻合。如果一个随机试验有 N 个可能结果或者说一个随机消息有 N 个可能值，若它们出现的概率分别为 $p_1, p_2, \cdots, p_N$，则这些事件的自信息的平均值

$$H = -\sum_{i=1}^{N} p_i \log p_i$$

作为这个随机试验或随机消息所提供的平均信息也是合理的。H 也称为熵，这是借助于统计物理学中的一个名词。事实上熵作为信息的代名词也是由20世纪伟大的数学家、物理学家冯·诺依曼向香农建议的。在物理学中熵是描述系统的不规则性或不确定性程度的一个物理量。

信息论所研究的通信系统基本模型如图1.1.1所示。

信源是产生消息(或消息序列)的源。消息通常由符号序列或时间函数组成。消息取值服从一定的统计规律，所以信源的数学模型可以是一个离散的随机序列或连续的随机过程。

信源编码器把信源产生的消息变换成数字序列。对无损信源编码来说，信源编码器的目的是在保证能从其输出数字序列中无错误地恢复出输入消息序列的前提下，减少输出数字序列的

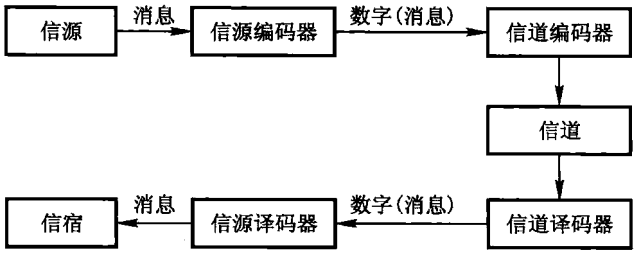

图 1.1.1 通信系统的基本模型

速率,也就是保证在不失真的条件下对输入消息序列进行压缩。在允许失真的情况下,信源编码的目的是对给定信源,在保证消息平均失真不超过某给定允许值 D 的条件下,尽量减少输出数字序列的速率。

信道在实际通信系统中是指传输信号的媒介或通道,如架空明线、电缆、电离层、人造卫星等。在信息论的模型中也把发送端和接收端的调制、解调器等归入信道,并把系统中各部分的噪声和干扰都归入信道中。在信道的输入、输出模型中,根据噪声和干扰的统计特性,用输入、输出的条件概率(或称转移概率)来描述信道特性。

信道编码器把信源编码输出的数字序列变换成适合于信道传输的,由信道入口符号组成的序列。信道编码器的最主要作用是要对其输出序列提供保护,以抵抗信道噪声和干扰。

信道译码器和信源译码器分别是信道编码和信源编码的反变换,信宿是消息的接收者,即消息的归宿。

信息论解决了通信中的两个基本问题。首先对于信源编码,信息论回答了"达到不失真信源压缩编码的极限(最低)编码速率是多少?"这一问题。香农的答复是这个极限速率等于该信源的熵 H。事实上香农认为每个随机过程,不管是音乐、语言、图像,都有一个固有的复杂性,该随机过程不能被无失真地压缩到该固有复杂性之下,这个固有复杂性就等于该随机过程的熵。信息论对通信解决的第二个问题是关于信道编码方面的。在香农以前,人们都认为增加信道的信息传输速率总要引起错误概率的增加,认为要使错误概率为零,则传输速率只能为零。但香农却出人意料地证明,只要信息传输速率小于信道容量 C,传输的错误概率可以任意地小,反过来如果超过信道容量,则传输的错误是不可避免的。对每个信道可以根据它的噪声干扰特征计算出它的容量 C。

香农信息论与信息编码技术是两个密不可分的学科领域,或者说它们是信息科学的两个不同方面。香农信息论指出了通信中信源编码和信道编码的极限速率。香农利用随机编码方法,证明了当码长趋于无限时,存在一种编码方式,能够达到这个理论上的极限速率。香农所使用的证明方法在理论上极为漂亮,但实际上无法实现。香农的证明方法是一种"存在性"证明方法。这种方法在计算上是不可实现的。对于实际的通信专家和编码专家来说必须去寻找有效的、可实现的编码方法。借助于电子科学技术的发展,无论对于信源编码,还是对于信道编码,目前都有许多具有实用价值的编、译码方案,它们的性能正逐步向香农指出的极限逼近。

编码理论工作者和通信工程师所追求的目标不仅仅是要寻找达到香农理论极限的编码方法,更重要的是要寻找可以实现的编码方法,因此,编码的复杂性是放在首位考虑的因素。在无

损压缩编码中，早期的 Huffman 编码被认为是最优的变长度压缩编码方法，但是它的复杂性随着码长的增大急剧增加，所以对于大的码长来说 Huffman 编码是不实际的。20 世纪 70 年代开始的算术编码，虽然按平均码长来说不是最佳的，但它是一个在线的算法，计算复杂性随码长线性增加，因此，算术码是一种实用的码。有人认为算术码的提出标志着无损压缩编码的一个突破。

众所周知，自然界的信号都是连续的，无论是语音信号、图像信号或各种传感信号，不可能用有限比特不失真地表示它们。因此问题在于如何设计一种编码方法，使其在给定的许可失真范围内，用最少的比特表示它们，或者说如何用给定的比特数来表示这个连续信号，使失真最小。这就是保真度意义下的压缩编码。几十年来，在这方面已发展了许多成功的实用压缩编码方法，比如矢量量化、预测编码、变换编码、子带编码等技术，其中许多技术已成为国际标准，例如 ITU 中关于语音压缩和图像压缩的标准。正是由于这些有损压缩编码技术的应用使得语音、图像信号的码率可以成十倍甚至上百倍地降低，同时使由压缩编码引起的信号质量下降不为人类感官所觉察。这些编码技术是当前各种多媒体技术的核心。

信道编码也就是通常所说的纠错编码，是另一大类信息编码技术。这类编码的目的在于检测或纠正传输中的错误，提高信息在传输中的可靠性。纠错编码中最早的 Hamming 码[55]是几乎与香农信息论同时被提出来的。早期纠错码研究集中在线性分组码，采用的数学工具是矩阵理论。到 20 世纪 60 年代，由于以有限域理论为主的抽象代数工具的引入使线性分组码的研究突飞猛进。循环码，特别是 BCH 码、RS 码等的研究，不仅为线性编码的研究打下坚实的基础，而且由于代数构造的引入使得译码复杂性大为下降。20 世纪 70 年代以后基于概率译码的序贯编码理论，特别是卷积码，获得了极大的发展。20 世纪 70 年代，纠错编码技术首先在宇宙飞船、深空通信中获得了成功应用，这极大地鼓舞了纠错编码研究者。今天由于微电子技术的发展，使以前难以实现的复杂译码算法在超大规模芯片中得到实现，从而使得纠错编码成为通信系统中不可缺少的一部分。以前由数学家们研究的技术，如 RS 码、Vitersi 算法等，已成为通信工程师的口头禅。纠错编码的成功刺激研究者寻找性能更优越的码，例如代数几何码、Turbo 码、低密度校验码（LDPC）等。这些码的性能已非常接近香农的极限。同时调制技术与纠错编码的结合，信源编码与信道编码的结合会产生一些性能更好的传输技术。相信随着科学的发展和需求的增长，新的、更好的码会不断涌现。

多个信源利用多个发信机和多个收信机在通信网络上进行信息传输会产生许多新的问题，如相互干扰、相互协作、相互叠加、反馈等。在多用户工作条件下的通信极限问题是单用户信息论的推广，称为多用户信息论或网络信息论。多用户信息论与单用户情况一样，主要研究两类问题，即多用户信源压缩编码和多用户信息在网络信道的传输。多用户信息编码中发展起来的许多思想，如叠加编码、嵌入编码、逐次抵消、时域灌水、脏纸编码、信息细化和边信息应用等，已经在理论和实际上得到了应用。有理由认为新一代的无线网络通信必须从多用户信息论中吸取思想精华，才可望获得新的突破。

香农信息论源于通信实践。它对通信领域的成功应用使得香农理论被称为通信的数学理论。但香农理论的思想、方法，甚至某些结论已渗透到许多其他学科中。

- 统计数学　香农理论本身就是一种数学理论，它与随机过程中 Ergodic（各态历经）理论有密切关系。香农编码定理的基本核心——渐近等分原理（AEP），实际上就是某种形式的大数定律。因此利用熵、互信息等概念来研究 Ergodic 系统是非常有效的。另外，用相对熵作为随机分

布之间的距离,在假设检验中、在大偏离理论中均有很好的应用。利用相对熵可以有效估计差错概率指数。

- 计算机科学 计算机和通信是密不可分的,计算能力受制于计算部件之间的通信能力,同时通信能力又受制于计算能力,所以计算和通信是一对双螺旋,信息论的每一步发展直接影响计算科学的发展。且不说各种信源编码、信道编码、存储编码技术的发展如何直接推动计算机技术的发展,即使计算机中"最佳随机数发生"这么一个简单问题也被证明与最佳信源编码等价。在计算科学中数据串的 Kolmogorov 复杂性被定义为利用通用计算机打印出这个数据串并停机所需的最短二元程序的长度。可以证明一个随机源所输出数据序列的 Kolmogorov 复杂性等于该随机源的香农熵,从而 Kolmogorov 复杂性理论与香农信息论建立了联系。

- 哲学和科学方法论 最大熵准则或最大信息原则是许多科学研究中常用的准则,实践证明这个准则是有效的、合理的。信息论赋予最大熵准则以明确的内涵。最大熵准则和最小描述长度准则都是一种科学的方法论,在信息论中可找到它们的联系。这给予相信"最简单的解释是最好的"信条的人们一个科学的佐证。

另外,信息论的思想和方法还在经济、生物等方面获得应用,已产生了"信息经济学"、"信息生物学"等边缘学科。因此,人们深信信息论的学习有助于对其他学科的研究,同时其他相关学科的研究也会促进信息论的发展。比如量子力学理论与经典信息论的结合已产生了目前发展迅速、前途不可限量的量子信息论、量子编码理论和量子计算理论等。完全可以相信这些理论是属于 21 世纪的工程科学理论,它们将对 21 世纪新科技产生巨大的作用。

第2章
熵和互信息

在本章,介绍信息论中的两个最重要的概念,即信源的熵和互信息,同时介绍它们的一些性质。信源的熵是用来刻画信源发出的消息(随机变量)的平均不确定性,而两个随机变量之间的互信息则表示一个随机变量对另一个随机变量所提供的信息量。因为对于2个随机变量来说,知道了其中一个的取值往往可以减少另一个随机变量的不确定性,这种不确定性的减少被认为是一个随机变量对另一个随机变量提供了互信息。

在本章中首先讨论事件的自信息和事件之间的互信息,然后介绍离散随机变量的平均自信息,即熵和平均互信息;进而把熵和互信息概念推广到连续随机变量和随机过程的情况。

2.1 随机变量的熵和互信息

通常的通信传输系统或信息处理系统可以用图2.1.1来表示。其中方框黑盒子中所谓的信息处理系统可以是某种确定性的处理,如线性滤波、放大等,也可以是某种不确定的处理,例如叠加上某种随机噪声、干扰,或者是经历某种随机衰落、失真等。一般用2个随机变量 X 和 Y 来表示它们的输入和输出。

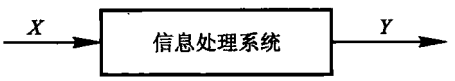

图2.1.1 通信传输系统或信息处理系统

对于随机变量 X,可以用3个量组成的三元组 $\{X,\mathscr{X},q(x)\}$ 来描述它,其中 X 是随机变量的名字(本书中一般用英文大写表示随机变量,用小写字母表示随机变量的取值),$\mathscr{X}$ 表示随机变量 X 的取值范围,$q(x)$ 表示 X 的概率分布。这样的三元组也称为概率空间。

如果 X 和 Y 是离散随机变量,它们的取值范围为

$$\mathscr{X}=\{x_k;\quad k=1,2,\cdots,K\}$$
$$\mathscr{Y}=\{y_j;\quad j=1,2,\cdots,J\}$$

对每个 $x_k\in\mathscr{X}$,X 取值为 x_k 的概率为

$$q(x_k)\overset{\text{def}}{=}q_k$$

由概率定义,q_k 应满足

$$\begin{cases} q_k\geqslant 0 & k=1,2,\cdots,K \qquad (2.1.1a)\\ \sum\limits_{k=1}^{K} q_k = 1 & \qquad (2.1.1b)\end{cases}$$

同样对每个 $y_j\in\mathscr{Y}$,随机变量 Y 取值 y_j 的概率 $\omega(y_j)$ 满足

$$\begin{cases} \omega_j \overset{\text{def}}{=} \omega(y_j) \geqslant 0 \qquad j=1,2,\cdots,J & (2.1.2a) \\ \sum_{j=1}^{J} \omega_j = 1 & (2.1.2b) \end{cases}$$

如果把一对随机变量(X,Y)看成一个新的二维随机矢量$\boldsymbol{Z}$,即$\boldsymbol{Z} \overset{\text{def}}{=} (X,Y)$,则与随机矢量$\boldsymbol{Z}$相应的概率空间为$\{Z,\mathscr{Z},p(z)\}$,它实际上代表

$$\{(X,Y),\mathscr{X}\times\mathscr{Y},p(x,y)\}$$

其中,$\mathscr{X}\times\mathscr{Y}$代表$\mathscr{X}$和$\mathscr{Y}$的直积,也就是随机矢量$(X,Y)$的取值范围,而

$$p(x,y) \overset{\text{def}}{=} P\{X=x,Y=y\}$$

由概率论知识,对每对$(x_k,y_j)\in\mathscr{X}\times\mathscr{Y}$,相应概率$p(x_k,y_j)$满足

$$\begin{cases} p(x_k,y_j) \geqslant 0 & (2.1.3a) \\ \sum_k \sum_j p(x_k,y_j) = 1 & (2.1.3b) \end{cases}$$

同时

$$\sum_k p(x_k,y_j) = \omega(y_j) \tag{2.1.4a}$$

$$\sum_j p(x_k,y_j) = q(x_k) \tag{2.1.4b}$$

相应的条件概率为

$$p(y_j|x_k) \overset{\text{def}}{=} p(Y=y_j|X=x_k) = \frac{p(x_k,y_j)}{q(x_k)} \tag{2.1.5a}$$

$$p(x_k|y_j) \overset{\text{def}}{=} p(X=x_k|Y=y_j) = \frac{p(x_k,y_j)}{\omega(y_j)} \tag{2.1.5b}$$

[例 2.1.1] 令随机变量X和Y分别代表某一城市市民的身高和体重,则$q(x_k)$和$\omega(y_j)$就分别代表随机地抽取一个市民,他身高为x_k的概率和体重为y_j的概率;$p(x_k,y_j)$则表示随机抽取一个市民,他身高为x_k,同时体重也正好为y_j的概率;而$p(x_k|y_j)$则表示在体重为y_j的人群中抽取一人,他身高为x_k的概率。

2.1.1 事件的自信息和互信息

在概率论中,如果做一个随机试验,结果发现随机变量X取某个特定值x,这时称发生了一个事件$\{X=x\}$,由于X取某个特定值x的概率为$q(x)$,若$q(x)$很小,则事件$\{X=x\}$是不太可能发生的;但如果做一次随机试验,结果出现了这个不太可能出现的事件,那么通常认为这次试验使人们获得了较多的信息。由此给出事件的自信息定义为:

定义 2.1.1 对于概率空间$\{X,\mathscr{X},q(x)\}$,事件$\{X=x_k;x_k\in\mathscr{X}\}$的自信息定义为

$$I(x_k) = -\log_a q(x_k) \tag{2.1.6}$$

当对数的底a取2时,自信息的单位为比特(bit),当对数底取e时,单位称为奈特(nat)。有时略

去了对数底 a，这时根据前后文来理解对数底。

为什么把自信息定义为概率的负对数呢？有 3 个原因。首先，这个定义符合概率越小，自信息越大的要求；其次，由于对数是数学上比较简单的函数，易于进行数学处理；第三，由这样的定义导出的一些性质与日常生活中关于信息的经验相吻合。例如在以后的章节中就可以看到信息量的可加性，可以由对数性质直接导出。

现在来考察 2 个随机变量 X 和 Y，一般来说两个随机变量之间会有不同程度的关联。比如，人的身高 X 和体重 Y 有一定关系。身高高的人一般会比矮的人的体重更重一点，所以知道了一个人的体重就或多或少获得一些关于身高的信息。下面定义 2 个事件 $\{X=x_k\}$ 和 $\{Y=y_j\}$ 之间的互信息。

定义 2.1.2 联合概率空间 $\{(X,Y),\mathscr{X}\times\mathscr{Y},p(x,y)\}$ 中 2 个事件 $\{X=x_k\}$ 和 $\{Y=y_j\}$ 之间的互信息定义为

$$I(x_k;y_j)=\log_a\frac{p(x_k|y_j)}{q(x_k)}$$

$$=-\log_a q(x_k)-\{-\log_a p(x_k|y_j)\} \tag{2.1.7}$$

在式(2.1.7)中 $-\log_a q(x_k)$ 表示事件 $(X=x_k)$ 的自信息，或者说表示事件 $\{X=x_k\}$ 的不确定性，而 $-\log_a p(x_k|y_j)$ 表示在已知 $Y=y_j$ 条件下，事件 $\{X=x_k\}$ 的不确定性，二者之差代表了不确定性的降低量。这个降低量可以被认为是由事件 $\{Y=y_j\}$ 提供的关于事件 $\{X=x_k\}$ 的互信息量。很易证明

$$I(x_k;y_j)=I(y_j;x_k) \tag{2.1.8}$$

也就是说事件 $\{X=x_k\}$ 提供关于 $\{Y=y_j\}$ 的互信息量等于 $\{Y=y_j\}$ 提供给事件 $\{X=x_k\}$ 的互信息量，所以互信息量具有对称性。

由于条件概率 $p(x_k|y_j)$ 可以大于、等于或者小于 $q(x_k)$，所以事件的互信息可正、可负，也可以为零。

如果 $\{X=x_k\}$ 和 $\{Y=y_j\}$ 是 2 个独立事件，即如果

$$p(x_k|y_j)=q(x_k) \tag{2.1.9a}$$

则

$$I(x_k;y_j)=0 \tag{2.1.9b}$$

所以，对于 2 个独立事件，它们彼此不提供任何信息量。

另一个极端的情况是 2 个事件完全相关，这时

$$p(x_k|y_j)=1 \tag{2.1.10a}$$

则

$$I(x_k;y_j)=\log_a\frac{p(x_k|y_j)}{q(x_k)}$$

$$=I(x_k) \tag{2.1.10b}$$

这也表示 $I(x_k)$ 等于要完全确定事件 $\{X=x_k\}$ 出现所需要的信息量。

[例 2.1.2] 令随机变量 X 表示人群中随机抽取的一人的性别，$X=0$ 表示抽取的为男性，$X=1$ 表示抽取的为女性。随机变量 Y 表示抽取人是否抽烟，$Y=0$ 表示抽烟，$Y=1$ 表示不抽烟。如果 (X,Y) 的联合分布如表 2.1.1 所示。