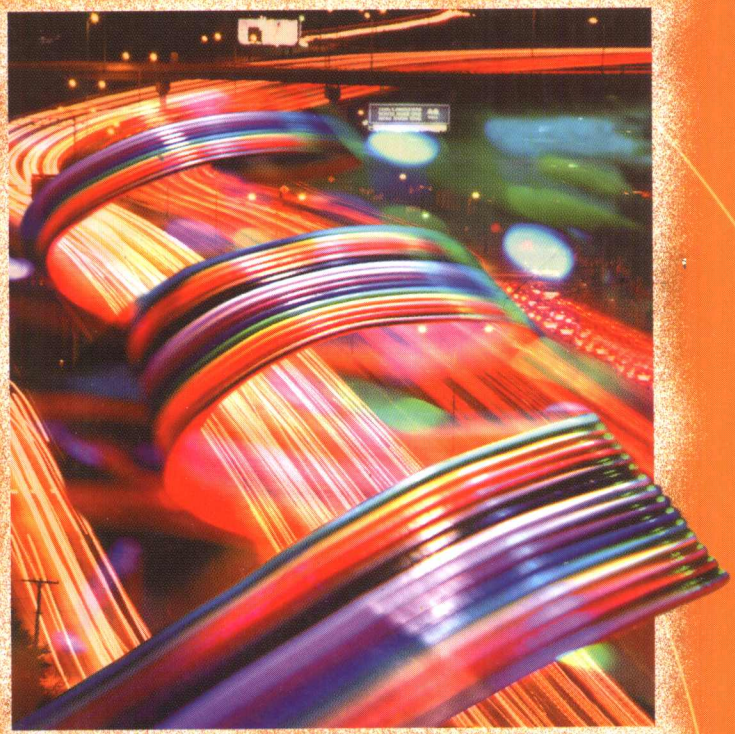
高等学校教材

信息论基础

叶中行　编



高等教育出版社

高等学校教材

信息论基础

叶中行 编

高等教育出版社

内容提要

信息论是20世纪40年代后期从长期通讯实践中总结出来的一门学科，是专门研究信息的有效处理和可靠传输的一般规律的学科。本书共分7章，内容包括：随机变量的信息度量；随机过程的信息度量；数据压缩和信源编码；数据可靠传输和信道编码；限失真数据压缩和率失真理论；网络信息理论；信息论应用等。既包括了信息论的基本理论，也涉及了一些信息处理的算法及信息论在其他领域的应用。

本书可作为数学类信息与计算科学专业的教材，也可为其他相关专业同类课程所选用。

图书在版编目(CIP)数据

信息论基础/叶中行编 .—北京：高等教育出版社，
2003.1

ISBN 7-04-011543-3

Ⅰ.信… Ⅱ.叶… Ⅲ.信息论-高等学校-教材
Ⅳ.G201

中国版本图书馆 CIP 数据核字(2002)第109649号

出版发行	高等教育出版社	购书热线	010-64054588
社　　址	北京市东城区沙滩后街55号	免费咨询	800-810-0598
邮政编码	100009	网　　址	http://www.hep.edu.cn
传　　真	010-64014048		http://www.hep.com.cn
经　　销	新华书店北京发行所		
印　　刷	北京机工印刷厂		
开　　本	787×960　1/16	版　　次	2003年1月第1版
印　　张	13.5	印　　次	2003年1月第1次印刷
字　　数	240 000	定　　价	14.60元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

前 言

(1) 信息论简史

信息论是 20 世纪 40 年代后期从长期通讯实践中总结出来的一门学科，是专门研究信息的有效处理和可靠传输的一般规律的科学。信息是系统传输和处理的对象，它载荷于语言、文字、数据、图像、影视、信号等之中，要研究信息处理和传输的规律，首先要对信息进行定量的描述，即信息的度量，这是信息论研究的出发点。但要对通常含义下的信息（如知识、情报、消息等）给出一个统一的度量是困难的，因为它涉及客观和主观两个标准，而迄今为止最成功、应用最广泛的是建立在概率模型基础上的信息度量，进而建立在此种信息量基础上的信息论成功地解决了信息处理和可靠传输中的一系列理论问题。

切略 (E.C.Cherry) 曾写过一篇早期信息理论史，他从石刻象形文字起，经过中世纪启蒙性语言学，直到 16 世纪吉尔伯特 (E.N.Gilbert) 等人在电报学方面的工作。

现代信息论实际是从 20 世纪 20 年代奈奎斯特 (H.Nyquist) 和哈特莱 (L.V.R. Hartley) 的工作开始的，他们最早研究了通信系统传输信息的能力，并试图度量系统的信道容量。

克劳德 · 申农 (Claude Shannon) 于 1948 年发表的具有里程碑性质的论文“通信的数学理论”是世界上首次将通讯过程建立了数学模型的论文，这篇论文和 1949 年发表的另一篇论文一起奠定了现代信息论的基础。申农开创性地定义了“信息”，他所定义的信息与语义无关，而是反映了将“信息”编码成由简单的 0 和 1 表示的语言的能力，由此整个通讯过程可表示成以下的过程，从一个信源发出的消息，经过编码后通过一个信道传输给接收者，接收者通过译码器将收到的信号复原成信源发出的原消息。

申农证明了任何的信息源 —— 文本、图像、影视、数据 —— 都伴随着一个可定量化的信息内涵，它表示了可以如何有效地表达这类信息的能力，此即信息压缩的基础，比如，申农证明了任何人再聪明也不可能把英文文本压缩到 1.5 比特 / 字母。申农还揭示了人们通过一个有噪音的信道传输信息的有效性的极限，称之谓信道容量。

申农以上的工作奠定了现在称之谓“信息论”的理论基础，他关于信息处理和可靠通讯的工作是当今无所不在的数据压缩、调制解调器、广播、电视、卫星通信、计算机存储乃至因特网通讯的理论基础。此外申农在第二次世界大战期间在密码方面的工作形成了现代信息安全系统的理论框架。

申农的工作也是因特网以及其他应用广泛的从音乐、影视到电子邮件等高

新技术的核心，他的工作在计算机科学、人工智能、基因工程、神经解剖学乃至金融投资学等众多领域也有广泛应用，因此有人称申农 1948 年的论文是信息时代的“基本法”。在过去的 20 世纪中只有少数几个科学家能称得上世纪科学巨人，申农就是其中之一。

自申农 1948 年的奠基性文章发表后，前苏联和美国的科学家采取了不同的研究途径进一步发展了信息论。在前苏联以辛钦 (Shiqin) 、柯尔莫哥洛夫 (Kolmogorov) 、宾斯基 (Pinsker) 和达布鲁新 (Dabrushin) 为首的一批著名数学家致力于信息论的公理化体系和更一般更抽象的数学模型，对信息论的基本定理给出了更为普遍的结果，为信息论发展成数学的一个分支作出了贡献。而在美国则是由一批数学修养很高的工程技术人员致力于信息有效处理和可靠传输的可实现性，为信息论转化为信息技术作出了贡献。

我国数学家和信息科学专家在 20 世纪 50 年代将信息论引进中国，经过 50 多年的不懈努力，特别是 20 世纪 80 年代中叶以来，一批华裔信息论专家在国际学术界崛起，为信息论的发展作出了自己的贡献。

(2) 关于申农

克劳德 · 申农, 1916 年 4 月 30 日生于美国密西根州的贝多斯克 (Petoskey) , 2001 年 2 月 24 日在长期患老年痴呆症后在美国马萨诸塞州的蒙得福特 (Medford) 去世，享年 85 岁，申农有 2 个儿子和 1 个女儿。

1936 年申农 20 岁时在密西根大学获数学和电机工程学学士学位，此后进入著名的麻省理工学院 (MIT),1938 年获硕士学位，他的硕士学位论文“延迟电路和开关电路的符号分析”后来获美国电机工程师协会优秀论文奖 (1940), 1940 年申农以题为“理论遗传学的代数学”的论文获 MIT 数学博士，但那以后他再也没写过遗传学方面的论文。

在 MIT , 申农跟随 V · 布什 (Vannevar Bush) 研究当时称为微分分析器的模拟计算机， 1940 年夏天，他曾在 AT&T 的贝尔电话实验室工作。之后，他在普林斯顿大学的高等研究院跟随著名数学家 H · 外尔 (Hermann Weyl) 工作了一年。在普林斯顿，他开始思考如何建立通信系统的恰当的数学基础。

1941 年，他回到贝尔实验室，在那里一直工作了 15 年。在第二次世界大战期间，他参与了数字密码系统的研究，包括曾用于邱吉尔和罗斯福的跨洋会议系统。 1945 年他完成了“密码学的数学理论”的报告，该文直到 1949 年才在 Bell System Technical Journal(BSTJ) 上发表，正是他对密码学的思考促进了他对通信理论的研究。 1948 年，申农发表了“通信的数学理论”这篇划时代的

杰作，奠定了信息论的基础。

1956 年，申农离开了贝尔实验室回到母校 MIT，担任通信科学方面的教授，1958 年起担任 Donner 讲座教授直到 1978 年退休。

国际电子和电机工程师协会 (IEEE) 在 1950 年初期成立了信息论学会，并于 1973 年设立申农讲座（现已更名申农奖），用以表彰对信息论发展作出卓越贡献的科学家，是国际信息论界的最高荣誉，而申农本人则是申农讲座的第一个得主。

申农后来的兴趣跃出了通信工程范围，曾致力于密码学，用概率论研究如何投资股票市场，试图在 DNA 复制和苟尔蒙信号研究中应用信息论方法，他还研究过早期的机器人，设计能在轮盘赌中取胜的计算机软件、智力游戏机。申农本人能一边骑自行车一边玩杂要，他设计了一个类似骑自行车玩杂要的机械。

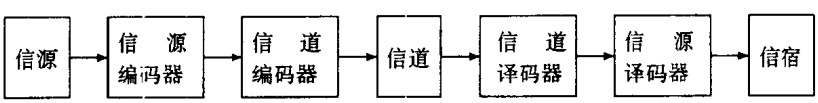
申农一生发表了一百多篇论文，1993 年出版的申农论文集收集了他 1938 - 1982 年期间发表的 127 篇论文。申农一生获得过无数的荣誉与奖励。他是 IEEE 的会士、美国科学院院士、美国艺术与科学院院士。他获得的奖励中包括 IEEE 成就奖（1966）、美国国家科学奖、以色列哈维 (Harvey) 奖（1972）和日本 Kyoto 奖（1985）。

(3) 信息论简介

由于现代通信技术的飞速发展及和其它学科的交叉渗透，信息论的研究已经从申农当年仅限于通信系统的数学理论的狭义范围扩展开来，而成为现在称之为信息科学的一个庞大体系。但是作为大学本科生课程的信息论基础，我们仍限于申农意义下的狭义信息论。一章简单介绍信息论在其它领域的应用，以供读者进一步学习的参考.

(a) 通信系统

作为通信系统的数学理论，申农在 1948 年的奠基性文章中提出了通信系统的一般模型（如下图所示）



信息的产生和发送者称之为信源，信源要将输出的消息通过某种通讯渠道传输给称为信宿的接收者．我们称信源发出的信息为消息，这里并不考虑其内含的语义信息，而只考虑它的统计特性，通信的主要目标之一是使接收端能尽

可能准确地复制信源发送的消息。为研究方便起见，先将信源与信道分开来介绍。

(b) 信源编码

通常信源发出的消息是有冗余的，比如普通语言与文字是有高冗余度的，为了有效地进行通信，往往对有冗余的消息先进行无冗余或少冗余编码，或称压缩编码，这就是信源编码器的任务。接收端的信源译码器对收到的压缩后的信息进行编码的逆运算，或称译码，将信源发送的原消息复现，此即信源译码器的任务。根据消息的不同特征及信宿对复制消息的不同要求，复制时可以无失真，如对文本消息；也可以允许有一定失真，如对语音、图像、影视信息。由此发展起来了无失真信源编码和允许失真的率失真理论，以寻求信息压缩的最优理论极限；同时也寻求压缩率尽可能接近这个理论极限的实用压缩编码技术，这两部分构成了信源编码理论的主体。

(c) 信道编码

信源发出的消息经过压缩后要通过某种渠道传输给接收者，这种渠道称之为信道，实际信道包括电缆、光纤、微波、无线通讯等。信道中通常会有噪声干扰，使传输的消息产生失真，如语音通信系统产生的噪音、电视系统中的“雪花”干扰等。由于噪声的存在，使信道能可靠传输信息的能力受到限制，信道的最大理论信息传输速率称之为信道容量。信道只能用低于信道容量的速率来可靠地传输信息，如传输速率超过了信道容量，就会出现错误。

研究各种特定信道的容量是信息论的另一基本问题。为了增加传输信息的抗干扰能力，就需增加信息的冗余度，这就是信道编码器的任务。经信道编码增加了冗余度的消息通过信道后，即便受噪声干扰可能会出现一些错误，但信道译码可利用增加的冗余信息进行纠错，尽可能正确地复制出信道编码前的消息。信道编码技术的研究已形成了现在称之为纠错码的一个完整体系。

信道容量和纠错码就构成了信道编码理论的主体。

(d) 复合信源 – 信道编码

对信源与信道分隔开讨论后再回到 (a) 中所示的通信系统，我们就可以发现，只要信源编码的压缩率不超过信道容量，就可以达到既有效又可靠地传输信息的目的，也把看似矛盾的信源编码（要减少冗余）和信道编码（要增加冗余）统一在一起。同时为提高效率，可以把信源与信道编码器合二而一为一个编码器，同理也可把信道与信源译码器合二而一为一个译码器。

(e) 网络通信

以上介绍的通信系统是点对点、一对一的通信模型。在 20 世纪 70 年代以

后，随着通信技术的发展，如卫星通信、广播通信、无线通信、计算机网络通信的发展，需要从理论上建立网络通信的数学模型，并回答在网络通信中信息压缩和可靠传输的理论极限以及实用编译码技术，由此发展起来的网络信息理论或称多用户信息理论。在 20 世纪最后 20 年中成为信息论界研究的一个热点，出现了众多的通信模型，涉及信源编码的有相关信源模型和信源多终端模型，涉及信道的有多址信道、广播信道、双向信道、串扰信道、中继（转播）信道、防窃听信道等。但由于网络通信模型在数学处理上的困难及实用编译码技术研究的滞后，上述模型还有相当一部分在理论上没有完全解决，或没有得到真正的实际应用，因此近年来对网络信息理论的研究有所降温。

(f) 密码学

信息在通信过程中的安全问题也是通信理论需研究的重要问题，申农在 1949 年在密码方面的工作形成了现代信息安全系统的理论基础。由于军事、经济、金融、计算机网络对信息安全的强烈需求使密码学迅速发展而成为信息论的一个重要分支。

传统的密码系统是通过对明文的变换使除拥有密钥的合法受信者外，其它任何人都不能从密文中获得明文与密钥。因此传统密码学研究分为密码的设计和密码的分析与破译两方面，但这类系统需通过一个安全信道传送所使用的密钥，代价昂贵，因此仅限于军事、政府等要害部门应用。

由于商业、金融、科技、计算机网络系统中日益增长的对于保密通信的需要，1976 年，迪飞 (Diffie) 和海尔曼 (Hellman) 提出了公开密钥系统的新概念，标志着密码学发展进入了一个新阶段，现在密码学已成为信息安全的研究主体。

(4) 信息论教材

信息论已经成为现代信息科学的一个重要组成部分，它是现代通信和信息技术的理论基础。现代信息论又是数学概率论下的一个分支，与遍历性理论、大偏差理论以及统计力学等都有密切关系，因此信息论已成为大学诸多专业的必修课或选修课，但以往较多的是在通信工程、电子工程、信息工程等专业开设. 近年来，由于专业调整，国内在数学类下开办了“信息与计算科学”专业，该专业的开办顺应了现代高科技对数学日益增长的需要，因此受到众多学子的青睐，国内已有近 200 所院校开办了此专业，“信息论”也自然成了该专业的必修课或主要选修课。

目前国内外已有一批关于信息论的教材，本书参考文献中列举了一些，这些教材各有所长，但其中大部分适用于工程学科的本科生或研究生。作者自 20

世纪 80 年代开始，长期从事信息理论与信息处理的科研与教学工作，积累了一定的经验，考虑到“信息与计算科学”专业对本课程的需求，编写了这本教材。本教材对信息论涉及的内容进行了精选，并将信息论研究的传统方法（如随机码方法）与现代数学方法（如典型序列法）相结合并贯穿始终，同时注意理论与应用的结合，在讨论信息论基本编码定理的基础上也介绍了一些实用算法，使之尽可能地适用于一般院校该专业本科生所用。

本书共分七章，第一章随机变量的信息度量，介绍了基本的信息量和它们的主要性质；第二章随机过程的信息度量和渐近等分性，将第一章的信息度量推广到随机过程，基于大数定理的渐近等分性是第一个信源编码定理的基础；第三章数据压缩和信源编码，讨论信源编码基本定理及数据压缩的实现技术，特别介绍了几种有很强实用价值的信源编码方法；第四章数据可靠传输和信道编码，主要讨论信道容量及基于它的信道编码定理，并简单介绍了线性分组纠错码，但没有进一步讨论卷积码和代数码；第五章限失真数据压缩和率失真理论，讨论了允许一定失真的信源编码的率失真理论，着重介绍了率失真函数的计算。

以上各章讨论的均是离散模型，第六章连续信源和信道编码理论，首先将信息度量推广到连续情形，然后着重讨论了高斯信道与高斯信源；第七章网络信息理论，在简述了各种多用户信源和信道通信模型后，只简单介绍了其中已得到完全解决的几个简单模型。

本教材主要对象是大学数学类数学与应用数学专业、信息与计算科学专业的本科生，也可作为其它相关专业（如通信工程、电子工程等）同类课程的教材。

本教材适用于 54 学时的课程，当然在增删部分内容后也可用于 36 学时的短课程或 72 学时的扩展课程。

(5) 致谢

本教材的编写得到了上海交通大学九五重点教材基金的资助，其出版则得到了高等教育出版社的大力支持。

美国南加州大学的张箴教授详细阅读了本书前四章的初稿，南开大学沈世镒教授阅读了全书，他们都提出了许多宝贵的修改意见；加拿大滑铁卢大学的杨恩辉教授提供了克佛 - 杨 (Kieffer-Yang) 通用信源编码方法的通俗描述；周煦、王俊、姚奕帮助用 CCTEX 打印了本书。德国康斯坦茨 (Konstanz) 大学数学与统计学系西格佛瑞特 · 海勒 (Siegfried Heiler) 教授盛情邀请作者访问该校，

本书的部分章节与习题就是 2002 年暑假在美丽的博登湖畔完成的。

对他们的帮助、支持和关心表示衷心感谢。

作者由衷地感谢攻读硕士和博士阶段的导师南开大学胡国定教授、沈世镒教授和美国康奈尔 (Cornell) 大学托比 · 贝尔格 (Toby Berger) 教授，是他们将作者引入信息论以至信息科学这个博大精深，使作者从中得到无穷乐趣并值得为之奋斗终身的科学领域。

作者特别感谢妻子毛经义和女儿叶雷对本人的关爱、理解和支持，没有良好的写作环境，本书的成书几无可能。

叶中行

2002 年 10 月于沪

目 录

第一章　随机变量的信息度量

本章主要介绍离散随机变量（或离散概率分布）的各种信息度量。首先引入信息的定义，然后定义了随机变量和随机向量基本的信息度量自信息、熵、联合熵、条件熵和相对熵，进而介绍互信息，最后介绍推广熵函数 —— 广义熵。在介绍这些信息度量的同时也讨论了它们的一些基本性质和计算方法。

§1.1　自信息

什么是“信息”？“信息”这个词被应用之广以至于很难对它下一个明确的定义，因为在不同的场合，它被赋予了不同的含义。在通讯领域，“信息”是指通信的消息，在信号处理方面，“信息”包含了数字、数据、图像等进行运算和处理所需的条件、内容和结果，当然“信息”也可以是人类对外部世界的感知。但是申农最初对“信息”进行定量化研究时是把“信息”限于“通信的消息”，并以此为出发点研究了通信的数学理论，形成了狭义的申农信息论，使通信技术从经验走向科学，开辟了通信科学的新纪元，同时也为整个信息科学的形成和发展奠定了必要的理论基础。本书主要研究作为“通信的消息”来理解的狭义的申农信息论，该理论对信息的形式化描述，抛开了信息的语义，给出了信息的度量，这种度量将使我们可以用来描述信源和信道的特征，优化信息的处理（压缩），改进信息传输的可靠性。

所谓“信源”，是指消息的来源，如信源输出的消息是以取值离散的符号形式出现，其不同符号数可以是有限个，或可列无限个，我们称其为离散信源。如信源输出的消息的取值是连续的，可取不可列无限多个值，我们称其为连续信源。本章首先讨论离散信源。

通常用随机变量 X 表示一个离散信源，X 的可能取值，即信源可能输出的不同符号用集合 $\mathcal{X}$ 表示。如掷硬币这个随机试验看作一个信源的话，其取值集合为 $\mathcal{X}$={ 正，反 }，掷骰子的结果可用集合 $\mathcal{X}=\{1,2,3,4,5,6\}$ 表示。

当信源发出某个信号 $x_0 \in \mathcal{X}$ 后，它提供了多少信息呢？即要解决信息的度量问题，我们把它称为 x_0 的自信息，记为 $I(x_0)$。它是信号 x_0 的不确定性的一种度量，而 x_0 的不确定性即是它发生的可能性，这可以用 x_0 发生的概率 $p(x_0)$ 的大小来描述，因此 $I(x_0)$ 应当是概率 $p(x_0)$ 的一个函数。那么 $I(x_0)$ 是一个什么样的函数呢？它应该满足哪些性质呢？

首先 $x \in \mathcal{X}$ 的概率 $p(x)$ 越大，其发生的可能性越大，不确定性越小，$I(x_0)$ 应当越小，因此 $I(x_0)$ 应当是概率 $p(x)$ 的单调减函数。其次如果信源连续独立

地发出 2 个信号 x,y，即它们的联合分布 $p(x,y)=p(x)p(y)$，则 x,y 的自信息应是它们各自信息量之和，即 $I(x,y)=I(x)+I(y)$。于是我们得到自信息应满足的几条公理：

(i) 非负性：$I(x)\geq 0$;

(ii) 如$p(x)=0$，则$I(x)\to\infty$;

(iii) 如$p(x)=1$，则$I(x)=0$;

(iv) 严格单调性：如果$p(x)>p(y)$，则$I(x)<I(y)$;

(v) 如果$p(x,y)=p(x)p(y)$，则$I(x,y)=I(x)+I(y)$。

根据这 5 条公理我们可以得到自信息量表示的唯一性定理。

定理 1.1.1 若自信息 $I(x)$ 满足上述 5 个条件，则

$$I(x)=C\log\frac{1}{p(x)}$$

其中 C 为常数。

我们只需先证明以下引理。

引理 1.1.2 如果实函数 $f(x)(1\leq x<\infty)$ 满足以下条件：

(i) $f(x)\geq 0$,

(ii) $f(x)$是严格单调增函数，即$x<y\Rightarrow f(x)<f(y)$,

(iii) $f(x\cdot y)=f(x)+f(y)$,

则 $f(x)=c\log x$。

证明 反复使用 (iii)，对任意自然数 k，我们有

$$f(x^k)=f(x\cdot x^{k-1})=f(x)+f(x^{k-1})=\cdots=kf(x) \tag{1.1}$$

从而 $f(1)=0$。进而由于 (i) 和 (ii)，对于任意 $x>1, f(x)>0$，对于任意大于 1 的 x,y 与任意自然数 k，总可以找到非负整数 n，使

$$y^n\leq x^k<y^{n+1}$$

取对数并除以 $k\log\ y$ 得

$$\frac{n}{k}\leq\frac{\log\ x}{\log\ y}<\frac{n+1}{k} \tag{1.2}$$

另一方面，由 (1.1) 及条件 (ii) 可得

$$nf(y)\leq kf(x)<(n+1)f(x)$$

或

$$\frac{n}{k}\leq\frac{f(x)}{f(y)}<\frac{n+1}{k} \tag{1.3}$$

由 (1.2)， (1.3) 我们有

$$|\frac{f(x)}{f(y)} - \frac{\log\ x}{\log\ y}| \leq \frac{1}{k}$$

当 $k \to \infty$ 时，

$$\frac{f(x)}{f(y)} = \frac{\log\ x}{\log\ y}$$

因此

$$\frac{f(x)}{\log\ x} = \frac{f(y)}{\log\ y} = c$$

或

$$f(x) = c\log\ x$$

为证明定理 1.1.1， 只需对 $f(\frac{1}{p(x)}) = I(p(x))$ 应用引理即可。

定义 1.1.1 设 $x \in \mathcal{X}$ 有概率 $p(x)$，则 x 的自信息定义为 $I(x) = \log \frac{1}{p(x)}$。

§1.2 熵、联合熵、条件熵

上一节我们定义了信源发出的每个信号的自信息，那么对整个信源来说，其每个信号的平均信息量是多少？我们把这个信息量称为熵。

如果用随机变量代表一个信源，则熵就是它的平均不肯定性的度量。设 X 是取值于离散字母集 $\mathcal{X}$ 的随机变量($\mathcal{X}$ 也称状态集)，其概率分布函数为 $p(x) = P_r\{X = x\}, x \in \mathcal{X}$，我们用 $p(x)$ 和 $p(y)$ 分别表示随机变量 X 和 Y 的概率分布函数，有时为明确区别起见，用 $p_X(x)$ 和 $p_Y(y)$ 表示。为简单起见，也用 $X \sim p(x)$ 表示 X 服从分布 $p(x)$。

定义 1.2.1 离散随机变量 X 的熵定义为

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)$$

我们也用 $H(p)$ 表示这个熵，有时也称它为概率分布 p 的熵，其中对数函数以 2 为底时，熵的单位为比特 (bit)，若对数以 e 为底，则熵的单位为奈特 (nat)，若对数以 10 为底，则熵的单位为哈特 (hartley)。注意熵只是概率分布 p 的函数，与 X 取什么值并无关系。用 E 表示数学期望，E_p 表示关于分布 p 的数学期望，即

$$E_p g(X) = \sum_{x\in\mathcal{X}} g(x)p(x)$$

则熵可表示为随机变量 $\log\frac{1}{p(X)}$ 的数学期望，即

$$H(X)=E_p\log\frac{1}{p(X)}$$

可见熵是自信息的概率加权平均值。熵有以下一些性质。

引理 1.2.1 $H(X)\geq 0$，且等号成立的充要条件是 X 有退化分布。

证明 因 $0\leq p(x)\leq 1$，从而 $-p(x)\log p(x)\geq 0$，由 $H(X)$ 定义即得 $H(X)\geq 0$，其中等号成立的充要条件为 $p(x)=0$ 或 $p(x)=1$。由概率分布的定义 $p(x)\geq 0,\sum_x p(x)=1$ 知，只能有一个 x_0 使 $p(x_0)=1$，而对其它 $x\in\mathcal{X}-\{x_0\}$，$p(x)=0$，即 p 为退化分布。

例 1.2.1 设

$$X=\begin{cases}1, \text{依概率}p\\ 0, \text{依概率}1-p\end{cases}$$

则 $H(X)=-p\log p-(1-p)\log(1-p)\overset{\text{def}}{=}h(p)$。函数 $h(p)$ 的图形见图 1.2.1，以后我们将经常用到这个函数，称之为二进熵函数。

例 1.2.2 设 X 服从有限集 $\mathcal{X}$ 上的均匀分布，即 $p(X=x)=\frac{1}{\|\mathcal{X}\|}$（其中 $\|\mathcal{X}\|$ 表示集合 $\mathcal{X}$ 中元素个数），则

$$H(X)=-\sum_{x\in\mathcal{X}}\frac{1}{\|\mathcal{X}\|}\log\frac{1}{\|\mathcal{X}\|}=\log\|\mathcal{X}\|$$

在本章习题中将计算一些常见离散分布的熵。

以下我们把熵的概念推广到随机向量，定义联合熵和条件熵。

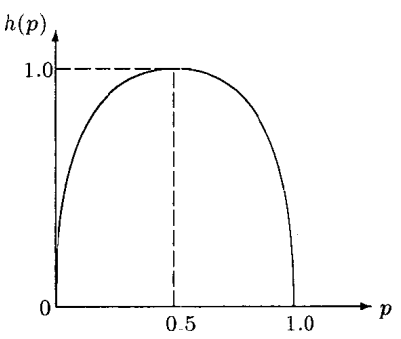


图 1.2.1 函数 $h(p)$ 的图形

定义 1.2.2 设一对随机变量 (X,Y) 的联合分布为

$$p(x,y)=P_r\{X=x,Y=y\},x\in\mathcal{X},y\in\mathcal{Y}$$