第 1 章

绪　　论

信息论或称为通信的数学理论，是应用近代数理统计方法研究信息的度量、传输、存储和处理的科学。信息是信息论中最基本、最重要的概念，是一个既复杂又抽象的概念。

本章介绍信息论和编码理论的基本情况，包括信息的概念、信息传输系统的构成、信息论和编码理论的研究内容及演变进程。

1.1　信息的概念

1.1.1　什么是信息

信息是消息或信号中的内容和意义，消息或信号是信息的载体。通信的本质在于传输信息。最简单的通信系统包括信源、信道和信宿三个部分，如图 1.1 所示。

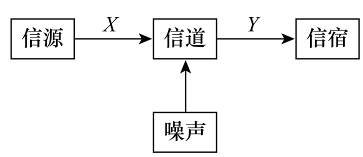


图 1.1　简单通信系统模型

信源产生能够被感觉器官所感知的消息，比如文字、符号、数据、语音、图像等，这些消息进一步转变成适于电子系统传输和处理的信号。

狭义信道从表面山看是将载荷消息的物理信号从发送端传送到接收端的传输媒质，从信息论角度看是信息传输通道，往往用符号的转移概率来表征信道。

信宿是消息传送的目的地，即接收消息的人或机器。

通信系统中形式上传输的是消息，但实质上传输的是信息。通信过程是对消息不确定性消除或部分消除的过程。不确定性消除的越多，获得的信息越多。

1.1.2　如何度量信息

事件信息量与不确定性消除程度有关，而事件的不确定度又可用其出现的概率来描述。要使得信息的定义符合人们的常识性认知，则消息事件 x_i 发生所提供的信息量 $I(x_i)$ 与事件发生概率 $P(x_i)$ 应具有如下规律：

（1）事件中所含信息量是该事件出现概率的函数。

（2）事件出现概率大小与该事件所含信息量多少成反比关系。即概率小，信息量大；概率大，信息量小。

（3）信息具有可加性，即彼此统计独立的消息提供的总信息量等于各消息提供的信息量之和。

引进对数函数定义自信息量，可满足上述三项要求：

$$I(x_i)=\log_a\frac{1}{P(x_i)}=-\log_a P(x_i)$$

这就是香农关于自信息的定义，单位与对数的底有关。$I(x_i)$可以表达两种含义：

（1）信源输出消息前，该消息客观存在的不确定度；

（2）信源输出消息后，该消息提供的信息量。

另外，通信系统接收端收到一个消息后，获得关于信源某个消息的信息量是多少呢？由此引入互信息的概念。

通信前 $I(x_i)$ 代表事件 x_i 的先验不确定度，通信后

$$I(x_i\mid y_j)=\log_a\frac{1}{P(x_i\mid y_j)}$$

代表事件 x_i 的后验不确定度。通信前后关于事件 x_i 不确定度的消除量定义为互信息。即

$$I(x_i;\ y_j)=I(x_i)-I(x_i\mid y_j)=\log_a\frac{1}{P(x_i)}-\log_a\frac{1}{P(x_i\mid y_j)}$$

为信宿获得的信息量。

1.2 信息传输系统

图 1.2 是一个完整的数字通信系统模型，是信息传输系统的一个典型应用，可将其抽象为信息传输系统，如图 1.3 所示。

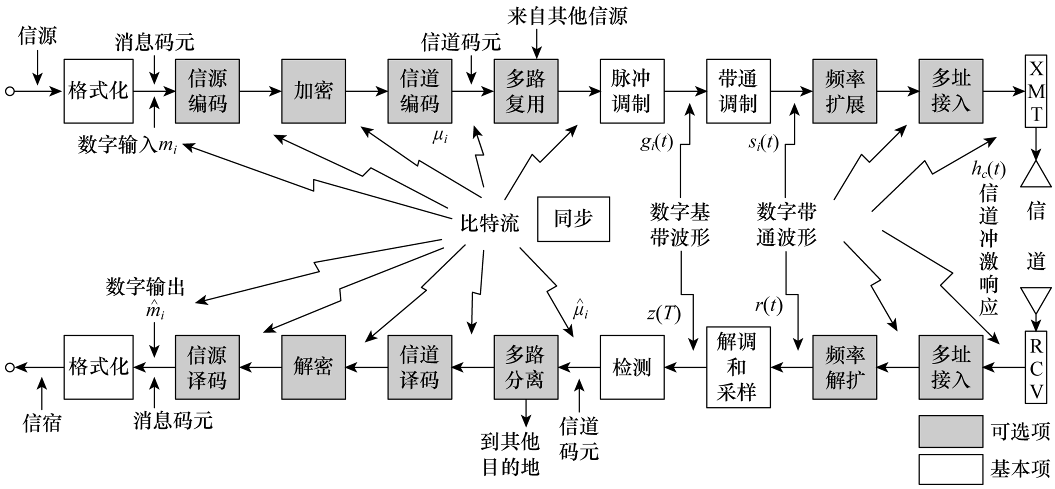


图 1.2 数字通信系统模型

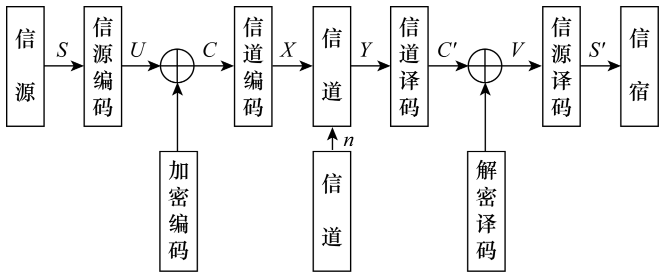


图 1.3　信息传输系统模型

该简化的信息传输系统模型可以概括为以下几部分。

1. 信源

产生消息的源，可以是人、机器或者其他事物，消息可以是文字、语言、图像等。信源可以用随机变量、随机矢量以及随机过程加以描述。

2. 编码

（1）信源编码。将信源输出的消息进行适当的变换和处理，提高信息传输的有效性。

（2）信道编码。对消息进行变换和处理，提高信息传输的可靠性。

（3）加密编码。确保消息只能被授权者接收，且接收到的消息是真实的，提高信息传输的保密性和认证性。

3. 信道

将调制和解调扩展为物理信道的一部分，只关注编码器输出符号和译码器输入符号的统计依赖关系，新的等效信道称为编码信道。

4. 译码

和编码相对应，是发端编码部分的逆变换，包括信源译码、解密译码、信道译码，最大可能正确地恢复出原始消息序列。

5. 信宿

信息传送的对象，即接收信息的人、机器或其他事物。

1.3　信息论的研究内容

信息论是信息科学的主要理论基础之一，是在长期通信工程实践和理论基础上发展起来的。发展历史虽然年轻，但它对科学技术的影响是相当深刻的，已发展为一门独立的理论学科。

1.3.1　狭义信息论（经典信息论）

狭义信息论主要研究信息的测度、信道容量以及信源和信道编码理论等问题。这部分内容是信息论的基础理论，又称为香农信息论。

香农理论的核心是：在通信系统中，采用适当的编码后，能够实现高效率和高可靠地传输信息，并给出了信源编码定理和信道编码定理。从数学观点看，这些定理是最优编码的存在定理。但从工程的观点看，这些定理不是结构性的，即不能从定理的结果直接得到实现最优编码的具体途径。然而，它们给出了编码的性能极限，在理论上阐明了通信系统中各种因素之间的相互关系，为人们寻找最佳通信系统提供了重要的理论依据。

1.3.2 一般信息论(工程信息论)

一般信息论主要研究信息传输和处理问题，除了香农理论外，还包括编码理论、噪声理论、信号滤波和预测、统计检测和估计理论、调制理论、信息处理理论、保密理论等。维纳和苏联数学家柯尔莫哥洛夫在最佳线形滤波理论、统计检测和估计理论、噪声理论等领域做出了重要贡献。

1.3.3 广义信息论

广义信息论是现代信息科学理论，它是一门新兴的综合性学科，不仅包括上述两方面的内容，而且包括所有与信息有关的领域，如模式识别、计算机翻译、心理学、遗传学、生物学、神经生理学、语言学、语义学等，甚至包括了社会、人文、经济等学科中有关信息的问题。

1.4 信息论的发展进程

信息论从诞生到今天，已有 60 多年历史，它在长期的通信工程实践和理论研究的基础上逐渐发展，已成为一门独立的理论科学。通信系统是人类社会的神经系统。纵观 100 多年电通信系统(电信系统)的发展历史，一个很有意义的事实是：一旦物理学中的电磁理论以及电子学理论有某些进展，很快就会促进电信系统的创造发明或改进。

1820—1830 年法拉第(M. Faraday)发现电磁感应的基本规律后，不久莫尔斯(F. B. Morse)就建立起电报系统。1876 年，贝尔(A. G. Bell)又发明了电话系统。

1864 年麦克斯韦(Maxell)预言了电磁波的存在，1888 年赫兹(H. Hertz)用实验证明了这一预言。接着 1895 年英国的马可尼(G. Marconi)和俄国的波波夫(A. C. Popov)发明了无线电通信系统。

1907 年福雷斯特(Lee de Forest)发明能把电磁波进行放大的电子管之后，出现了远距离无线电通信系统。大功率超高频电子管的发明则促成了电视广播系统的建立(1925—1927 年)。之后，随着微波电子管的出现，在 30 年代末和 40 年代的第二次世界大战初期，微波通信系统、微波雷达等得到迅速发展。

20 世纪 50 年代后期发明的量子放大器，60 年代初发明的光技术，使人类进入了光纤通信的时代。

现代信息论的理论研究可以认为开始于 20 世纪 20 年代奈奎斯特和哈特莱的工作。

1832 年莫尔斯电报系统中的高效率编码方法对后来香农的编码理论是有启发的。

1885 年凯尔文(L. Kelvin) 曾经研究过一条电缆上的极限传信问题。

1922 年卡逊(J. R. Carson) 对调幅信号的频谱结构做了研究，并明确了边带的概念。

1924 年奈奎斯特(H. Nyquist) 的“影响电报速率因素的确定”一文，1928 年哈特莱(R. V. Hartley) 的“信息传输”一文研究了通信系统传输信息的能力，并给出了信息度量的初步方法。

1936 年阿姆斯特朗(E. H. Armstrong) 提出增加信号带宽可以增强系统抑制噪声干扰的能力，推动了调频通信的发展。

20 世纪 40 年代初期，维纳把随机过程和数理统计的观点引入通信和控制系统中，揭示了信息传输和处理过程的统计本质。他还利用自己在 30 年代提出的广义谐波分析理论对信息系统中的谐波过程进行谱分析。

信息论产生后的重要发展：

信道编码定理：1952 年费诺(R. M. Fano) 给出并证明了费诺不等式，并给出了关于香农信道编码逆定理的证明。1961 年费诺描述了分组码的码率、码长和错误概率的关系，并提供了香农信道编码定理的充要性证明。1965 年格拉格尔(R. G. Gallager) 提供了更为简明的证明方法。1972 年阿莫托(S. Arimoto) 和布莱哈特(R. Blahut) 分别发展了信道容量的迭代算法。

信道容量：1964 年霍尔辛格(J. L. Holsinger) 继续香农的工作，开展对有色高斯噪声信道容量的研究。1969 年平斯克(M. S. Pinsker) 提出了具有反馈的非白噪声高斯信道容量问题，并由科弗尔(T. M. Cover) 在 1989 年证明。

无失真信源编码：1952 年费诺提出了一种费诺码，同年霍夫曼(D. A. Huffman) 构造了一种 Huffman 编码方法，并证明了它是一种最优码。1956 年麦克米伦(B. Mcmillan) 首先证明了唯一可译变长码的 Kraft 不等式。1968 年艾利斯(P. Elias) 在香农-费诺码的基础上提出了算术编码的初步思路。1976 年瑞斯桑尼(J. Rissanen) 给出了算术编码方案，并于 1982 年和兰登(G. G. langdon) 合作将算术编码系统化，省去了乘法运算。1977 年齐弗(J. Ziv) 和兰佩尔(A. Lempel) 提出 LZ 码，并证明此方法可以达到信源的熵值。1990 年贝尔(T. C. Bell) 对 LZ 算法做了一系列的改进，现已广泛用于文本的数据压缩中。

纠错码理论：1950 年汉明(R. W. Hamming) 为使贝尔实验室的计算机具备有检测错误能力的运行程序，首先提出了纠正一位错误的编码方法，建立了线性分组码的基本思想。随后格雷(Marcel J. E. Golay) 提出了纠正二位和三位错误的格雷码。1954 年 Reed 和 Muller 提出新的分组码 RM 码。(1969—1977 年，RM 码在火星探测方面得到了极为广泛的应用。即使在今天，RM 码也具有很大的研究价值，其快速的译码算法非常适合于光纤通信系统)。1957 年 E. Prange 提出循环码，代数编码理论成型。循环码的一个非常重要的子集就是分别由 Hocquenghem 在 1959 年、Bose 和 Ray-Chaudhuri 研究组在 1960 年几乎同时提出的 BCH 码(Bose-Chaudhuri-Hocquenghem)。

1960 年 Reed 和 Solomon 将 BCH 码扩展到非二元的情况，得到了 RS (Reed-Solomon) 码。1967 年，Berlekamp 给出了一个非常有效的译码算法后，RS 码得到了广泛的应用。此后，RS 码在 CD 播放器、DVD 播放器中得到了很好的应用。虽然分组码在理论分析和数学描

述方面已经非常成熟，并且在实际的通信系统中也已经得到了广泛的应用，但分组码固有的缺陷大大限制了它的进一步发展。首先，由于分组码是面向数据块的，因此，在译码过程中必须等待整个码字全部接收到之后才能开始进行译码。在数据块长度较大时，引入的系统延时是非常大的。分组码的第二个缺陷是它要求精确的帧同步，即需要对接收码字或帧的起始符号时间和相位精确同步。

1955 年 Elias 等人提出卷积码以改善分组码所存在的固有缺点。卷积码充分利用了各个信息块之间的相关性，编码过程连续进行。同样，在卷积码的译码过程中，不仅要从本码中提取译码信息，还要充分利用以前和以后时刻收到的码组，从这些码组中提取译码相关信息，而且译码也是可以连续进行的，这样可以保证卷积码的译码延时相对比较小。通常，在系统条件相同的条件下，在达到相同译码性能时，卷积码的信息块长度和码字长度都要比分组码的信息块长度和码字长度小，相应译码复杂性也小一些。

1961 年由 Wozencraft 和 Reiffen 提出，Fano 和 Jelinek 分别在 1963 年和 1969 年进行改进了卷积码的序贯译码算法，是基于码字树图结构的一种次最优概率译码算法。1963 年 Massey 提出门限译码算法，利用码字的代数结构进行代数译码。1967 年 Viterbi 提出 Viterbi 最优算法，是基于码字格图结构的一种最大似然译码算法。在 Viterbi 译码算法提出之后，卷积码在通信系统中得到了极为广泛的应用，如 GSM、3G、商业卫星通信系统等。(A. J. Viterbi 也是高通公司(Qualcomm)的创始人之一。高通是最早实现商用 CDMA 蜂窝移动系统的公司，因此 Viterbi 被世界公认为 CDMA 之父)

近年来，在信道编码定理的指引下，人们一直致力于寻找能满足现代通信业务要求，结构简单、性能优越的优秀编码方案，并在分组码、卷积码等基本编码方法和最大似然译码算法的基础上提出了许多构造优秀编码及简化译码复杂性的方法，提出了乘积码、代数几何码、低密度校验码(LDPC，Low Density Parity Check)、分组-卷积级联码等编码方法和逐组最佳译码、软判决译码等译码方法以及编码与调制相结合的网格编码调制(TCM，Trellis Coded Modulation)技术。其中级联码、软判决译码和 TCM 技术对纠错码的发展有较大影响。

1993 年 C. Berrou、A. Glavieux 和 P. Thitimajshima 首次提出了一种新型信道编码方案-Turbo 码。它很好地应用了 Shannon 信道编码定理中的随机性编译码条件，获得了几乎接近 Shannon 理论极限的译码性能。1997 年 Host、Johannesson、Ablov 提出了编织卷级码(WCC，Woven Convolutional Code)的概念。它是一种组合码，其系统结构可完全包容传统分组码、卷级码以及各类 Turbo 码，结构综合了并行级联卷级码(Turbo 码)和串行级联卷级码的结构特点，当外编码器个数足够多时，该码型完全拥有了 Shannon 编码定理中随机长码的特性，因此，其纠错性能理论上比 Turbo 码要优异。

最佳噪声通信系统模型：在香农理论基础上给出的最佳噪声通信系统模型近年来正在成为现实。

信号检测理论：在噪声中信号过滤与检测基础上发展起来的信号检测理论和抗干扰编码基础上发展起来的编码理论已成为现代信息论的两个重要分支。

网络信息论：1961 年 Shannon 的论文“网络通信通道”开拓了网络信息论的研究。从

70 年代开始这一领域的研究十分活跃，理论日益完善。

保密理论：1976 年 Diffe 和 Hellman 提出公开密钥密码体系后，保密通信问题得到广泛研究，形成综合线形代数、初等数论、矩阵、近世代数等相关内容的密码学理论分支。

新兴信息工程领域：光通信、空间通信、计算机互联网、移动通信、多媒体通信等领域的应用与理论研究。

信息科学：信息论与自动控制、系统工程、人工智能、仿生学、电子计算机等学科相互渗透结合形成的一门独立的新兴学科。信息科学以信息为主要研究对象，以信息的运动规律和利用信息的原理作为主要的研究内容，以信息科学方法论作为主要的研究手段，以扩大人类的信息功能为主要的研究目标。由于其研究对象(信息)的特征，信息科学区别于传统自然科学而具有独立存在性和广阔的发展前景。

信息科学由信息科学理论、信息应用技术和信息科学方法三者组成。信息科学理论主要包含信息定性理论、信息定量理论和信息应用理论。信息应用技术包括信息的获取、传递、加工处理、存储等代替和延伸人的感官及大脑的信息功能的技术，可以细分为信息获取技术(感测技术)、信息传递技术(电信技术)、信息加工处理技术(计算机技术)及信息控制技术(自动智能控制技术)。信息科学方法包括信息分析方法和信息加工方法，指导人类通过信息窗口去认识世界、改造世界。

此外，模糊信息处理、相对信息处理、主观信息处理、智能信息处理、自动化信息控制等大量崭新课题的研究相继展开。

1.5 香农简介

克劳德・香农(Claude Elwood Shannon，1916—2001)1916 年 4 月 30 日诞生于美国密歇根州的 Petoskey。在 Gaylord 小镇长大，当时镇里只有 3000 居民。父亲是该镇的法官，他们父子的姓名完全相同，都是 Claude Elwood Shannon。母亲是镇里的中学校长，姓名是 Mabel Wolf Shannon。他生长在一个有良好教育的环境中，不过父母给他的科学影响好像还不如祖父的影响大。香农的祖父是一位农场主兼发明家，发明过洗衣机和许多农业机械，这对香农的影响比较直接。此外，香农的家庭与大发明家爱迪生(Thomas Alva Edison，1847—1931)还有远亲关系。

香农的大部分时间是在贝尔实验室和 MIT(麻省理工学院)度过的。在“功成名就”后，香农与玛丽(Mary Elizabeth Moore)于 1949 年 3 月 27 日结婚，他们是在贝尔实验室相识的，玛丽当时是数据分析员。他们共有四个孩子：三个儿子 Robert、James、Andrew Moore 和一个女儿 Margarita Catherine。后来身边还有两个可爱的孙女。

2001 年 2 月 24 日，香农在马萨诸塞州 Medford 辞世，享年 85 岁。贝尔实验室和 MIT 发表的讣告都尊崇香农为信息论及数字通信时代的奠基人。

1936 年，香农在密西根大学获得数学与电气工程学士学位，然后进入 MIT 念研究生。

1938 年，香农在 MIT 获得电气工程硕士学位，硕士论文题目是 *A Symbolic Analysis of Relay and Switching Circuits*(继电器与开关电路的符号分析)。当时他已经注意到电话交换

电路与布尔代数之间的类似性，即把布尔代数的“真”与“假”和电路系统的“开”与“关”对应起来，并用 1 和 0 表示。于是他用布尔代数分析并优化开关电路，这就奠定了数字电路的理论基础。哈佛大学的 Howard Gardner 教授说，“这可能是本世纪最重要、最著名的一篇硕士论文。”

1940 年，香农在 MIT 获得数学博士学位，而他的博士论文却是关于人类遗传学的，题目是 *An Algebra for Theoretical Genetics*（理论遗传学的代数学）。这说明香农的科学兴趣十分广泛，后来他在不同的学科方面发表过许多有影响力的文章。

在读学位的同时，他还用部分时间跟温尼法·布什（Vannevar Bush）教授进行微分分析器的研究。这种分析器是早期的机械模拟计算机，用于获得常微分方程的数值解。1941 年香农发表了 *Mathematical theory of the differential analyzer*（微分分析器的数学理论），他写道：“大多数结果通过证明的定理形式给出。最重要的是处理了一些条件，有些条件可以生成一个或多个变量的函数，有些条件可使常微分方程得到解。还给出了一些注意事项，给出求函数的近似值（不能产生精确值）、求调整率的近似值以及自动控制速率的方法。”

1941 年，香农以数学研究员的身份进入新泽西州的 AT&T 贝尔电话公司，并在贝尔实验室工作到 1972 年，从 24 岁到 55 岁，整整 31 年。1956 年，他当了 MIT 的访问教授，1958 年，成为正式教授，1978 年退休。

人们描述香农的生活，白天他总是关起门来工作，晚上则骑着他的独轮车来到贝尔实验室。他的同事 D. Slepian 写道：“我们大家都带着午饭来上班，饭后在黑板上玩玩数学游戏，但克劳德很少过来。他总是关起门来工作。但是，如果你要找他，他会非常耐心地帮助你。他能立刻抓住问题的本质。他真是一位天才，在我认识的人中，我只对他一人使用这个词。”

香农与 John Riordan 一起工作，1942 年发表了一篇关于串并联网络的双终端数的论文。这篇论文扩展了麦克马洪（Percy A. MacMahon，1854—1929）1892 年在 *Electrician* 上发表的论文理论。1948 年香农则创立了信息论（information theory）。

在漫长的岁月，他思考过许多问题。除在普林斯顿高等研究院工作过一年外，主要都在 MIT 和 Bell Lab 度过。需要说明的是，在第二次世界大战时，香农博士也是一位著名的密码破译者。他在 Bell Lab 的破译团队主要是追踪德国飞机和火箭，尤其是在德国火箭对英国进行闪电战时起了很大作用。1949 年香农发表了另外一篇重要论文 *Communication Theory of Secrecy Systems*（保密系统的通信理论），正是基于这种工作实践，它的意义是使保密通信由艺术变成科学。

1948 年香农在 Bell System Technical Journal 上发表了 *A Mathematical Theory of Communication*。论文由香农和威沃共同署名。前辈威沃（Warren Weaver，1894—1978）当时是洛克菲勒基金会自然科学部的主任，他为文章写了序言。后来，香农仍然从事技术工作，而威沃则研究信息论的哲学问题。顺便提一句，该论文刚发表时，使用的是不定冠词 A，收入论文集时改为定冠词 The。

熵的概念

香农理论的重要特征是熵（entropy）的概念，他证明熵与信息内容的不确定程度有等价

关系。熵曾经是波尔兹曼在热力学第二定律引入的概念，我们可以把它理解为分子运动的混乱度。信息熵也有类似意义，例如在中文信息处理时，汉字的静态平均信息熵比较大，中文是 9.65 比特，英文是 4.03 比特。这表明中文的复杂程度高于英文，反映了中文词义丰富、行文简练，但处理难度也大。信息熵大，意味着不确定性也大。因此，我们应该深入研究，以寻求中文信息处理的深层突破。不能盲目认为汉字是世界上最优美的文字，从而引申出汉字最容易处理的错误结论。

众所周知，质量、能量和信息量是三个非常重要的量。

人们很早就知道用秤或者天平计量物质的质量，而热量和功的关系则是到了 19 世纪中叶，随着热功当量的明确和能量守恒定律的建立才逐渐清楚。能量一词就是它们的总称，而能量的计量则通过“卡、焦耳”等新单位的出现而得到解决。

然而，关于文字、数字、图画、声音的知识已有几千年历史了。但是它们的总称是什么，它们如何统一地计量，直到 19 世纪末还没有被正确地提出来，更谈不上如何去解决了。20 世纪初期，随着电报、电话、照片、电视、无线电、雷达等的发展，如何计量信号中信息量的问题被隐约地提上日程。

1928 年，哈特利(R. V. H. Harley)考虑到从 D 个彼此不同的符号中取出 N 个符号并且组成一个“词”的问题。如果各个符号出现的概率相同，而且是完全随机选取的，就可以得到 DN 个不同的词。从这些词里取了特定的一个就对应一个信息量 I。哈特利建议用 $N\log D$ 这个量表示信息量，即 $I=N\log D$。这里的 log 表示以 10 为底的对数。后来，1949 年控制论的创始人维纳也研究了度量信息的问题，还把它引向热力学第二定律。

但是就信息传输给出基本数学模型的核心人物还是香农。1948 年，香农长达数十页的论文“通信的数学理论”成了信息论正式诞生的里程碑。在他的通信数学模型中，清楚地提出信息的度量问题，他把哈特利的公式扩大到概率 pi 不同的情况，得到了著名的计算信息熵 H 的公式：

$$H=\sum -pi\log pi$$

如果计算中的对数 log 是以 2 为底的，那么计算出来的信息熵就以比特(bit)为单位。今天在计算机和通信中广泛使用的字节(Byte)、KB、MB、GB 等都是从比特演化而来。“比特”的出现标志着人类知道了如何计量信息量。香农的信息论为明确什么是信息量概念做出决定性的贡献。

香农在进行信息的定量计算的时候，明确地把信息量定义为随机不定性程度的减少。这就表明了他对信息的理解：信息是用来减少随机不定性的东西。或香农逆定义：信息是确定性的增加。

虽然仙农的信息概念比以往的认识有了巨大的进步，但仍存在局限性，这一概念同样没有包含信息的内容和价值，只考虑了随机型的不定性，没有从根本上回答“信息是什么”的问题。

事实上，香农最初的动机是把电话中的噪音除掉，他给出通信速率的上限，这个结论首先用在电话上，后来用到光纤上，现在又用在无线通信上。我们今天能够清晰地打越洋电话或卫星电话，都与通信信道质量的改善密切相关。

克劳德·香农在公众中并不特别知名，但他是使我们的世界能进行即时通信的少数科学家和思想家之一。他是美国科学院院士、美国工程院院士、英国皇家学会会员、美国哲学学会会员。他获得过许多荣誉和奖励。例如 1949 年 Morris 奖、1955 年 Ballantine 奖、1962 年 Kelly 奖、1966 年的国家科学奖章、IEEE 的荣誉奖章、1978 年 Jaquard 奖、1983 年 Fritz 奖、1985 年基础科学京都奖。他接受的荣誉学位不胜枚举，不再赘述。

今天，我们怀念香农，要熟悉他的两大贡献：一是信息理论、信息熵的概念；二是符号逻辑和开关理论。我们更应该学习他好奇心强、重视实践、追求完美、永不满足的科学精神，这是他获得成功的重要经验。

习 题

1.1 从信息论的观点简述典型数字通信系统各部分的功能。

1.2 查阅资料，简述信息论的发展现状和未来发展趋势。

第 2 章

单符号离散信源

最简通信系统由信源、信道和信宿 3 部分构成。离散信源可用离散随机矢量来描述，其中每个随机变量都是取值离散的。当随机矢量只含一个离散随机变量时，这种信源就称为单符号离散信源。

2.1 单符号离散信源的数学模型

如果信源 X 符号集为 $A=\{a_1, a_2, \cdots, a_r\}$，$r$ 为符号集大小，信源符号对应某一概率分布，$\{p(a_i), i=1, 2, \cdots, r\}$，称此信源为单符号离散信源，信源空间数学模型可由式(2.1)描述。

$$\begin{bmatrix} X \\ P \end{bmatrix}=\begin{bmatrix} a_1 & \cdots & a_n \\ p(a_1) & \cdots & p(a_n) \end{bmatrix} \tag{2.1}$$

其中，$p(a_i) \geqslant 0$，$\sum_{i=1}^{n} a_i = 1$

【例 2.1】一个三元无记忆信源 X，符号集 $A=\{0, 1, 2\}$，概率分布为 $p(0)=1/2$，$p(1)=1/3$，$p(2)=1/6$，写出信源空间模型。

解：信源空间模型如下式所示：

$$\begin{bmatrix} X \\ P \end{bmatrix}=\begin{bmatrix} 0, & 1, & 2 \\ 1/2, & 1/3, & 1/6 \end{bmatrix}$$

【例 2.2】某箱子中共有 32 个质地均匀的球，其中，红球 16 个，黄球 8 个，蓝球和白球各 4 个。每次拿出一个球，拿出后又放回，各种彩球出现的先验概率分别为

$$P(\text{红})=16/32=1/2$$
$$P(\text{黄})=8/32=1/4$$
$$P(\text{蓝})=4/32=1/8$$
$$P(\text{白})=4/32=1/8$$

用随机变量 X 表示这个信源，其信源空间为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} \text{红} & \text{黄} & \text{蓝} & \text{白} \\ 1/2 & 1/4 & 1/8 & 1/8 \end{bmatrix}$$

由上述两个例子看出：信源空间包含两个基本要素：随机变量 X 的状态空间和概率空间，而概率空间又是决定性要素。概率可测是香农信息论的基本前提。

2.2 信源符号的自信息量

设信源发出某符号 a_i，由于信道中存在噪声或其他干扰，收端收到的是 a_i 的某种变型 b_j，收信者收到 b_j 后，从 b_j 中获取关于 a_i 的信息量，用 $I(a_i;b_j)$ 表示，则

$$\begin{aligned} I(a_i;b_j) &= [\text{收到 } b_j \text{ 前，收信者对信源发 } a_i \text{ 的不确定性}] \\ &\quad -[\text{收到 } b_j \text{ 后，收信者对信源发 } a_i \text{ 仍然存在的不确定性}] \\ &= \text{收信者收到 } b_j \text{ 前、后，对信源发 } a_i \text{ 的不确定性的消除} \end{aligned} \tag{2.2}$$

若信道是无噪无损信道，即一一对应信道，收信者收到 b_j 就是 a_i 本身，那么就完全消除了对信源符号 a_i 的不确定性，即

$$I(a_i;a_i) = [\text{收到 } a_i \text{ 前，收信者对信源发 } a_i \text{ 的不确定性}] \tag{2.3}$$

把 $I(a_i;a_i)$ 定义为信源符号 a_i 的自信息量，并用 $I(a_i)$ 表示，即

$$I(a_i) = [\text{收到 } a_i \text{ 前，收信者对信源发出 } a_i \text{ 的不确定性}] \tag{2.4}$$

信源符号的自信息量度量问题，相当于信源发符号的不确定性度量问题，是先验概率的函数：

$$I(a_i) = f[p(a_i)] \tag{2.5}$$

根据自信息量定义所需的 4 个公理性条件，自信息量函数 $I(a_i)$ 定义如下：

$$I(a_i) = \log \frac{1}{p(a_i)} = -\log p(a_i) \tag{2.6}$$

自信息量 $I(a_i)$ 随 $p(a_i)$ 变化的曲线如图 2.1 所示。

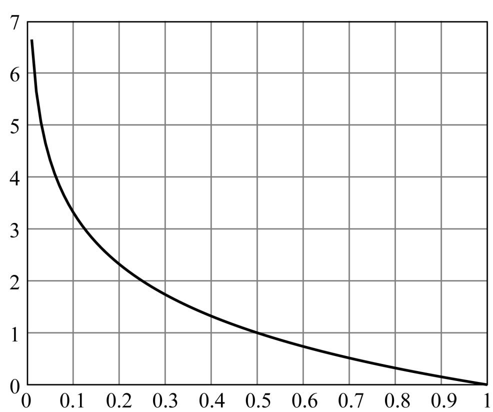


图 2.1　自信息量 $I(a_i)$ 随 $p(a_i)$ 变化的曲线

自信息量的单位与对数的底有关

若以“2”为底，自信息量单位为“比特”(bit-binary unit)，即

$$I(a_i)=\log_2\frac{1}{p(a_i)}=-\log_2 p(a_i)\quad(\text{比特})$$

若以“e”为底，自信息量单位为“奈特”(nat-nature unit)，即

$$I(a_i)=\ln\frac{1}{p(a_i)}=-\ln p(a_i)\quad(\text{奈特})$$

若以“10”为底，自信息量单位为“哈特”(Hart-Hartley)，即

$$I(a_i)=\log_{10}\frac{1}{p(a_i)}=-\log_{10} p(a_i)\quad(\text{哈特})$$

若以正整数“r”为底，自信息量单位为“r 进制信息单位”，即

$$I(a_i)=\log_r\frac{1}{p(a_i)}=-\log_r p(a_i)\quad(r\text{ 进制信息单位})$$

不同信息单位之间可以进行换算。本书如不加说明，信息度量默认为以“2”为底，且略去底数“2”不写。

香农信息量的度量，是信息论发展史上的里程碑，奠定了信息论发展成为一门学科的基础。

【例 2.3】箱中 80 个红球，20 个白球，现从箱中随机取出两个球。求：

(1) 求解“两个球中红、白各一个”的不确定性；

(2) 求解“两个球都是白球”所提供的信息量；

(3) 求解“两个球都是白球”和“两个球都是红球”相比较，哪个事件发生的更难测？

解：设 x 为红球数，y 为白球数，则

$$P_{XY}(1,1)=\frac{C_{80}^1C_{20}^1}{C_{100}^2}=\frac{32}{99},\quad I(1,1)=0.490(\text{bit})$$

$$P_{XY}(0,2)=\frac{C_{20}^2}{C_{100}^2}=\frac{19}{495},\quad I(0,2)=1.416(\text{bit})$$

$$P_{XY}(2,0)=\frac{C_{80}^2}{C_{100}^2}=\frac{316}{495},\quad I(2,0)=0.195(\text{bit})$$

从上述计算结果来看，1.416bit 大于 0.195bit，显然，事件“两个球都是白球”要比事件“两个球都是红球”更难测。

【例 2.4】设二元随机矢量 $X^N=(X_1,X_2,\cdots,X_N)$，矢量中每个变量 $X_i(i=1,2,\cdots,N)$ 为独立同分布随机变量，且 0 符号的概率 $\omega(0\leqslant\omega\leqslant1)$。求序列 x＝101100100 的自信息量。

解：$I(x)=-\log p(x)=-\log\omega^4(1-\omega)^5=-4\log\omega-5\log(1-\omega)$

2.3　信源的信息熵

自信息量 $I(a_i)$ 是针对具体的特定事件 a_i 在发生前存在的不确定度的度量，或发生后，提供给接收者的信息量。自信量的度量具有个体差异性，不具备对信源总体不确定度

的度量能力。对信源 X 总体进行信息度量，即信源熵，它表征信源平均每个符号所含有的不确定性或平均输出一个符号提供给观测者的信息量，是自信息量的统计均值，即有

$$\begin{aligned}H(X) &= E[I(a_i)] \\ &= p(a_1)I(a_1) + p(a_2)I(a_2) + \cdots + p(a_r)I(a_r) \\ &= -p(a_1)\log p(a_1) - p(a_2)\log p(a_2) - \cdots - p(a_r)\log p(a_r) \\ &= -\sum_{i=1}^{r} p(a_i)\log p(a_i) \quad (\text{比特/信源符号})\end{aligned} \tag{2.7}$$

熵的单位即自信息量的单位，取决于对数的底。

【例 2.5】(a) 设有一信源 X 由两个事件 a_1，a_2 组成，其概率空间如下

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} a_1, & a_2 \\ 0.99, & 0.01 \end{bmatrix}$$

则其信源熵为

$$H(X) = -0.99\times\log 0.99 - 0.01\times\log 0.01 = 0.08 \text{ bit/sig}$$

(b) 设有另一信源 Y 如下

$$\begin{bmatrix} X \\ P(y) \end{bmatrix} = \begin{bmatrix} b_1, & b_2 \\ 0.5, & 0.5 \end{bmatrix}$$

则

$$H(Y) = -0.5\times\log 0.5 - 0.5\times\log 0.5 = 1 \text{ bit/sig}$$

可见有 $H(Y)>H(X)$，即信源 Y 的平均不确定性要大于信源 X 的平均不确定性。直观地分析也易得出这样一个结论：信源中各事件的出现概率越接近，则事先猜测某一事件发生的把握性越小，即不确定性越大。从后述信息熵的极值性质，我们可知：当信源各事件等概率出现时具有最大的信源熵，也即信源的平均不确定性最大。

【例 2.6】设二元信源 X 的信源空间为$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$，其中，$0\leqslant p\leqslant 1$。

(1) 试计算当 $p=1/2$；$p=0$(或 $p=1$)时信息熵 $H(X)$ 的值；

(2) 信息熵 $H(X)$ 是概率分量 p 的函数 $H(p)$。画出 $H(p)$ 的函数曲线，并指明函数 $H(p)$ 的特性。

解：信源 X 的信息熵

$$H(X) = -p\times\log p - (1-p)\times\log(1-p) = H(p)$$

这表明，二元信源 X 的信息熵 $H(X)$ 是概率分量 p 的函数。

(1) 当 $p=(1-p)=1/2$ 时，二元信源 X 称为二维等概信源，其信息熵

$$H(X) = H(p) = H\left(\frac{1}{2}\right) = -\frac{1}{2}\times\log\frac{1}{2} - \left(\frac{1}{2}\right)\times\log\left(\frac{1}{2}\right) = 1 \quad (\text{比特/信源符号})$$

这表明，二元等概信源 X，每一个符号(不论是“0”还是“1”)含有的平均信息量是 1 比特，每发一个符号(不论是“0”还是“1”)提供的平均信息量是 1 比特；每一个符号(不论是“0”还是“1”)存在的平均不确定性是 1 比特；收信者能确切无误的收到信源 X 的“0”或“1”时，收信者每收到信源 X 的一个符号，获取的平均信息量是 1 比特。

（2）当 $p=0(p=1)$ 时，二元信源 X 是一个确知信源，其信息熵

$$H(X)=H(p)=-p\log p-(1-p)\log(1-p)=-0\log 0-1\log 1$$

因为，对任意小的正数 $\varepsilon>0$，以“e”为底的对数 $\ln\varepsilon$ 可展开成级数

$$\ln\varepsilon=2\cdot\left\{\frac{\varepsilon-1}{\varepsilon+1}+\frac{1}{3}\left(\frac{\varepsilon-1}{\varepsilon+1}\right)^3+\frac{1}{5}\left(\frac{\varepsilon-1}{\varepsilon+1}\right)^5+\cdots\right\}$$

当 $\varepsilon\to 0$ 时，$\varepsilon\ln\varepsilon$ 的极限值等于零，即有

$$\lim_{\varepsilon\to 0}\{\varepsilon\ln\varepsilon\}=0$$

所以

$$\lim_{\varepsilon\to 0}\{\varepsilon\log\varepsilon\}=\lim_{\varepsilon\to 0}\left\{\varepsilon\frac{\ln\varepsilon}{\ln 2}\right\}=\frac{1}{\ln 2}\lim_{\varepsilon\to 0}\{\varepsilon\ln\varepsilon\}=0$$

这样，可合理的约定

$$0\log 0=0$$

则当 $p=0(p=1)$ 时，信源 X 的信息熵

$$H(X)=H(0)=H(1)=0$$

这表明，以概率 1 发符号“0”或“1”的二元确知信源，不存在任何不确定性，不含有任何信息量。

（3）当 $0<p<1$ 时，取不同的 p 值，可计算出二元信源 X 的信息熵

$$H(p)=-p\log p-(1-p)\log(1-p)$$

相应地，得到 $H(p)$ 函数曲线如图 2.2 所示

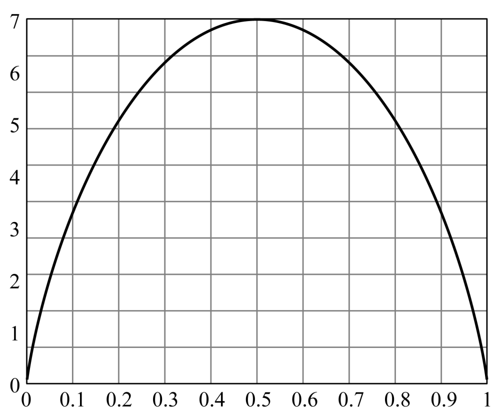


图 2.2　$H(p)$ 函数曲线

从图中可见，若二元信源 X 发符号“0”（或“1”）是确定事件，即 $p=1(p=0)$ 时，则 $H(p)=H(1)$（或 $H(0)$）$=0$。这意味着确知信源 X 不提供任何信息量；若二元信源 X 以相同概率 0.5 发符号“0”或“1”，即 $p=1/2$ 时，$H(p)=1$ 比特/信源符号。这意味着等概信源 X 每发一个符号提供的平均信息量达到最大值 1 比特/信源符号。若二元信源 X 发“0”（或“1”）的概率 $p(0<p<1)$ 越接近 1/2，则发“1”（或“0”）的概率 $(1-p)$ 亦越接近 1/2。若“0”和“1”的概率越接近则 $H(p)$ 的值越大。这意味着二元信源 X 的平均不确定性越大，每符号（不论是“0”还是“1”）含有的平均信息量越大；相反，二元信源 X 发“0”（或“1”）的

概率 $p(0<p<1)$ 越远离 1/2，则发“1”(或“0”)的概率 $(1-p)$ 亦越远离 1/2. 若发“0”和“1”的概率相差越大，则 $H(p)$ 的值就越小。这意味着二元信源 X 的平均不确定性越小，每符号(不论是“0”还是“1”)含有的平均信息量就越小。这体现出 $H(p)$ 是概率分量 p 的上凸形函数的基本特性。

【例 2.7】在一个箱子中装有 m 个黑球和 $(n-m)$ 个白球。设实验 X：随机的从箱子中取出一个球而不再放回箱子；实验 Y：从箱子中取出第二个球。

(1) 计算实验 X 所获取的平均信息量。

(2) 若实验 X 摸取的第一个球的颜色不知道，计算实验 Y 所获取的平均信息量。

解：令 W 代表白球；B 代表黑球。

(1) 设实验 X 中取出的球是黑球的概率为 $P_X(B)$；实验 X 中取出的球是白球的概率为 $P_X(W)$，则有

$$P_X(B)=\frac{m}{n}$$

$$P_X(W)=\frac{n-m}{n}$$

信源 X 的信源空间可表示为

$$\begin{bmatrix} X \\ P \end{bmatrix}=\begin{bmatrix} B & W \\ \dfrac{m}{n} & \dfrac{n-m}{n} \end{bmatrix}$$

实验 X 所获取的平均信息量，就是信源 X 的信息熵

$$H(X)=-\frac{m}{n}\log\frac{m}{n}-\left(\frac{n-m}{n}\right)\log\left(\frac{n-m}{n}\right)$$

(2) 若实验 X 中取出的第一个球是白球，令实验 Y 中取出的白球的概率为 $P_Y(W/W)$；取出黑球的概率为 $P_Y(B/W)$，则有

$$P_Y(W/W)=\frac{n-m-1}{n-1}$$

$$P_Y(B/W)=\frac{m}{n-1}$$

若实验 X 中取出的第一个球是黑球，令实验 Y 中取出白球的概率为 $P_Y(W/B)$；取出黑球的概率为 $P_Y(B/B)$，则有

$$P_Y(W/B)=\frac{n-m}{n-1}$$

$$P_Y(B/B)=\frac{m-1}{n-1}$$

设实验 Y 中出现白球的概率为 $P_Y(W)$；出现黑球的概率为 $P_Y(B)$，则有

$$\begin{aligned} P_Y(W) &= P_X(W)P_Y(W/W)+P_X(B)P_Y(W/B) \\ &= \frac{(n-m)}{n}\cdot\frac{(n-m-1)}{n-1}+\frac{m}{n}\cdot\frac{(n-m)}{n-1}=\frac{n-m}{n}=P_X(W) \end{aligned}$$

$$P_Y(B)=P_X(W)P_Y(B/W)+P_X(B)P_Y(B/B)$$
$$=\frac{(n-m)}{n}\cdot\frac{m}{(n-1)}+\frac{m}{n}\cdot\frac{(m-1)}{(n-1)}=\frac{m}{n}=P_X(B)$$

信源 Y 的信源空间可表示为

$$\begin{bmatrix}Y\\P\end{bmatrix}=\begin{bmatrix}B & W\\ \frac{m}{n} & \frac{n-m}{n}\end{bmatrix}$$

实验 Y 所获取的平均信息量，就是信源 Y 的信息熵

$$H(Y)=-\frac{m}{n}\log\frac{m}{n}-\left(\frac{n-m}{n}\right)\log\left(\frac{n-m}{n}\right)=H(X)$$

这说明，实验 X 和实验 Y 的平均不确定性或获取的平均信息量是相等的。

2.4　信息熵的性质

在介绍熵的性质之前，给出几个在信息论证明中有用的几个定义和定理。并将单个随机变量的熵推广到两个或者多个变量熵的情形。

2.4.1　联合熵与条件熵

定义 2.4.1　对于服从联合分布为 $p(x, y)$ 的一对离散随机变量 (X, Y)，其联合熵 $H(X, Y)$ (joint entropy) 定义为

$$H(X, Y)=-\sum_{x\in X}\sum_{y\in Y}p(x, y)\log p(x, y) \tag{2.8}$$

上式亦可表示为 $$H(X, Y)=-E\log p(X, Y) \tag{2.9}$$

也可以定义一个随机变量在给定另一随机变量下的条件熵，先求条件分布熵 $H(Y|X=x)$；然后对条件分布熵再求统计均值既得条件熵。

定义 2.4.2　若 $(X, Y)\sim p(x, y)$，条件熵 (conditional entropy) $H(Y/X)$ 定义为

$$\begin{aligned}H(Y\backslash X) &= \sum_{x\in X}p(x)H(Y|X=x)\\ &=-\sum_{x\in X}p(x)\sum_{y\in Y}p(y|x)\log p(y|x)\\ &=-\sum_{x\in X}\sum_{y\in Y}p(x, y)\log p(y|x)\\ &=-E\log p(Y|X)\end{aligned} \tag{2.10}$$

联合熵和条件熵定义的这种自然性可由一个事实得到体现，它就是一对随机变量的熵等于其中一个随机变量的熵加上另一个随机变量的条件熵。其证明见如下定理。

定理 2.1(链式法则)

$$H(X, Y)=H(X)+H(Y|X) \tag{2.11}$$

【证明】

$$H(X, Y)=-\sum_{x\in X}\sum_{y\in Y}p(x, y)\log p(x, y)$$

$$
\begin{aligned}
&= -\sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(x)p(y|x) \\
&= -\sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(x) - \sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(y|x) \\
&= -\sum_{x\in X} p(x)\log p(x) - \sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(y|x) \\
&= H(X) + H(Y|X) \qquad (2.12)
\end{aligned}
$$

等价地记为

$$\log p(X, Y) = \log p(X) + \log p(Y|X) \qquad (2.13)$$

等式的两边同时取对数期望，即得本定理。

推论

$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z) \qquad (2.14)$$

2.4.2 相对熵

若 P 和 Q 为定义在同一概率空间的两个概率测度，定义 P 相对于 Q 的相对熵为

$$D(P\|Q) = \sum_{x} P(x)\log\frac{P(x)}{Q(x)} \qquad (2.15)$$

相对熵又称散度、鉴别信息、方向散度、交叉熵、Kullback_leibler 距离等。注意，在式(2.15)中，概率分布的维数不限，可以是一维，也可以是多维，也可以是条件概率。

在证明下面的定理前，首先介绍一个在信息论中常用的不等式。

对于任意正实数 x，下面不等式成立：

$$1 - 1/x \leqslant \ln x \leqslant x - 1 \qquad (2.16)$$

实际上，设 $f(x) = \ln x - x + 1$，可求得函数的稳定点为 $x = 1$，并可求得在该点的二阶导数小于 0，从而可得 $x = 1$ 为 $f(x)$ 取极大值的点，即 $f(x) = \ln x - x + 1 \leqslant 0$，仅当 $x = 1$ 时，式(2.16)右边等号成立。令 $y = 1/x$，可得 $1 - 1/y \leqslant \ln y$，再将 y 换成 x，就得到左边的不等式。

定理 2.2 如果在一个共同有限字母表概率空间上给定两个概率测度 $P(x)$ 和 $Q(x)$，那么

$$D(P\|Q) \geqslant 0 \qquad (2.17)$$

仅当对所有 x，$P(x) = Q(x)$ 时，等式成立。

【证明】 因 $P(x)$，$Q(x) \geqslant 0$，$\sum_x P(x) = \sum_x Q(x) = 1$，所以根据式(2.16)，有

$$
\begin{aligned}
-D(P\|Q) &= \sum_x P(x)\log\frac{Q(x)}{P(x)} \leqslant \sum_x P(x)(\log e)\left[\frac{Q(x)}{P(x)} - 1\right] \\
&= (\log e)\left[\sum_x Q(x) - \sum_x P(x)\right] = 0 \qquad (2.18)
\end{aligned}
$$

仅当对所有 x，$P(x) = Q(x)$ 时，等式成立。

式(2.17)称为散度不等式(divergence inequality)，该式说明，一个概率测度相对于另一个概率测度的散度是非负的，仅当两测度相同时，散度为零。散度可以解释为两个概率

测度之间的“距离”，即两概率测度不同程度的度量。不过，散度并不是通常意义下的距离，因为它不满足对称性，也不满足三角不等式。

【例 2.8】设一个二元信源的符号集为{0，1}，有两个概率分布 p 和 q，并且 $p(0)=1-r$，$p(1)=r$，$q(0)=1-s$，$q(1)=s$，求散度 $D(p\|q)$ 和 $D(q\|p)$，并分别求当 $r=s$ 和 $r=2s=1/2$ 时散度的值。

解：根据式(2.15)，得

$$D(p\|q)=(1-r)\log\frac{1-r}{1-s}+r\log\frac{r}{s}$$

$$D(q\|p)=(1-s)\log\frac{1-s}{1-r}+s\log\frac{s}{r}$$

当 $r=s$ 时，有 $D(p\|q)=D(q\|p)=0$

当 $r=2s=1/2$ 时，有

$$D(p\|q)=(1-1/2)\log\frac{1-1/2}{1-1/4}+1/2\log\frac{1/2}{1/4}=1-(\log 3)/2=0.2075\text{bit};$$

$$D(q\|p)=(1-1/4)\log\frac{1-1/4}{1-1/2}+1/4\log\frac{1/4}{1/2}=\frac{3}{4}(\log 3)-1=0.1887\text{bit}。$$

定理 2.3(熵的不增原理)

$$H(Y|X)\leqslant H(X) \tag{2.19}$$

【证明】设 $p(y)=\sum_x p(x)p(y/x)$，那么

$$\begin{aligned}H(Y)-H(Y/X)&=-\sum_y p(y)\log p(y)+\sum_x\sum_y p(x)p(y/x)\log p(y/x)\\&=-\sum_y\sum_x p(x)p(y/x)\log p(y)+\sum_x\sum_y p(x)p(y/x)\log p(y/x)\\&=\sum_x p(x)\sum_y p(y/x)\log\frac{p(y/x)}{p(y)}=\sum_x p(x)D(p(y/x)\|p(y))\geqslant 0\end{aligned}$$

上面利用了散度不等式，仅当 X，Y 相互独立时，等式成立。

式(2.19)表明，条件熵总是不大于无条件熵，这就是熵的不增原理：在信息处理过程中，已知条件越多，结果的不确定性越小，也就是熵越小。

2.4.3 Jensen 不等式及其结果

定义 2.4.3 若对于任意的 x_1，$x_2\in(a, b)$及 $0\leqslant\lambda\leqslant 1$，满足

$$f(\lambda x_1+(1-\lambda)x_2)\leqslant\lambda f(x_1)+(1-\lambda)f(x_2) \tag{2.20}$$

则称函数 $f(x)$在区间(a, b)上是凸的(convex)。如果仅当 $\lambda=0$ 或 $\lambda=1$，上式成立，则称函数 f 是严格凸的(strictly convex)。

定义 2.4.4 如果$-f$为凸函数，则称函数 f 是凹的。如果函数总是位于任何一条弦的下面，则该函数是凸的；如果函数总是位于任何一条弦的上面，则该函数是凹的。

定理 2.4(Jensen 不等式) 若给定凸函数 f 和一个随机变量 X，则

$$Ef(X) \geqslant f(EX) \tag{2.21}$$

进一步，若 f 是严格凸的，那么式(2.21)中的等式蕴含 $X=EX$ 的概率为 1(即 X 是个常量)。

【证明】我们只证明离散分布情形，且对分布点的个数进行归纳证明。当 f 为严格凸函数时，等号成立条件的证明留给读者。对于两点分布，不等式变为

$$p_1 f(x_1)+p_2 f(x_2) \geqslant f(p_1 x_1+p_2 x_2) \tag{2.22}$$

这由凸函数的定义可直接得到。假定当分布点的个数为 $k-1$ 时，定理成立，此时记 $p_i'=p_i/(1-p_k)(i=1, 2, \cdots, k-1)$，则有

$$\begin{aligned}\sum_{i=1}^{k} p_i f(x_i) &= p_k f(x_k) + (1-p_k)\sum_{i=1}^{k-1} p'_i f(x_i)\\ &\geqslant p_k f(x_k) + (1-p_k) f\Big(\sum_{i=1}^{k-1} p'_i x_i\Big)\\ &\geqslant f\Big(p_k x_k + (1-p_k)\sum_{i=1}^{k-1} p'_i x_i\Big)\\ &= f\Big(\sum_{i=1}^{k} p_i x_i\Big)\end{aligned} \tag{2.23}$$

2.4.4 信息熵的基本性质

1. 对称性

概率矢量 $p=(p_1, p_2, \cdots, p_n)$中，各分量的次序任意改变，熵不变，即

$$H(p_1, p_2, \cdots, p_n)=H(p_{j_1}, p_{j_2}, \cdots, p_{j_n}) \tag{2.24}$$

其中，j_1，j_2，$\cdots$，j_n 是 1，2，$\cdots$，n 的任何一种 n 级排列。该性质说明熵仅与随机变量总体概率特性(即概率分布)有关，而与随机变量的取值及符号排列顺序无关。

2. 非负性

$$H(p)=H(p_1, p_2, \cdots, p_n) \geqslant 0 \tag{2.25}$$

仅当对某个 $p_i=1$ 时，等式成立。

因为自信息量是非负的，熵为自信息的平均，所以也是非负的。不过，非负性仅对离散熵有效，而对连续熵来说这一性质并不成立。

3. 确定性

$$H(1, 0)=H(1, 0, 0)=H(1, 0, \cdots, 0)=0 \tag{2.26}$$

这就是说，当随机变量集合中任一事件概率为 1 时，熵就为 0。这个性质意味着，从总体来看，事件集合中虽含有许多事件，但是如果只有一个事件几乎必然出现，而其他事件几乎都不出现，那么，这就是一个确知的变量，其不确定性为 0。

4. 扩展性

$$\lim_{\varepsilon\to 0} H_{n+1}(p_1, p_2, \cdots, p_n-\varepsilon, \varepsilon)=H_n(p_1, p_2, \cdots, p_n) \tag{2.27}$$

利用$\lim\limits_{\varepsilon \to 0} \varepsilon \log \varepsilon = 0$可得到上面的结果，其含义是，虽然小概率事件自信息量大，但在计算熵时所占比重很小，可以忽略。

5. 极值性

定理 2.5（离散最大熵定理）　对于有限离散随机变量，当符号集中的符号等概率发生时，熵达到最大值。

【证明】设随机变量有 n 个符号，概率分布为 $P(x)$；$Q(x)$ 为等概率分布，即 $Q(x)=1/n$。根据散度不等式有

$$
\begin{aligned}
D(P \| Q) &= \sum_x P(x) \log \frac{P(x)}{Q(x)} \\
&= \sum_x P(x) \log P(x) - \sum_x P(x) \log(1/n) \\
&= -H(X) + \log n \geqslant 0
\end{aligned} \tag{2.28}
$$

即，仅当 $P(x)$ 等概布时等号成立。

注意：离散最大熵定理仅适用于有限离散随机变量，对于无限可数符号集，只有附加其他约束求最大熵才有意义。

6. 上凸性

$H(p)=H(p_1, p_2, \cdots, p_n)$是概率矢量 p 的严格的上凸函数。

这就是说，若 $p=\theta p_1+(1-\theta)p_2$，那么 $H(p)>\theta H(p_1)+(1-\theta)H(p_2)$，其中 p，p_1，p_2 均为 n 维概率矢量，$0 \leqslant \theta \leqslant 1$。该性质可用凸函数性质(1)来证明(提示：先证明$-p_i \log p_i$ 是严格上凸的)。

7. 一一对应变换下的不变性

离散随机变量的变换包含两种含义，一是符号集中符号到符号的映射，二是符号序列到序列的变换。首先研究第一种情况。设两随机变量 X，Y，符号集分别为 A，B，其中 Y 是 X 的映射，可以表示为 $A \to B$，$x \to f(x)$。因此有

$$
p(y/x)=\begin{cases} 1 & y=f(x) \\ 0 & y \neq f(x) \end{cases} \tag{2.29}
$$

所以 $H(Y/X)=0$；$H(XY)=H(X)+H(Y/X)=H(X)$，而另一方面 $H(XY)=H(Y)+H(X/Y) \geqslant H(Y)$，所以 $H(X) \geqslant H(Y)$，仅当 f 是一一对应映射时等号成立，此时 $H(X/Y)=0$。应用类似的论证也可推广到多维随机矢量的情况，因此得到如下定理。

定理 2.6　离散随机变量(或矢量)经符号映射后的熵不大于原来的熵，仅当一一对应映射时熵不变。

【例 2.9】设二维随机矢量 XY，其中 X，Y 为独立同分布随机变量，符号集为 $A=\{0, 1, 2\}$，对应的概率为$\{1/3, 1/3, 1/3\}$，做变换 $u=x+y$，$v=x-y$，得到二维随机矢量 UV；求 $H(U)$，$H(V)$，$H(UV)$。

解：U，V 取值空间如表 2.1、表 2.2 所示。

表 2.1　U 取值

$U=X+Y$		Y		
		0	1	2
X	0	0	1	2
	1	1	2	3
	2	2	3	4

U 的符号集为$\{0，1，2，3，4\}$

$$P_u(0)=\frac{1}{3}\times\frac{1}{3}=\frac{1}{9}$$

$$P_u(1)=\frac{1}{3}\times\frac{1}{3}\times2=\frac{2}{9}$$

$$P_u(2)=\frac{1}{3}\times\frac{1}{3}\times3=\frac{1}{3}$$

$$P_u(3)=\frac{1}{3}\times\frac{1}{3}\times2=\frac{2}{9}$$

$$P_u(4)=\frac{1}{3}\times\frac{1}{3}=\frac{1}{9}$$

$$H(U)=H\left(\frac{1}{9},\ \frac{2}{9},\ \frac{1}{3},\ \frac{2}{9},\ \frac{1}{9}\right)=2.1972\ \text{bit/符号}$$

表 2　V 取值

$V=X-Y$		Y		
		0	1	2
X	0	0	−1	−2
	1	1	0	−1
	2	2	1	0

V 的符号集为$\{-2，-1，0，1，2\}$

$$P_v(-2)=\frac{1}{3}\times\frac{1}{3}=\frac{1}{9}$$

$$P_v(-1)=\frac{1}{3}\times\frac{1}{3}\times2=\frac{2}{9}$$

$$P_v(0)=\frac{1}{3}\times\frac{1}{3}\times3=\frac{1}{3}$$

$$P_v(1)=\frac{1}{3}\times\frac{1}{3}\times2=\frac{2}{9}$$

$$P_v(2)=\frac{1}{3}\times\frac{1}{3}=\frac{1}{9}$$

$H(V)=H\left(\frac{1}{9}, \frac{2}{9}, \frac{1}{3}, \frac{2}{9}, \frac{1}{9}\right)=2.1972$ bit/符号

因为是一一对应变换，所以

$H(UV)=H(XY)=H(X)+H(Y)=2\log 3=3.1699$ bit/2 个符号

看到 $H(UV)<H(U)+H(V)$，所以 U，V 不独立。

本章要点

1. 单符号离散信源数学模型

$$\begin{bmatrix} X \\ P \end{bmatrix}=\begin{bmatrix} a_1 & \cdots & a_n \\ p(a_1) & \cdots & p(a_n) \end{bmatrix}$$

2. 定义　自信息量函数 $I(a_i)$

$$I(a_i)=\log\frac{1}{p(a_i)}=-\log p(a_i)$$

3. 定义　信源的信息熵

$$\begin{aligned} H(X) &= E[I(a_i)] \\ &= -\sum_{i=1}^{r} p(a_i)\log p(a_i) \quad (\text{比特 / 信源符号}) \end{aligned}$$

4. 定义　联合熵 $H(X, Y)$

$$H(X, Y)=-\sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(x, y)$$

5. 定义　条件熵

$$\begin{aligned} H(Y|X) &= \sum_{x\in X} p(x)H(Y|X=x) \\ &= -E\log p(Y|X) \end{aligned}$$

6. 链式法则

$$H(X, Y)=H(X)+H(Y|X)$$

7. 定义 P 相对于 Q 的相对熵

$$D(P\|Q)=\sum_{x} P(x)\log\frac{P(x)}{Q(x)}$$

8. Jensen 不等式

若给定凸函数 f 和一个随机变量 X，则　$Ef(X)\geqslant f(EX)$

9. 信息熵的性质

对称性　$H(p_1, p_2, \cdots, p_n)=H(p_{j_1}, p_{j_2}, \cdots, p_{j_n})$

非负性　$H(p)=H(p_1, p_2, \cdots, p_n)\geqslant 0$

确定性　$H(1, 0)=H(1, 0, 0)=H(1, 0, \cdots, 0)=0$

扩展性　$\lim\limits_{\varepsilon\to 0} H_{n+1}(p_1, p_2, \cdots, p_n-\varepsilon, \varepsilon)=H_n(p_1, p_2, \cdots, p_n)$

极值性　$H(X)\leqslant \log n$

上凸性　　$H(p)=H(p_1, p_2, \cdots, p_n)$是概率矢量 p 的严格上凸函数。

变换不变性　离散信源经过一一对应变换后熵不变

习　题

2.1　同时掷一对均匀的骰子，试求：

(1)“2 和 6 同时出现”这一事件的自信息量；

(2)“两个 5 同时出现”这一事件的自信息量；

(3)两个点数的各种组合的熵；

(4)两个点数之和的熵；

(5)“两个点数中至少有一个是 1”的自信息量。

2.2　某一无记忆信源的符号集为{0, 1}，已知 $p_0=\frac{1}{3}$，$p_1=\frac{2}{3}$。

(1)求符号的平均自信息量；

(2)由 1000 个符号构成的序列，求某一特定的序列(例如有 m 个“0”，$(1000-m)$个“1”)的自信息量的表达式；

(3)计算(2)中的序列的熵。

2.3　设信源 X 的信源空间为

$$[X \cdot P]:\begin{cases} X: & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ P(X): & 0.17 & 0.19 & 0.18 & 0.16 & 0.18 & 0.3 \end{cases}$$

求信源熵，并解释为什么 $H(X)>\log 6$，不满足信源熵的极值性。

2.4　每帧电视图像可以认为是由 3×10^5 个像素组成，所有像素均是独立变化，且每像素又取 128 个不同的亮度电平，并设亮度电平是等概出现。问每帧图像含有多少信息量？若现有一个广播员，在约 10 000 个汉字中选 1000 个字来口述这一电视图像，试问若要恰当地描述此图像，广播员在口述中至少需要多少个汉字？

2.5　设有一离散无记忆信源，其概率空间为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix}=\begin{bmatrix} a_1=0, & a_2=1, & a_3=2, & a_4=3 \\ \frac{3}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \end{bmatrix}$$

该信源发出的消息符号序列为{202　120　130　213　001　203　210　110　321　010　021　032　011　223　210}

试求：

(1)此消息的自信息是多少？

(2)在此消息中平均每个符号携带的信息量是多少？

第 3 章 单符号离散信道

信息论中有关信道问题的讨论，使用编码信道模型，即用信道输入输出符号转移概率或转移概率密度函数表征信道特性。

如果信道输入和输出的是随机过程，则对应的是波形信道；如果信道的输入和输出是随机矢量，而且矢量中每个随机变量的取值是连续的，则对应的信道是连续信道，每个随机变量的取值是离散的，则对应的信道是离散信道。离散信道中，若输入输出分别仅有一个随机变量，这种信道称为单符号离散信道。

本章将对单符号离散信道的信息传输、信道容量计算等一系列基本问题展开谈论。

3.1 信道的数学模型

单符号离散信道模型如图 3.1 所示。

$$X\left\{\begin{matrix} a_1 \\ a_2 \\ \cdots \\ a_r \end{matrix}\right. \to P(b_j/a_i) \to \left.\begin{matrix} b_1 \\ b_2 \\ \cdots \\ b_s \end{matrix}\right\}Y$$

图 3.1 单符号离散信道模型

输入变量 X 有 r 种取值，即输入符号集 X：$\{a_1, a_2, \cdots, a_r\}$，输出变量 Y 有 s 种取值，即输出符号集$\{b_1, b_2, \cdots, b_s\}$。信道转移概率

$$P(Y=b_j/X=a_i)=P(b_j/a_i)=P_{ij}$$

共有 $r\times s$ 个取值，体现了信道的符号传递特性。

写成矩阵形式，形成一个$(r\times s)$阶矩阵，矩阵行数代表信道输入符号个数，矩阵列数代表输出符号个数。

$$P=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \cdots & \cdots & \cdots & \cdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{bmatrix} \text{而且满足 } 0\leqslant p_{ij}\leqslant 1 \quad \begin{cases} i \text{ 是行的标号} \\ j \text{ 是列的标号} \end{cases}$$

$\sum_{j=1}^{s} p(b_j/a_i)=1(i=1, 2, \cdots, r)$，表明信道矩阵行和是1。信道特性也可以形象、直观地用信道转移图表示。

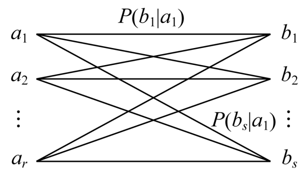


图 3.2　信道转移图

【例 3.1】二进制对称信道，简记为 BSC(Binary Semmetric Channel)，输入输出符号集分别为 X：{0，1}，Y：{0，1}，信道转移概率满足 $P_{Y/X}(0/0)=P_{Y/X}(1/1)=1-p$，$P_{Y/X}(1/0)=P_{Y/X}(0/1)=p$，$p$ 为错误传输概率。写出信道的转移概率矩阵，并画出转移概率图。

解：转移概率矩阵为 $p=\begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$，相应的信道转移概率矩阵如图 3.3 所示。

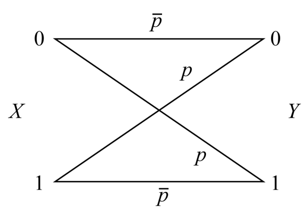


图 3.3　二进制对称信道转移概率图

【例 3.2】二进制删除信道，简记为 BEC(Binary Erasure Channel)其中输入集 X：{0，1}，输出集 Y：{0，2，1}，转移概率如图 3.4 所示。

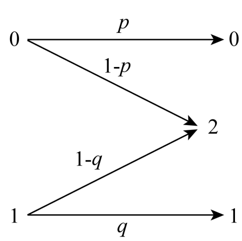


图 3.4　二进制删除信道转移概率图

写出信道的转移概率矩阵。

解：信道的转移概率矩阵为 $p=\begin{bmatrix} p & 1-p & 0 \\ 0 & 1-q & q \end{bmatrix}$

3.2　信道的交互信息量

在求解信道输入输出单个符号对应互信息量前，先讨论信道输入输出变量的统计特性。通信系统模型如图 3.5 所示。

$$X\left\{\begin{matrix} a_1 \\ a_2 \\ \cdots \\ a_r \end{matrix}\right. \to P(b_j/a_i) \to \left.\begin{matrix} b_1 \\ b_2 \\ \cdots \\ b_s \end{matrix}\right\}Y$$

图 3.5　系统模型

为方便讨论，首先定义几种概率的名称。

1. 先验概率

信源 X 输出符号 a_i 的概率 $p(x=a_i)=p(a_i)$　$(i=1，2，\cdots，r)$称为先验概率。

2. 正向转移概率

从信道输入符号 a_i 到信道输出符号 b_j 的条件概率

$$P(Y=b_j/X=a_i)=P(b_j/a_i)=P_{ij}\quad (i=1，2，\cdots，r；j=1，2，\cdots，s) \tag{3.1}$$

称为正向转移概率。

3. 后验概率

从信道输出符号 b_j 到输出符号 a_i 的条件概率

$$P(X=a_i/Y=b_j)=P(a_i/b_j)=P_{ji}\quad (i=1，2，\cdots，r；j=1，2，\cdots，s) \tag{3.2}$$

称为后验概率，又称之为反向转移概率。

利用 $P(X/Y)=\dfrac{P(XY)}{P(Y)}$，则有

$$p(a_i/b_j)=\frac{p(a_ib_j)}{p(b_j)}=\frac{p(a_i)p(b_j/a_i)}{\sum_{i=1}^{r}p(a_i)p(b_j/a_i)}\quad (i=1，2，\cdots，r；j=1，2，\cdots，s) \tag{3.3}$$

看到，已知先验概率和信道转移概率、后验概率即为确定。

4. 联合概率

信源符号 a_i 和信道输出符号 b_j 的联合概率

$$P(X=a_i；Y=b_j)=P(a_ib_j)=P(a_i)P(b_j/a_i)\quad (i=1，2，\cdots，r；j=1，2，\cdots，s) \tag{3.4}$$

由信源概率和信道转移概率唯一确定。

类似信道转移概率矩阵可以写出联合概率矩阵，也是 r 行 s 列矩阵。

$$[p_{XY}(a_ib_j)]=\begin{bmatrix} p(a_1b_1) & p(a_1b_2) & \cdots & p(a_1b_s) \\ p(a_2b_1) & p(a_2b_2) & \cdots & p(a_2b_s) \\ \cdots & \cdots & \cdots & \cdots \\ p(a_rb_1) & p(a_rb_2) & \cdots & p(a_rb_s) \end{bmatrix} \tag{3.5}$$

5. 信宿概率

信道输出符号概率 $p(Y=b_j)=p(b_j)$，$p(b_j)=\sum_{i=1}^{r}p(a_ib_j)=\sum_{i=1}^{r}p(a_i)p(b_j/a_i)$ 已知信源

符号概率和信道转移概率，即可求得信道输出符号概率。

现在转入互信息量的讨论。

在通信系统中，信源发出某符号 a_i，由于受到噪声的随机干扰，在信道的输出端输出符号 a_i 的某种变型 b_j，按信息的定义，信宿收到 b_j 后，从 b_j 中获取关于 a_i 的信息量 $I(a_i;b_j)$，等于信宿收到 b_j 前、后，对符号 a_i 的不确定性的消除，即有

[信宿收到 b_j 后，从 b_j 中获取关于 a_i 的信息量] $I(a_i;b_j)$

= [收到 b_j 前，收信者对信源发 a_i 的不确定性]

−[收到 b_j 后，收信者对信源发 a_i 仍然存在的不确定性] (3.6)

= 收信者收到 b_j 前、后，对信源发 a_i 的不确定性的消除

信宿收到 b_j 前，对信源发符号 a_i 的先验不确定性

$$I(a_i)=\log\frac{1}{p(a_i)}\quad(i=1,2,\cdots,r)\tag{3.7}$$

信宿收到 b_j 后，对信源发出的符号 a_i 的后验不确定性

$$I(a_i/b_j)=\log\frac{1}{p(a_i/b_j)}\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\tag{3.8}$$

则可得互信息为

$$\begin{aligned}I(a_i;b_j)&=I(a_i)-I(a_i/b_j)=\log\frac{1}{p(a_i)}-\log\frac{1}{p(a_i/b_j)}\\&=\log\frac{p(a_i/b_j)}{p(a_i)}\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\end{aligned}\tag{3.9}$$

信宿收到 b_j 后，从 b_j 中获取关于的信息量 $I(a_i;b_j)$ 称为输入符号 a_i 和输出符号 b_j 之间的交互信息量，简称为互信息。它表示信道在把输入符号 a_i 传递为输出符号 b_j 的过程中，信道所传递的信息量。(3.9)式称为符号 a_i 和 b_j 之间的互信息函数。

现在就互信息量表达式所代表的物理含义做进一步说明。

1. 当 $p(a_i/b_j)=1$ 时，有

$$I(a_i;b_j)=\log\frac{1}{p(a_i)}=I(a_i)\quad(i=1,2,\cdots,r)\tag{3.10}$$

$p(a_i/b_j)=1$ 表明收到的 $b_j(j=1,2,\cdots,s)$ 后，即可确切无误地判断发端符号为 a_i，消除对 a_i 的全部不确定性，收端获得关于 a_i 的全部信息量 $I(a_i)$ $(i=1,2,\cdots,r)$。

2. 当 $p(a_i)<p(a_i/b_j)<1$ 时，有

$$I(a_i;b_j)=\log\frac{p(a_i/b_j)}{p(a_i)}>0\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\tag{3.11}$$

这就意味着，收信者收到 b_j 后，判断信源发 $a_i(i=1,2,\cdots,r)$ 的可能性，比对于收到 $b_j(j=1,2,\cdots,s)$ 前判断信源发 $a_i(i=1,2,\cdots,r)$ 的可能性更大；也就是说收到 $b_j(j=1,2,\cdots,s)$ 后对信源发 $a_i(i=1,2,\cdots,r)$ 的不确定性，比收到 $b_j(j=1,2,\cdots,s)$ 前有所减小，收信者从 $b_j(j=1,2,\cdots,s)$ 中就可获取关于 $a_i(i=1,2,\cdots,r)$ 的一定信息量。

3. 当 $p(a_i/b_j)=p(a_i)$ 时，有

$$I(a_i;b_j)=\log\frac{p(a_i/b_j)}{p(a_i)}=\log 1\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\tag{3.12}$$

由 $p(a_ib_j)=p(a_i)p(b_j/a_i)=p(a_i)p(b_j)$ 得到，符号 a_i 与符号 b_j 统计独立。

说明，收信者收到 $b_j(j=1,2,\cdots,s)$ 前、后，判断信源发 $a_i(i=1,2,\cdots,r)$ 的可能性大小没有任何变化，收信者在收到 $b_j(j=1,2,\cdots,s)$ 前、后，对判断信源发 $a_i(i=1,2,\cdots,r)$ 的不确定性没有任何减小，收信者从 $b_j(j=1,2,\cdots,s)$ 中得不到关于 $a_i(i=1,2,\cdots,r)$ 的任何信息量。

4. 当 $0<p(a_i/b_j)<p(a_i)$ 时，有

$$I(a_i;b_j)=\log\frac{p(a_i/b_j)}{p(a_i)}<0\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\tag{3.13}$$

式(3.13)表明，由于信道噪声的干扰，收到符号 $b_j(j=1,2,\cdots,s)$ 后，猜测符号 $a_i(i=1,2,\cdots,r)$ 的难度反而加大。

互信息量 $I(a_i;b_j)$ 的另外两种表达形式

1. 第一种表达形式

$$\begin{aligned}I(a_i;b_j)&=\log\frac{p(a_i/b_j)}{p(a_i)}=\log\frac{p(a_ib_j)}{p(a_i)p(b_j)}\\&=\log\frac{1}{p(a_i)p(b_j)}-\log\frac{1}{p(a_ib_j)}\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\end{aligned}\tag{3.14}$$

其含义是：通信前，条件 a_i 与 b_j 统计独立。即 $p(a_ib_j)=p(a_i)p(b_j)$，$\log\frac{1}{p(a_i)p(b_j)}$ 表达联合条件 $I(a_ib_j)$ 不确定性；通信后，条件 a_i 与条件 b_j 建立统计关联，$\log\frac{1}{p(a_ib_j)}$ 表达通信后，联合条件 $I(a_ib_j)$ 的确定性。二者差，同样表明，信宿收到 b_j 后，从 b_j 中获取关于 a_i 的信息量 $I(a_i;b_j)$ 等于通信前后不确定性的消除。

2. 第二种表达形式

$$\begin{aligned}I(a_i;b_j)&=\log\frac{p(a_ib_j)}{p(a_i)p(b_j)}=\log\frac{p(b_j/a_i)}{p(b_j)}\\&=\log\frac{1}{p(b_j)}-\log\frac{1}{p(b_j/a_i)}=I(b_j;a_i)\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\end{aligned}\tag{3.15}$$

$a_i\to b_j$，构成正向信道；

$b_j\to a_i$，构成反向信道。

说明在反向信道中，从 a_i 中获得关于 b_j 的信息量等于在正向信道中从 b_j 中获得关于 a_i 的信息量，即 $I(a_i;b_j)=I(b_j;a_i)$。

【例 3.3】二元删除信道，简记为 BEC(Binary Erasure Channel)：其中输入符号集 $A=$

$\{a_0, a_1\}=\{0, 1\}$，输出符号集 $B=\{b_0, b_1, b_2\}=\{0, 2, 1\}$，$p(a_0)=p(a_1)=0.5$，写出信道的转移概率矩阵及互信息量 $I(a_i; b_j)$。

解： 设信道的转移概率矩阵为

$$P=\begin{bmatrix} p & 1-p & 0 \\ 0 & 1-q & q \end{bmatrix}$$

联合概率为

$$P(a_ib_j)=\begin{bmatrix} \frac{1}{2}p & \frac{1}{2}(1-p) & 0 \\ 0 & \frac{1}{2}(1-q) & \frac{1}{2}q \end{bmatrix}$$

则 $P(b_j)$ 为

$$P(b_j)=\begin{bmatrix} \frac{1}{2}p & 1-\frac{1}{2}(p+q) & \frac{1}{2}q \end{bmatrix}$$

那么互信息量 $I(a_i; b_j)$ 用矩阵形式简写为

$$[I(a_i; b_j)]=\left[\log\frac{p(a_ib_j)}{p(a_i)p(b_j)}\right]=\begin{bmatrix} 2 & \dfrac{(1-p)}{1-\frac{1}{2}(p+q)} & 0 \\ 0 & \dfrac{(1-p)}{1-\frac{1}{2}(p+q)} & 2 \end{bmatrix}$$

【例 3.4】 4 个等概率消息，编程的码字为 $M_1=000$，$M_2=011$，$M_3=101$，$M_4=110$，通过如图 3.6 所示二元对称无记忆信道（$\varepsilon<0.5$）传输，求：(1) 事件“接收到第一个数字为 0”与发送 M_1 之间的互信息；(2) 当接收到第二个数字也为 0 时，关于 M_1 的附加信息；(3) 当接收到第三个数字也为 0 时，又增加了多少关于 M_1 的信息？

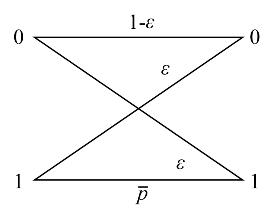


图 3.6　二元对称无记忆信道

解： 记“0”表示第一个接收数字为 0，“00”表示第一、二个接收数字都为 0，“000”表示前三个接收数字都为 0；$q(\cdot)$ 表示接收符号的概率；$p(y/x)$ 为信道的转移概率。

(1) $q(''0'')=\sum_{i=1}^{4}p(M_i)p(0/M_i)=\frac{1}{4}[2(1-\varepsilon)+2\varepsilon]=\frac{1}{2}$；

互信息：$\mathrm{I}(M_1;''0'')=\log\dfrac{p(''0''\mid M_1)}{q(''0'')}=\log\dfrac{1-\varepsilon}{1/2}=\log[2(1-\varepsilon)]$。

$$(2)\ q(''00'') = \sum_{i=1}^{4} p(M_i)p(00/M_i)$$

$$= \frac{1}{4}[p(0\,|\,0)p(0\,|\,0)+p(0\,|\,0)p(0\,|\,1)+p(0\,|\,1)p(0\,|\,0)+p(0\,|\,1)p(0\,|\,1)]$$

$$= \frac{1}{4}[(1-\varepsilon)^2+2\varepsilon(1-\varepsilon)+\varepsilon^2] = \frac{1}{4};$$

互信息：$\mathrm{I}(M_1;''00'') = \log\dfrac{p(''00''\,|\,M_1)}{q(''00'')} = \log\dfrac{(1-\varepsilon)^2}{1/4} = 2\log[2(1-\varepsilon)];$

附加信息：$\log[2(1-\varepsilon)]$

$$(3)\ q(''000'') = \sum_{i=1}^{4} p(M_i)p(000/M_i) = \frac{1}{4}[(1-\varepsilon)^3 + 3\varepsilon^2(1-\varepsilon)]$$

$$= \frac{1}{4}(1-\varepsilon)(4\varepsilon^2 - 2\varepsilon + 1);$$

互信息：$\mathrm{I}(M_1;''000'') = \log\dfrac{p(''000''\,|\,M_1)}{q(''000'')} = \log\dfrac{(1-\varepsilon)^3}{(1-\varepsilon)(4\varepsilon^2-2\varepsilon+1)/4}$

$$= 2\log[2(1-\varepsilon)] - \log(4\varepsilon^2-2\varepsilon+1);$$

又增加的信息为$-\log(4\varepsilon^2-2\varepsilon+1)$

3.3　条件互信息量

现在把目光转到级联信道的交互信息的讨论上。

如图 3.7 所示信道 *I* 与信道 *II* 串接。信道 *I* 的输入符号集 $X=\{a_1, a_2, \cdots, a_r\}$，输出符号集 $Y=\{b_1, b_2, \cdots, b_s\}$，信道 *II* 的输入符号集 $Y=\{b_1, b_2, \cdots, b_s\}$，输出符号集 $Z=\{c_1, c_2, \cdots, c_t\}$

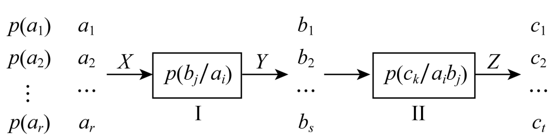


图 3.7　信道 *I* 与信道 *II* 串接

信道 *I* 的传递概率：

$$\mathrm{P}(\mathrm{Y/X}) = \{p(b_j/a_i)\} \quad (i=1, 2, \cdots, r;\ j=1, 2, \cdots, s)$$

信道 *II* 的传递概率：

$$\mathrm{P}(\mathrm{Z/XY}) = \{p(c_k/a_ib_j)\} \quad (i=1, 2, \cdots, r;\ j=1, 2, \cdots, s;\ k=1, 2, \cdots, t)$$

则 $p(a_ib_jc_k)=p(a_i)p(b_j/a_i)p(c_k/a_ib_j)$ $(i=1, 2, \cdots, r;\ j=1, 2, \cdots, s;\ k=1, 2, \cdots, t)$

由 X、Y、Z 的联合概率 $p(a_ib_jc_k)$ 可以求得其他各种概率分布。

一维分布：

$$p(a_i) = \sum_{j=1}^{s}\sum_{k=1}^{t} p(a_ib_jc_k)$$

$$p(b_j)=\sum_{i=1}^{r}\sum_{k=1}^{t}p(a_ib_jc_k)$$

$$p(c_k)=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_jc_k)$$

二维分布：

$$p(a_ib_j)=\sum_{k=1}^{t}p(a_ib_jc_k)$$

$$p(b_jc_k)=\sum_{i=1}^{r}p(a_ib_jc_k)$$

$$p(a_ic_k)=\sum_{j=1}^{s}p(a_ib_jc_k)$$

进一步求得条件概率分布：

$$p(a_ib_j/c_k)=\frac{p(a_ib_jc_k)}{p(c_k)}$$

$$p(a_ic_k/b_j)=\frac{p(a_ib_jc_k)}{p(b_j)}$$

$$p(b_jc_k/a_i)=\frac{p(a_ib_jc_k)}{p(a_i)}$$

$$p(a_i/b_jc_k)=\frac{p(a_ib_jc_k)}{p(b_jc_k)}$$

$$p(b_j/a_ib_j)=\frac{p(a_ib_jc_k)}{p(a_ic_k)}$$

$$p(c_k/a_ib_j)=\frac{p(a_ib_jc_k)}{p(a_ib_j)}$$

在串接信道输出某符号 c_k 条件下，从符号 b_j 中获取符号 a_i 的信息量定义为

$$\begin{aligned}I(a_i;\ b_j/c_k)&=\log\frac{p(a_i/b_jc_k)}{p(a_i/c_k)}\\&=\log\frac{1}{p(a_i/c_k)}-\log\frac{1}{p(a_i/b_jc_k)}\\&=I(a_i/c_k)-I(a_i/b_jc_k)\end{aligned}$$

$$(i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s;\ k=1,\ 2,\ \cdots,\ t)\quad(3.16)$$

该式表明，c_k 已知条件下，b_j 出现前后对 a_i 的条件不确定性的消除。

用 $p(b_jc_k)$ 同时乘上式右边的分子、分母，则有

$$\begin{aligned}I(a_i;\ b_j/c_k)&=\log\frac{p(b_jc_k)p(a_i/b_jc_k)}{p(b_jc_k)p(a_i/c_k)}=\log\frac{p(a_ib_jc_k)}{p(c_k)p(b_j/c_k)p(a_i/c_k)}\\&=\log\frac{p(a_ib_j/c_k)}{p(b_j/c_k)p(a_i/c_k)}=\log\frac{1}{p(a_i/c_k)}+\log\frac{1}{p(b_j/c_k)}-\log\frac{1}{p(a_ib_j/c_k)}\\&=I(a_i/c_k)+I(b_j/c_k)-I(a_ib_j/c_k)\end{aligned}$$

$$(i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.17}$$

上式表明，I 信道通信前、后，输入输出端同时出现 a_i 和 b_j 的条件不确定性的消除。

由 3.17 式，做进一步变换，有

$$\begin{aligned}
I(a_i;b_j/c_k) &= \log\frac{p(a_ib_j/c_k)}{p(b_j/c_k)p(a_i/c_k)} = \log\left\{\frac{p(a_ib_j/c_k)}{p(a_i/c_k)}\cdot\frac{1}{p(b_j/c_k)}\right\} \\
&= \log\frac{p(b_j/a_ic_k)}{p(b_j/c_k)} = \log\frac{1}{p(b_j/c_k)} - \log\frac{1}{p(b_j/a_ic_k)} \\
&= I(b_j/c_k) - I(b_j/a_ic_k) \\
&\quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t)
\end{aligned}$$

所以有 $I(a_i;b_j/c_k)=I(b_j;a_i/c_k)$

在图 3.7 中，随机变量 c_k 与随机变量 X 和 Y 的联合符号 (a_ib_j) 之间的相关交互信息量

$$\begin{aligned}
I(a_ib_j;c_k) &= \log\frac{p(c_k/a_ib_j)}{p(c_k)} = \log\frac{p(c_k/b_j)p(c_k/a_ib_j)}{p(c_k/b_j)p(c_k)} \\
&= \log\frac{p(c_k/b_j)}{p(c_k)} + \log\frac{p(c_k/a_ib_j)}{p(c_k/b_j)} \\
&= I(b_j;c_k) + I(b_j;c_k/b_j) \\
&\quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t)
\end{aligned}$$

这表明，相关交互信息量 $I(a_ib_j;c_k)$ 等于 b_j 与 c_k 之间的交互信息量 $I(b_j;c_k)$，再加上 b_j 已知的条件下，a_i 与 c_k 之间的条件交互信息量 $I(a_i;c_k/b_j)$ 所得之和。同样

$$\begin{aligned}
I(a_ib_j;c_k) &= \log\frac{p(c_k/a_ib_j)}{p(c_k)} = \log\frac{p(c_k/a_i)p(c_k/a_ib_j)}{p(c_k/a_i)p(c_k)} \\
&= \log\frac{p(c_k/a_i)}{p(c_k)} + \log\frac{p(c_k/a_ib_j)}{p(c_k/a_i)} \\
&= I(a_i;c_k) + I(b_j;c_k/a_i) \\
&\quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t)
\end{aligned}$$

这表明，相关交互信息量 $I(a_ib_j;c_k)$ 也等于 a_i 与 c_k 之间的交互信息量 $I(a_i;c_k)$，再加上 a_i 已知的条件下，b_j 与 c_k 之间的条件交互信息量 $I(b_j;c_k/a_i)$ 所得之和。

【例 3.5】 如表 3.1 所示，列出了无失真信源编码的信源消息、消息的先验概率以及每一个消息所对应的码字。

表 3.1　信源输出符号的条件概率

信源消息	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
码字	000	001	010	011	100	101	110	111
消息概率	1/4	1/4	1/8	1/8	1/16	1/16	1/16	1/16

试以消息 a_5 及相应码字(100)为例，分别说明码字(100)中每一个码符号对消息 a_5 提供的信息量。

解：根据相关交互信息量的理论可有

$$I(a_5;\ 100)=\log\frac{p(a_5/100)}{p(a_5)}=\log\frac{p(a_5/100)p(a_5/1)p(a_5/10)}{p(a_5)p(a_5/1)p(a_5/10)}$$

$$=\log\frac{p(a_5/1)}{p(a_5)}+\log\frac{p(a_5/10)}{p(a_5/1)}+\log\frac{p(a_5/100)}{p(a_5/10)}$$

$$=I(a_5;\ 1)+I(a_5;\ 0/1)+I(a_5;\ 0/10)$$

以下分别计算其中各项条件互信息。

(1) $I(a_5;\ 1)=\log\frac{p(a_5/1)}{p(a_5)}$

其中，$p(a_5/1)$表示收到码符号“1”后，判断信源发消息 a_5 的后验概率。因为收到码符号“1”后，再收到码符号序列“00”就构成码字(100)，即消息 a_5 出现。所以后验概率 $p(a_5/1)=p(00/1)$，即有

$$p(a_5/1)=p(00/1)=\frac{p(100)}{p(1)}$$

其中，码字(100)出现的概率 $p(100)$ 等于消息 a_5 出现的概率，即有 $p(100)=p(a_5)=\frac{1}{16}$

从表 3.1 中可看出，8 个码字中有 4 个码字(100)、(101)、(110)、(111)的第一个码符号是“1”，所以有

$$p(1)=p(100)+p(101)+p(110)+p(111)$$

$$=\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}=\frac{1}{4}$$

即可得 $p(a_5/1)=\frac{p(100)}{p(1)}=\frac{\frac{1}{16}}{\frac{1}{4}}=\frac{1}{4}$

这样就有 $I(a_5;\ 1)=\log\frac{p(a_5/1)}{p(a_5)}=\log\frac{\frac{1}{4}}{\frac{1}{16}}=2$(比特)

(2) $I(a_5;\ 0/1)=\log\frac{p(a_5/10)}{p(a_5/1)}$

其中，$p(a_5/10)$表示收到码符号“10”后，判断信源发消息 a_5 的后验概率。因为收到码符号“10”后，再收到码符号序列“0”就构成码字(100)，即消息 a_5 出现。所以后验概率 $p(a_5/10)=p(0/10)$，即有 $p(a_5/10)=p(0/10)=\frac{p(100)}{p(10)}$

从表 3.1 中可看出，8 个码字中有 2 个码字(100)、(101)的前两个码符号序列是“10”，所以有

$$p(10)=p(100)+p(101)=\frac{1}{16}+\frac{1}{16}=\frac{1}{8}$$

即可得 $p(a_5/10)=\dfrac{p(100)}{p(10)}=\dfrac{\frac{1}{16}}{\frac{1}{8}}=\dfrac{1}{2}$

这样就有 $I(a_5;0/1)=\log\dfrac{p(a_5/10)}{p(a_5/1)}=\log\dfrac{\frac{1}{2}}{\frac{1}{4}}=1$（比特）

（3）$I(a_5;0/10)=\log\dfrac{p(a_5/100)}{p(a_5/10)}$

其中，$p(a_5/100)$ 表示收到码符字（100）后，判断信源发消息 a_5 的后验概率。显然收到（100）后也就是收到了消息 a_5，所以有 $p(a_5/100)=1$

这样就有 $I(a_5;0/10)=\log\dfrac{p(a_5/100)}{p(a_5/10)}=\log\dfrac{1}{p(a_5/10)}=\log\dfrac{1}{\frac{1}{2}}=1$（比特）

以上三项计算结果表明，在消息 a_5 相对应的码字（100）中：第一个码符号“1”提供关于消息 a_5 的信息量 $I(1;a_5)=I(a_5;1)=2$（比特）；在收到第一个码符号“1”的条件下，第二个码符号“0”提供关于 a_5 的条件交互信息量 $I(0;a_5/1)=I(a_5;0/1)=1$（比特）；在收到第一个码符号“1”和第二个码符号“0”组成的码符号序列“10”的条件下，第三个码符号“0”提供关于 a_5 的条件交互信息量 $I(0;a_5/10)=I(a_5;0/10)=1$（比特）。所以，从码字（100）中的三个码符号总共提供关于消息 a_5 的相关交互信息量

$$\begin{aligned}I(100;a_5)=I(a_5;100)&=I(a_5;1)+I(a_5;0/1)+I(a_5;0/10)\\&=2+1+1=4\text{（比特）}\end{aligned}$$

另一方面，消息 a_5 的自信息量

$$I(a_5)=\log\frac{1}{p(a_5)}=\log\frac{1}{\frac{1}{16}}=\log 16=4\text{（比特）}$$

这从信息测量的角度证实了由于消息 a_5 与相应的码字（100）是一一对应的确定关系，相关交互信息量 $I(a_5;100)$ 就是消息 a_5 的自信息量 $I(a_5)$。

3.4　平均交互信息量

$I(a_i;b_j)$ 表示单个事件间的交互信息量，要求信道两端平均一对符号传递信息量的多少，就要计算平均交互信息量，如图 3.8 和图 3.9 所示。

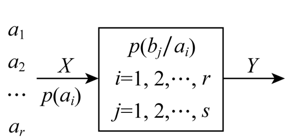


图 3.8　信息传输方向为 X 到 Y

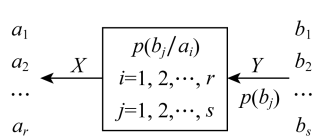


图 3.9　信息传输方向为 Y 到 X

首先给出两种特殊形式

1. $$I(X;b_j)=\sum_{i=1}^{r}p(a_i/b_j)I(a_i;b_j)=\sum_{i=1}^{r}p(a_i/b_j)\log\frac{p(a_i/b_j)}{p(a_i)}\quad (j=1,2,\cdots,s)\tag{3.18}$$

这表示信宿收到 b_j 后，获得有关信源的信息量。

2. $$I(a_i;Y)=\sum_{j=1}^{s}p(b_j/a_i)I(a_i;b_j)=\sum_{j=1}^{s}p(b_j/a_i)\log\frac{p(b_j/a_i)}{p(b_j)}\quad (i=1,2,\cdots,r)\tag{3.19}$$

这表示反向信道中信宿收到 a_i 后获得有关信源 Y 的信息量。

由上述两个式子可知，X 与 Y 间平均传递一个符号所传输的平均信息量 $I(X;Y)$ 应该是 $I(a_i;b_j)$ 在联合集 XY 中的统计均值，即

$$I(X;Y)=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)I(a_i;b_j)=I(Y;X)\tag{3.20}$$

同样，$I(X;Y)$ 也有三种不同的表达形式

1. 正向信道中收到 Y 前后关于 X 的不确定性的消除

$$\begin{aligned}I(X;Y)&=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)I(a_i;b_j)=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log\frac{p(a_i/b_j)}{p(a_i)}\\&=-\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log p(a_i)-\left\{-\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log p(a_i/b_j)\right\}\\&=-\sum_{i=1}^{r}p(a_i)\log p(a_i)-\sum_{j=1}^{s}p(b_j)\left\{-\sum_{i=1}^{r}p(a_i/b_j)\log p(a_i/b_j)\right\}\\&=H(X)-\sum_{j=1}^{s}p(b_j)H(X/Y=b_j)=H(X)-H(X/Y)\end{aligned}\tag{3.21}$$

其中，$$H(X/Y=b_j)=-\sum_{i=1}^{r}p(a_i/b_j)\log p(a_i/b_j)\quad (j=1,2,\cdots,s)\tag{3.22}$$

表示在随机变量$Y=b_j$　$(j=1,2,\cdots,s)$的前提下，对随机变量 X 仍然存在的平均不确定性。而条件熵

$$\begin{aligned}H(X/Y)&=\sum_{j=1}^{s}p(b_j)H(X/Y=b_j)=-\sum_{j=1}^{s}p(b_j)\sum_{i=1}^{r}p(a_i/b_j)\log p(a_i/b_j)\\&=-\sum_{i=1}^{r}\sum_{j=1}^{s}p(b_j)p(a_i/b_j)\log p(a_i/b_j)\\&=-\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log p(a_i/b_j)\end{aligned}\tag{3.23}$$

表示收到随机变量 Y 后，对随机变量 X 仍然存在的平均不确定性，通常称它为疑义度。

(3.21)式表明，从收到 Y 中获取关于 X 的平均交互信息量 $I(X;Y)$，等于收到 Y 前对 X 的平均不确定性 $H(X)$，与收到 Y 后对 X 仍然存在的平均不确定性 $H(X/Y)$之差，即收到 Y 前、后，关于 X 的平均不确定性的消除。

2. 反向信道中收到 X 前后关于 Y 的不确定性的消除

$$
\begin{aligned}
I(X;Y) &= \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)I(a_i;b_j) = \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log\frac{p(b_j/a_i)}{p(b_j)} \\
&= -\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(b_j) - \left\{-\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(b_j/a_i)\right\} \\
&= -\sum_{j=1}^{s} p(b_j)\log p(b_j) - \sum_{i=1}^{r} p(a_i)\left\{-\sum_{j=1}^{s} p(b_j/a_i)\log p(b_j/a_i)\right\} \\
&= H(Y) - \sum_{i=1}^{r} p(a_i)H(Y/X=a_i) = H(Y) - H(Y/X) = I(Y;X)
\end{aligned}
\tag{3.24}
$$

其中，
$$
H(Y/X=a_i) = -\sum_{j=1}^{s} p(b_j/a_i)\log p(b_j/a_i) \quad (i=1,2,\cdots,r) \tag{3.25}
$$
表示在随机变量 $X=a_i$ $(i=1, 2, \cdots, r)$ 的前提下，随机变量 Y 仍然存在的平均不确定性。而条件熵

$$
\begin{aligned}
H(Y/X) &= \sum_{i=1}^{r} p(a_i)H(Y/X=a_i) = -\sum_{i=1}^{r} p(a_i)\sum_{j=1}^{s} p(b_j/a_i)\log p(b_j/a_i) \\
&= -\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_i)p(b_j/a_i)\log p(b_j/a_i) \\
&= -\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(b_j/a_i)
\end{aligned}
\tag{3.26}
$$

表示表示在反向信道中，收到随机变量 X 后，对随机变量 Y 仍然存在的平均不确定性。这个“反向疑义度”一般称为噪声熵。

式(3.24)表明，对于反向信道来说，从输出随机变量 X 中，获取关于 Y 的平均交互信息量 $I(Y;X)$，等于信宿收到 X 前，对 Y 的先验不确定性 $H(Y)$，与信宿收到 X 后，对 Y 仍然存在的后验平均不确定性 $H(Y/X)$ 之差，即通信前、后，关于 Y 的平均不确定性的消除。

3. 通信前后，随机变量 X 和 Y 同时出现的平均不确定性的消除

$$
\begin{aligned}
I(X;Y) &= \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)I(a_i;b_j) = \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log\frac{p(a_ib_j)}{p(a_i)p(b_j)} \\
&= -\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(a_i) - \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(b_j) - \left\{-\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(a_ib_j)\right\} \\
&= -\sum_{i=1}^{r} p(a_i)\log p(a_i) - \sum_{j=1}^{s} p(b_j)\log p(b_j) - \left\{-\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(a_ib_j)\right\} \\
&= H(X) + H(Y) - H(XY)
\end{aligned}
\tag{3.27}
$$

其中，
$$
H(XY) = -\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(a_ib_j)
$$
表示通信后，信道两端同时出现 X 和 Y 的后验平均不确定性，通常称它为共熵。

式(3.27)表明，信源 X 通过传递概率为 $P(Y/X)$ 的信道输出随机变量 Y，信道传递的平均交互信息量 $I(Y;X)$，等于通信前(随机变量 X 和 Y 统计独立)随机变量 X 和 Y 同时出现的平均不确定性 $\{H(X)+H(Y)\}$，与通信后(随机变量 X 和 Y 由信道传递概率 $P(Y/X)$ 相联系)信道两端同时出现随机变量 X 和 Y 的平均不确定性 $H(XY)$ 之差，即通信

前、后，随机变量 X 和 Y 同时出现的平均不确定性的消除。

上述讨论的通信系统中各类熵的关系可用维拉图形象表示，如图 3.10 所示。

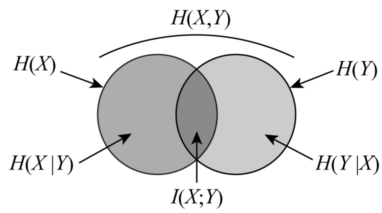


图 3.10　通信系统中各类商关系图

【例 3.6】设信源 X 的符号集 X：$\{a_1, a_2\}$，先验概率分布为 $p(a_1)=\omega(0<\omega<1)$；$p(a_2)=1-\omega$。信道的输入符号集 X：$\{a_1, a_2\}$，输出符号集：Y：$\{a_1, a_2\}$，传递概率 $P(Y/X)$：$\{p(a_j/a_i)=p_{ij}(i=1, 2; j=1, 2)\}$。现将信源 X 与如图 3.11 信道相接。

（一）试写出平均交互信息量 $I(X;Y)$ 的一般表达式；

（二）若 $\omega=\dfrac{1}{2}$，$p_{11}=p_{22}=\bar{p}$；$p_{12}=p_{21}=p$　$(0\leqslant\bar{p}, p\leqslant 1; \bar{p}+p=1)$。试计算如图 3.12 所示反向信道的 $I(X;Y)$。

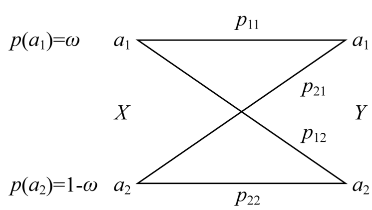


图 3.11　正向信道传递概率图

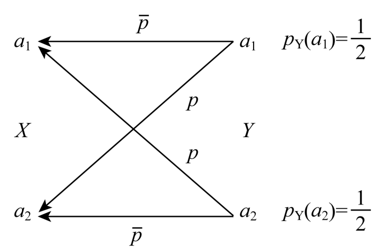


图 3.12　反向信道传递概率图

解：（一）平均交互信息量 $I(X;Y)$ 的一般表达式

（1）各联合概率

$$p(a_1a_1)=p(a_1)p(a_1/a_1)=\omega p_{11}$$
$$p(a_1a_2)=p(a_1)p(a_2/a_1)=\omega p_{12}$$
$$p(a_2a_1)=p(a_2)p(a_1/a_2)=(1-\omega)p_{21}$$
$$p(a_2a_2)=p(a_2)p(a_2/a_2)=(1-\omega)p_{22}$$

（2）随机变量 Y 的概率分布

$$p_Y(a_1)=p(a_1a_1)+p(a_2a_1)=\omega p_{11}+(1-\omega)p_{21}$$
$$p_Y(a_2)=p(a_1a_2)+p(a_2a_2)=\omega p_{12}+(1-\omega)p_{22}$$

（3）求随机变量 Y 的熵

$$\begin{aligned}H(Y) &= -\sum_{j=1}^{2}p_Y(a_j)\log p_Y(a_j) = -p_Y(a_1)\log p_Y(a_1) - p_Y(a_2)\log p_Y(a_2)\\ &= -\{[\omega p_{11}+(1-\omega)p_{21}]\log[\omega p_{11}+(1-\omega)p_{21}]\} - \{[\omega p_{12}+(1-\omega)p_{22}]\log[\omega p_{12}+(1-\omega)p_{22}]\}\end{aligned}$$

（4）由 $p(a_j/a_i)=p_{ij}(i=1, 2; j=1, 2)$ 求得条件熵

$$H(Y/X)=-\sum_{i=1}^{2}\sum_{j=1}^{2}p(a_ia_j)\log p(a_j/a_i)$$
$$=-\{p(a_1a_1)\log p(a_1/a_1)+p(a_1a_2)\log p(a_2/a_1)+p(a_2a_1)\log p(a_1/a_2)+p(a_2a_2)\log p(a_2/a_2)\}$$
$$=-\{\omega p_{11}\log p_{11}+\omega p_{12}\log p_{12}+(1-\omega)p_{21}\log p_{21}+(1-\omega)p_{22}\log p_{22}\}$$

（5）求得平均交互信息量

$$I(X;Y)=H(Y)-H(Y/X)$$
$$=-\{[\omega p_{11}+(1-\omega)p_{21}]\log[\omega p_{11}+(1-\omega)p_{21}]\}$$
$$-\{[\omega p_{12}+(1-\omega)p_{22}]\log[\omega p_{12}+(1-\omega)p_{22}]\}$$
$$+\{\omega p_{11}\log p_{11}+\omega p_{12}\log p_{12}+(1-\omega)p_{21}\log p_{21}+(1-\omega)p_{22}\log p_{22}\}$$

这说明平均交互信息量是信源概率分布 ω 和信道传递概率 $p_{ij}(i=1, 2; j=1, 2)$ 的函数。

（二）计算反向信道的 $I(X;Y)$

（1）各联合概率

$$p(a_1a_1)=\frac{1}{2}\bar{p}$$
$$p(a_1a_2)=\frac{1}{2}p$$
$$p(a_2a_1)=\frac{1}{2}p$$
$$p(a_2a_2)=\frac{1}{2}\bar{p}$$

（2）随机变量 Y 的概率分布

$$p_Y(a_1)=\frac{1}{2}\bar{p}+\frac{1}{2}p=\frac{1}{2}$$
$$p_Y(a_2)=\frac{1}{2}p+\frac{1}{2}\bar{p}=\frac{1}{2}$$

（3）求随机变量 Y 的熵

$$H(Y)=-\left\{\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right\}=1\quad（比特/符号）$$

（4）由 $p(a_j/a_i)=p_{ij}(i=1, 2; j=1, 2)$，求得噪声熵

$$H(Y/X)=-\left\{\frac{1}{2}\bar{p}\log\bar{p}+\frac{1}{2}p\log p+\frac{1}{2}p\log p+\frac{1}{2}\bar{p}\log\bar{p}\right\}$$
$$=-\{\bar{p}\log\bar{p}+p\log p\}=H(\bar{p}, p)\quad（比特/符号）$$

（5）求得平均交互信息量

$$I(X;Y)=H(Y)-H(Y/X)=1-H(\bar{p}, p)\quad（比特/符号）$$

3.5 平均交互信息量的性质

如上节所述，平均交互信息量 $I(X;Y)$ 除具有对称性以外，即 $I(X;Y)=I(Y;X)$，还具有如下基本性质。

1. 平均互信息的非负性

$$I(X;Y)\geqslant 0$$

当且仅当 X 和 Y 统计独立时，等式成立。

【证明】利用詹森不等式得

$$\begin{aligned}-I(X;b_j) &= \sum_{i=1}^{r} p(a_i/b_j)\log\frac{p(a_i/b_j)}{p(a_i)}\\ &\leqslant \log\left\{\sum_{i=1}^{r} p(a_i/b_j)\frac{p(b_j)}{p(a_i/b_j)}\right\}\\ &= \log\left\{\sum_{i=1}^{r} p(a_i)\right\} = \log 1 = 0\end{aligned}$$

即有 $I(X;b_j)\geqslant 0 \quad (j=1,2,\cdots,s)$

$$I(X;Y)=\sum_{j=1}^{s} p(b_j)I(X;b_j)\geqslant 0$$

当且仅当对一切 i，j 都有 $p(a_ib_j)=p(a_i)p(b_j) \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s)$

即当 X 和 Y 统计独立时，$I(X;Y)=0$

2. 平均互信息的极值性

由上述性质，直接得到

$$I(X;Y)\leqslant H(X)$$

$$I(X;Y)\leqslant H(Y)$$

【证明】由于 $\log\dfrac{1}{p(a_i/b_j)}\geqslant 0$，而 $H(X/Y)$ 是对 $\log\dfrac{1}{p(a_i/b_j)}$ 求统计平均，即

$$H(X/Y)=\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log\frac{1}{p(a_i/b_j)}$$

因此有
$$H(X/Y)\geqslant 0$$

同理
$$H(Y/X)\geqslant 0$$

所以
$$I(X;Y)=H(X)-H(X/Y)\leqslant H(X)$$

$$I(X;Y)=H(Y)-H(Y/X)\leqslant H(Y)$$

即
$$I(X;Y)\leqslant \min\{H(X),H(Y)\}$$

3. 平均互信息的凸函数性

定理 3.1 信道两端随机变量 X 和 Y 之间的平均互信息量 $I(X;Y)$，在信道转移概率 $p(b_j/a_i)$ 给定条件下，是输入随机变量 X 的概率分布 $p(X)$：$\{p(a_i),\ i=1,2,\cdots,r\}$ 的∩型凸函数。

【证明】当条件概率 $P(y/x)$ 是固定时，平均互信息 $I(X;Y)$ 只是 $P(x)$ 的函数。简写成 $I[P(x)]$。现选择输入信源 X 的两种已知的概率分布 $P_1(x)$ 和 $P_2(x)$。其对应的联合概率分布为 $P_1(xy)=P_1(x)P(y/x)$ 和 $P_2(xy)=P_2(x)P(y/x)$，因而平均互信息分别为 $I[P_1(x)]$ 和 $I[P_2(x)]$。再选择输入变量 X 的另一种概率分布 $P(x)$，令 $0<\theta<1$，和 $\theta+\bar{\theta}=1$，而 $P(x)=\theta P_1(x)+\bar{\theta}P_2(x)$，因而得其相应的平均互信息为 $I[P(x)]$。

根据平均互信息的定义得

$$\begin{aligned}
&\theta I[P_1(x)]+\bar{\theta}I[P_2(x)]-I[P(x)]\\
&=\sum_{X,Y}\theta P_1(xy)\log\frac{P(y/x)}{P_1(y)}+\sum_{X,Y}\bar{\theta}P_2(xy)\log\frac{P(y/x)}{P_2(y)}-\sum_{X,Y}P(xy)\log\frac{P(y/x)}{P(y)}\\
&=\sum_{X,Y}\theta P_1(xy)\log\frac{P(y/x)}{P_1(y)}+\sum_{X,Y}\bar{\theta}P_2(xy)\log\frac{P(y/x)}{P_2(y)}-\sum_{X,Y}[\theta P_1(xy)+\bar{\theta}P_2(xy)]\log\frac{P(y/x)}{P(y)}
\end{aligned}$$

上式中根据概率关系

$$P(xy)=P(x)P(y/x)=\theta P_1(x)P(y/x)+\bar{\theta}P_2(x)P(y/x)=\theta P_1(xy)+\bar{\theta}P_2(xy)$$

所以得

$$\begin{aligned}
&\theta I[P_1(x)]+\bar{\theta}I[P_2(x)]-I[P(x)]\\
&=\theta\sum_{X,Y}P_1(xy)\log\frac{P(y)}{P_1(y)}+\bar{\theta}\sum_{X,Y}P_2(xy)\log\frac{P(y)}{P_2(y)}
\end{aligned}$$

因为 $\log x$ 是 x 的∩型函数，所以对上式中第一项，根据詹森不等式得

$$\begin{aligned}
&\sum_{X,Y}P_1(xy)\log\frac{P(y)}{P_1(y)}\leqslant\log\sum_{X,Y}P_1(xy)\log\frac{P(y)}{P_1(y)}\\
&=\log\sum_{Y}\frac{P(y)}{P_1(y)}\sum_{X}P_1(xy)=\log\sum_{Y}\frac{P(y)}{P_1(y)}P_1(y)=\log\sum_{Y}P_1(y)=0
\end{aligned}$$

同理
$$\sum_{X,Y}P_2(xy)\log\frac{P(y)}{P_2(y)}\leqslant 0$$

又因为 θ 和 $\bar{\theta}$ 都小于 1 且大于 0，即得 $\theta I[P_1(x)]+\bar{\theta}I[P_2(x)]-I[P(x)]\leqslant 0$，因而

$$I[\theta P_1(x)+\bar{\theta}P_2(x)]\geqslant\theta I[P_1(x)]+\bar{\theta}I[P_2(x)]$$

因此根据凸函数的定义知，$I(X;Y)$ 是概率分布 $P(x)$ 的∩型凸函数。

定理 3.2　信道两端随机变量 X 和 Y 之间的平均交互信息量 $I(X;Y)$，在信源概率分布 $P(a_i)$ 给定的条件下，是信道转移概率 $p(Y/X)$：$\{p(b_j/a_i),\ i=1,2,\cdots,r;\ j=1,2,\cdots,s\}$ 的∪型凸函数。

【证明】当概率分布 $P(x)$ 固定时，平均互信息 $I(X;Y)$ 只是条件概率 $P(y|x)$ 的函数，简写成 $I[P(y/x)]$。选择两种条件概率分别为 $P_1(y|x)$ 和 $P_2(y|x)$。相对应的平均互信息分别为 $I[P_1(y|x)]$ 和 $I[P_2(y|x)]$，再选择第三种条件概率满足 $P(y|x)=\theta P_1(y|x)+\bar{\theta}P_2(y|x)$。设相应的平均互信息为 $I[P(y/x)]$。其中 $0<\theta<1$，和 $\theta+\bar{\theta}=1$。因而求得

$$I[P(y|x)]-\theta I[P_1(y|x)]-\bar{\theta}I[P_2(y|x)]$$

$$=\sum_{X,Y}[\theta P_1(xy)+\bar{\theta}P_1(xy)]\log\frac{P(x|y)}{P(x)}-\sum_{X,Y}\theta P_1(xy)\log\frac{P_1(x|y)}{P(x)}-\sum_{X,Y}\bar{\theta}P_2(xy)\log\frac{P_2(x|y)}{P(x)}$$

$$=\theta\sum_{X,Y}P_1(xy)\log\frac{P(x|y)}{P_1(x|y)}+\sum_{X,Y}\bar{\theta}P_2(xy)\log\frac{P(x|y)}{P_2(x|y)}$$

运用詹森不等式，上式中第一项为

$$\theta\sum_{X,Y}P_1(xy)\log\frac{P(x|y)}{P_1(x|y)}\leqslant\theta\log\left[\sum_{X,Y}P_1(xy)\frac{P(x|y)}{P_1(x|y)}\right]=\theta\log\left[\sum_{X,Y}P_1(y)P(x|y)\right]$$

$$=\theta\log\left[\sum_{Y}P_1(y)\sum_{X}P(x|y)\right]=\theta\log\sum_{Y}P_1(y)=\theta\log1=0$$

同理 $$\sum_{X,Y}\theta P_2(xy)\log\frac{P(x|y)}{P_2(x|y)}\leqslant 0$$

所以 $I[P(y|x)]-\theta I[P_1(y|x)]-\bar{\theta}I[P_2(y|x)]\leqslant 0$，即

$$I[\theta P_1(y/x)+\theta P_2(y/x)]\leqslant\theta I[P_1(y/x)]+\bar{\theta}I[P_2(y/x)]$$

根据凸函数的定义得：平均互信息 $I(X;Y)$ 是条件概率 $P(y|x)$ 的∪型凸函数。

3.6 信道容量及其一般算法

3.6.1 信道容量的定义

信道的信息传输率定义为信道中平均每个符号所传送的信息量，即平均互信息。

$$R=I(X;Y)=H(X)-H(X/Y)\quad(\text{比特/符号})$$

信道的信息传输速率定义为信道平均每秒传输的信息量。若传输一个符号平均需要 t 秒，则信道的信息传输速率表示为

$$R_t=\frac{1}{t}I(X;Y)=\frac{1}{T_s}[H(X)-H(X/Y)]$$

上式中 T_s 为信道码元传输时间

给定某个信道，平均交互信息量 $I(X;Y)$ 是信源概率分布 $P(x)$ 的上凸型函数，存在极大值，这个极大值就定义为信道容量。

$$C=R_{\max}=\max_{p(x)}\{I(X;Y)\}\quad(\text{比特/符号})$$

信道的最大信息传输速率是信道容量的另一种表述形式

$$C_t=R_{t\max}=\frac{1}{t}\max\{I(X;Y)\}\quad(\text{比特/秒})$$

3.6.2 信道容量的一般算法

平均互信息量 $I(X;Y)$ 是输入信源概率分布 $p(X)$ 的∩型凸函数，所以极大值是一定

存在的。而 $I(X;Y)$ 是 r 个输入信号变量 $\{p(a_1),p(a_2),\cdots,p(a_r)\}$ 的多元函数，并且任何信源概率分布都必须遵循约束条件

$$\sum_{i=1}^{r} p(a_i)=1$$

所以求信道容量 C 就是在约束条件式的约束下，求 $I(X;Y)$ 的最大值问题，并导出取最大值时的条件 $p(a_i)(i=1,2,\cdots,r)$。

此类问题可以通过拉格朗日乘子法来计算。为此，作辅助函数

$$F[P(a_1),p(a_2),\cdots,P(a_r)]=I(X;Y)-\lambda\sum_{X}P(a_i)$$

式中，λ 为拉格朗日乘子。

当

$$\frac{\partial F}{\partial P(a_i)}=\frac{\partial\left[I(X;Y)-\lambda\sum_{X}P(a_i)\right]}{\partial P(a_i)}=0 \tag{3.28}$$

时求得 $I(X;Y)$ 的值即为信道容量。

由于

$$\begin{aligned}I(X;Y)&=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_i)p(b_j/a_i)\log\frac{p(b_j/a_i)}{p(b_j)}\\&=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_i)p(b_j/a_i)(\log p(b_j/a_i)-\log p(b_j))\end{aligned}$$

而 $p(b_j)=\sum_{i=1}^{r}p(a_i)p(b_j/a_i)$

所以 $$\frac{\partial F}{\partial P(a_i)}\log P(b_j)=\left[\frac{\partial\ln P(b_j)}{\partial P(a_i)}\right]\log e=\frac{P(b_j/a_i)}{p(b_j)}\log e$$

对式(3.28)整理得

$$\begin{aligned}\frac{\partial F}{\partial P(a_i)}&=\frac{\partial\left[I(X;Y)-\lambda\sum_{X}P(a_i)\right]}{\partial P(a_i)}\\&=\sum_{j=1}^{s}P(b_j/a_i)\log\frac{P(b_j/a_i)}{P(b_j)}-\sum_{k=1}^{r}\sum_{j=1}^{s}P(a_k)P(b_j/a_k)\frac{P(b_j/a_i)}{P(b_j)}\log e-\lambda\end{aligned} \tag{3.29}$$

而

$$\begin{aligned}\sum_{k=1}^{r}\sum_{j=1}^{s}P(a_k)P(b_j/a_k)\frac{P(b_j/a_i)}{P(b_j)}&=\sum_{k=1}^{r}\sum_{j=1}^{s}P(a_kb_j)\frac{P(b_j/a_i)}{P(b_j)}\\&=\sum_{j=1}^{s}P(b_j)\frac{P(b_j/a_i)}{P(b_j)}=\sum_{j=1}^{s}P(b_j/a_i)=1\end{aligned}$$

因此,式(3.29)可以简化为

$$\frac{\partial F}{\partial P(a_i)}=\frac{\partial\left[I(X;Y)-\lambda\sum_{X}P(a_i)\right]}{\partial P(a_i)}$$

$$= \sum_{j=1}^{s} P(b_j/a_i)\log\frac{P(b_j/a_i)}{P(b_j)} - \log e - \lambda$$

令$\frac{\partial F}{\partial P(a_i)}=0$，得

$$\sum_{j=1}^{s} P(b_j/a_i)\log\frac{P(b_j/a_i)}{P(b_j)} = \lambda + \log e \quad (i=1,2,\cdots,r) \tag{3.30}$$

式(3.30)两边分别乘以$P(a_i)$，并求和得

$$\sum_{i=1}^{r}\sum_{j=1}^{s} P(a_i)P(b_j/a_i)\log\frac{P(b_j/a_i)}{P(b_j)} = \lambda + \log e \tag{3.31}$$

式(3.31)左边即为平均互信息的极大值C，即

$$C=\lambda+\log e \tag{3.32}$$

结合式(3.32)，把式(3.30)中前r个方程改写成

$$\sum_{j=1}^{s} P(b_j/a_i)\log P(b_j/a_i) - \sum_{j=1}^{s} P(b_j/a_i)\log P(b_j) = C \quad (i=1,2,\cdots,r)$$

移项后得

$$\sum_{j=1}^{s} P(b_j/a_i)\log P(b_j/a_i) = \sum_{j=1}^{s} P(b_j/a_i)[\log P(b_j) + C] \quad (i=1,2,\cdots,r)$$

令

$$\beta_j = C+\log P(b_j)$$

得

$$\sum_{j=1}^{s} P(b_j/a_i)\log P(b_j/a_i) = \sum_{j=1}^{s} P(b_j/a_i)\beta_j \quad (i=1,2,\cdots,r)$$

这是含有s个未知参数β_j，有r个方程的非齐次线性方程组。

如果设$r=s$，信道传递矩阵P是非奇异矩阵，则此方程组有解，并且可以求出β_j的数值，然后根据$\sum_{j=1}^{s} P(b_j)=1$的附加条件求得信道容量

$$C=\log\sum_{j} 2^{\beta_j} \quad (比特/符号)$$

由这个C值就可以解得对应的输出概率分布$p(b_j)$

$$p(b_j)=e^{(\beta_j-C)}$$

再根据$p(b_j)=\sum_{i=1}^{r} p(a_i)p(b_j/a_i) \quad (j=1,2,\cdots,s)$，即可解出最佳输入分布$P(a_i)$。

观察式(3.30)可以发现，该式左边正好是输出端接收到符号Y后，获得的关于输入符号x_i的信息量，结合式(3.32)可知

$$I(x_i;Y)=\sum_{j=1}^{s} P(b_j/a_i)\log\frac{P(b_j/a_i)}{P(b_j)} = C$$

由此可以导出以下定理

定理 3.3 一般离散信道的平均互信息$I(X;Y)$达到信道容量的充要条件是输入概率分布$\{p_i\}$满足

$$\begin{cases} I(x_i;Y)=C, & p_i\neq 0 \\ I(x_i;Y)\leqslant C, & p_i=0 \end{cases}$$

这时的 $I(X;Y)$ 就是信道容量 C。

如果求解达到信道容量时，最佳概率分布中，某些 $P(a_i)<0$，则这些解无效。它表明所求极大值 C 出现的区域不满足概率条件，这时最大值必须在边界上，即某些 a_i 的概率 $P(a_i)=0$。因此，必须设某些信源符号 a_i 的概率 $P(a_i)=0$，然后重新进行运算。当 $r<s$ 时，求解非其次线性方程就比较困难，即使求出解，也无法保证求得的信源符号概率都大于或等于零。因此，必须反复进行运算，这就使运算变得非常复杂。

【例 3.7】设离散信道的输入符号集为 $\{a_1, a_2, a_3, a_4\}$，输出符号集为 $\{b_1, b_2, b_3, b_4\}$，其信道转移概率矩阵为 $[P(y/x)]=\begin{bmatrix} 0.25 & 0.5 & 0 & 0.25 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0.125 & 0.125 & 0.25 & 0.5 \end{bmatrix}$，求其信道容量及最佳输入分布。

解：列得

$$\begin{cases} 0.25\beta_1+0.5\beta_2+0.25\beta_4=0.25\log 0.25+0.5\log 0.5+0.25\log 0.25 \\ 0.5\beta_1+0.5\beta_3=0.5\log 0.5+0.5\log 0.5 \\ 0.5\beta_2+0.5\beta_3=0.5\log 0.5+0.5\log 0.5 \\ 0.125\beta_1+0.125\beta_2+0.25\beta_3+0.5\beta_4=-(0.125\times 3+0.125\times 3+0.25\times 2+0.5\times 1) \end{cases}$$

解得

$$\beta_1=\beta_2=-\frac{7}{6},\quad \beta_3=-\frac{5}{6},\quad \beta_4=-\frac{5}{2}$$

则信道的信道容量 $C=\log\sum_j 2^{\beta_j}=0.7039$ （比特/符号）

由 $p(b_j)=e^{(\beta_j-C)}$ 得 $p(b_1)=p(b_2)=0.2735$，$p(b_3)=0.3445$ $\quad p(b_4)=0.1085$

由于 $p(b_j)=\sum_{i=1}^{r}p(a_i)p(b_j/a_i)\quad (j=1, 2, \cdots, r)$ 可写成下式

$$[p(a_1), p(a_2), p(a_3), p(a_4)]p(y/x)=[p(b_1), p(b_2), p(b_3), p(b_4)]$$

因此

$$\begin{aligned}&[p(a_1), p(a_2), p(a_3), p(a_4)]\\ &=[0.2735 \quad 0.2735 \quad 0.3445 \quad 0.1085]\begin{bmatrix} 0.25 & 0.5 & 0 & 0.25 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0.125 & 0.125 & 0.25 & 0.5 \end{bmatrix}^{-1}\\ &=[0.2698 \quad 0.3915 \quad 0.2566 \quad 0.0821]\end{aligned}$$

因此，最佳分布为 $p(a_1)=0.2698 \quad p(a_2)=0.3915 \quad p(a_3)=0.2566 \quad p(a_4)=0.0821$

现在验证以上结果

$$
\begin{aligned}
I(x=a_1;\ Y) &= \sum_{j=1}^{4} p(b_j/a_1)\log\frac{p(b_j/a_1)}{p(b_j)} \\
&= \frac{1}{4}\log\frac{1/4}{0.2735}+\frac{1}{2}\log\frac{1/2}{0.2735}+\frac{1}{4}\log\frac{1/4}{0.1085} \\
&= 0.7039(\text{比特/符号})
\end{aligned}
$$

同理可以计算

$$I(x=a_2;\ Y) = \sum_{j=1}^{4} p(b_j/a_2)\log\frac{p(b_j/a_2)}{p(b_j)} = 0.7039(\text{比特/符号})$$

$$I(x=a_3;\ Y) = \sum_{j=1}^{4} p(b_j/a_3)\log\frac{p(b_j/a_3)}{p(b_j)} = 0.7039(\text{比特/符号})$$

$$I(x=a_4;\ Y) = \sum_{j=1}^{4} p(b_j/a_4)\log\frac{p(b_j/a_4)}{p(b_j)} = 0.7039(\text{比特/符号})$$

显然,

$$\begin{cases} I(x_i;\ Y)=0.7039, & p_i\neq 0 \\ I(x_i;\ Y)\leqslant C, & p_i=0 \end{cases}$$

而每个符号贡献的互信息也正好是前文求解出的信道容量，证实了该求解过程是正确的。

3.7 几种特殊结构的信道容量计算

3.7.1 无噪无损信道

如图 3.13 所示

$$X\left\{\begin{array}{ccc} a_1 & \text{——} & b_1 \\ a_2 & \text{——} & b_2 \\ a_3 & \text{——} & b_3 \end{array}\right\}Y$$

图 3.13 无噪无损信道

输入输出符号一一对应，信道转移概率矩阵为单位矩阵

$$P=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

此时平均互信息满足

$$I(X;\ Y)=H(X)=H(Y)$$

噪声熵 $H(Y/X)$ 和损失熵 $H(X/Y)$ 都为 0。

此时输入符号的概率分布为等概分布。所以无噪无损信道的信道容量

$$C=\max_{p(x)}\{I(X;\ Y)\}=\max_{p(x)}\{H(X)\}=\log r(\text{bit/符号})$$

3.7.2　有噪无损信道

此时噪声熵 $H(Y/X)\neq 0$，损失熵为 0，对应信道的一个输入符号 a_i 有多个输出符号 b_j 与之对应，如图 3.14 所示。

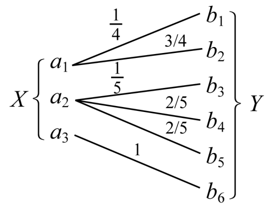


图 3.14　有噪无损信道

信道转移概率矩阵

$$P=\begin{bmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{5} & \frac{2}{5} & \frac{2}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

信道反向转移概率 $p(a_i/b_j)=1$，$I(X;Y)=H(X)<H(Y)$，信道容量

$$C=\max_{p(x)}H(X)=\log r\text{(比特/符号)}$$

输入符号等概分布时达到信道容量。

3.7.3　无噪有损信道

属于多个输入对应一个输出的信道如图 3.15 所示。

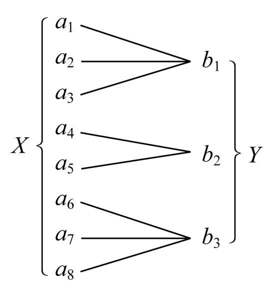


图 3.15　无噪有损信道

顾名思义，这种信道噪声熵 $H(Y/X)=0$，损失熵 $H(X/Y)\neq 0$

$$I(X;Y)=H(Y)<H(X)$$

信道容量

$$C=\max_{p(x)}H(Y)=\log s\text{(比特/符号)}$$

一定存在一种输入符号分布，使得输出符号 $p(b_j)=\frac{1}{s}$，达到等概分布。

3.7.4 对称离散信道的信道容量

若信道转移概率矩阵每行元素构成相同，称为输入对称；若转移矩阵中，每列元素构成相同，称为输出对称。若输入和输出都对称，此时信道称为对称信道。比如

$$P_1=\begin{bmatrix}\frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3}\end{bmatrix},\quad P_2=\begin{bmatrix}\frac{1}{2} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2}\\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3}\end{bmatrix}$$

都是对称信道矩阵

若输入符号和输出符号个数相同，都等于 r，那么信道矩阵为

$$P=\begin{bmatrix}\bar{p} & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \frac{p}{r-1}\\ \frac{p}{r-1} & \bar{p} & \frac{p}{r-1} & \cdots & \frac{p}{r-1}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ \frac{p}{r-1} & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \bar{p}\end{bmatrix}$$

式中，$0\leqslant p$，$\bar{p}\leqslant 1$，且 $p+\bar{p}=1$，则此信道称为强对称信道或均匀信道。这类信道中总的错误概率为 p，对称地平均分配给 $r-1$ 个输出符号。它是对称离散信道的一类特例。二元对称信道就是 $r=2$ 的均匀信道。对于均匀信道，其信道矩阵中各列之和也等于 1(一般信道的信道矩阵中各列之和不一定等于 1)。

对于对称信道，噪声熵 $H(Y/X)$ 为

$$\begin{aligned}H(Y/X) &= -\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_i)P(b_j/a_i)\log P(b_j/a_i)\\ &= \sum_{i=1}^{r}P(a_i)\left\{-\sum_{j=1}^{s}P(b_j/a_i)\log P(b_j/a_i)\right\}\\ &= \sum_{i=1}^{r}P(a_i)H(p_1',\ p_2',\ \cdots,\ p_s')=H(p_1',\ p_2',\ \cdots,\ p_s')\end{aligned}$$

可见，离散对称信道的噪声熵就是矩阵某一行元素所对应的熵 $H(p_1',\ p_2',\ \cdots,\ p_s')$。信道容量为

$$\begin{aligned}C &= \max_{P(x)}\{I(X;\ Y)\}=\max_{P(x)}\{H(Y)-H(Y/X)\}\\ &=\max_{P(x)}\{H(Y)\}-H(p_1',\ p_2',\ \cdots,\ p_s')\\ &=\log s-H(p'_1,\ p_2',\ \cdots,\ p_s')=\log s-H(P\ \text{的行矢量})\end{aligned}\tag{3.33}$$

当信道输入符号等概，即 $p(a_i)=\dfrac{1}{r}$，信道输出符号概率

$$
\begin{aligned}
p(b_j) &= \sum_{i=1}^{r} p(a_i)p(b_j/a_i) \\
&= \frac{1}{r}\sum_{i=1}^{r} p(b_j/a_i) = 常数 \quad (j=1,\ 2,\ \cdots,\ s)
\end{aligned}
$$

式中，$\sum_{i=1}^{r} p(b_j/a_i)$ 是对称信道转移概率中的列和，是一个固定值。所以当信道输入符号等概时，输出符号也等概，达到(3.33)式所对应的信道容量，类似可以得到强对称信道的信道容量。

$$
\begin{aligned}
C &= \log r - H\left(\bar{p},\ \frac{p}{r-1},\ \frac{p}{r-1},\ \cdots,\ \frac{p}{r-1}\right) \\
&= \log r + \bar{p}\log\bar{p} + \underbrace{\frac{p}{r-1}\log\frac{p}{r-1} + \cdots + \frac{p}{r-1}\log\frac{p}{r-1}}_{共(r-1)项} \\
&= \log r + \bar{p}\log\bar{p} + p\log\frac{p}{r-1} = \log r - p\log(r-1) - H(p)
\end{aligned}
$$

3.7.5　准对称信道的信道容量

若信道转移概率矩阵中，每行都是同一行元素的不同排列，每列元素构成不同，但该信道矩阵按列可以划分为互不相交的子矩阵，每个子矩阵都是对称矩阵，则该信道称为准对称信道。例如，信道矩阵 $P=\begin{bmatrix}0.7 & 0.1 & 0.2\\ 0.2 & 0.1 & 0.7\end{bmatrix}$ 可以划分成两个对称的子矩阵 $P_1=\begin{bmatrix}0.7 & 0.2\\ 0.2 & 0.7\end{bmatrix}$ 和 $P_2=\begin{bmatrix}0.1\\ 0.1\end{bmatrix}$，因此它是对称信道。

定理 3.4　对于准对称离散信道，当输入等概率时达到信道容量，其信道容量为

$$
C = H(Y) - H(p'_1,\ p'_2,\ \cdots,\ p'_s) = \log r - \sum_{k=1}^{n} N_k\log M_k - H(p'_1,\ p'_2,\ \cdots,\ p'_s)
$$

式中，$H(Y)$ 为输入等概时信道输出的熵，n 为准对称离散信道矩阵按列可以划分成互不相交的子集，N_k 是第 k 个子矩阵中的行元素之和，M_k 是第 k 个子矩阵中的列元素之和，$H(p'_1,\ p'_2,\ \cdots,\ p'_s)$ 就是信道矩阵 P 中行元素集合 $\{p'_1,\ p'_2,\ \cdots,\ p'_s\}$ 的 s 个元素构成的熵函数。

【例 3.8】 设某信道的转移矩阵为

$$
P=\begin{bmatrix}0.5 & 0.3 & 0.2\\ 0.3 & 0.5 & 0.2\end{bmatrix}
$$

计算其信道容量。

分析：该信道为一个准对称信道，计算信道容量即为输入等概时的平均互信息量。

解： 将 $P=\begin{bmatrix}0.5 & 0.3 & 0.2\\ 0.3 & 0.5 & 0.2\end{bmatrix}$ 划分为两个对称子矩阵

$$P_1=\begin{bmatrix}0.5 & 0.3\\ 0.3 & 0.5\end{bmatrix},\quad P_2=\begin{bmatrix}0.2\\ 0.2\end{bmatrix}$$

因为

$$r=2,\ N_1=0.5+0.3=0.8,\ M_1=0.5+0.3=0.8,$$
$$N_2=0.2,\ M_2=0.2+0.2=0.4,\ n=2$$

则，该准对称离散信道的信道容量为

$$\begin{aligned}C&=\log r-\sum_{k=1}^{n}N_k\log M_k-H(p'_1,\ p'_2,\ \cdots,\ p'_s)\\&=\log 2-(0.8\log 0.8+0.2\log 0.4)-H(0.5,\ 0.3,\ 0.2)\\&=0.036(\text{比特}/\text{符号})\end{aligned}$$

需要指出的是，由于本题的信道为准对称信道，信道容量也可以通过计算输入等概时的平均互信息得到。

3.8　信道容量的迭代计算

对于一般离散信道来说，信道容量的计算比较复杂。迭代计算是一种常用的近似方法。

设单符号离散信道的输入符号集 X：$\{a_1,\ a_2,\ \cdots,\ a_r\}$、输出符号集 Y：$\{b_1,\ b_2,\ \cdots,\ b_s\}$，信道的传递概率 $P(Y/X)$：$\{p(b_j/a_i)(i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s)\}$。若输入符号集 X：$\{a_1,\ a_2,\ \cdots,\ a_r\}$的概率分布 $P(X)$：$\{p(a_i)(i=1,\ 2,\ \cdots,\ r)\}$，则信道的平均交互信息量

$$I(X;\ Y)=H(X)-H(X/Y)=-\sum_{i=1}^{r}p(a_i)\ln p(a_i)+\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_i)p(b_j/a_i)\ln p(a_i/b_j)\tag{3.34}$$

是输入信源 X：$\{a_1,\ a_2,\ \cdots,\ a_r\}$的概率分布 $P(X)$：$\{p(a_i)(i=1,\ 2,\ \cdots,\ r)\}$和后验概率 $P(X/Y)$：$\{p(a_i/b_j)(i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s)\}$的函数

$$I(X;\ Y)=I[p(a_i),\ p(a_i/b_j)]\tag{3.35}$$

而先验概率 $p(a_i)$和后验概率 $p(a_i/b_j)$不是两个独立的变量，他们之间按照如下关系式

$$p(a_i/b_j)=\frac{p(a_i)p(b_j/a_i)}{\sum_{i=1}^{r}p(a_i)p(b_j/a_i)}\quad(i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s)\tag{3.36}$$

发生相应的变化。

为了导出近似算法，暂时把后验概率 $p(a_i/b_j)$当作自变量，而把本来要随之发生相应变化的应急变量 $p(a_i)$近似看作固定不变量。由于 $p(b_j/a_i)$也固定不变，则(3.34)式所示的平均交互信息量 $I(X;\ Y)$就可以看作是后验概率 $p(a_i/b_j)$的函数

$$I(X;\ Y)=I[p(a_i/b_j)]\tag{3.37}$$

由于“底”大于 1 的对数是∩型凸函数，所以 $I[p(a_i/b_j)]$具有上凸性。这样，就可以

在条件

$$\sum_{i=1}^{r} p(a_i/b_j) = 1 \quad (j=1,2,\cdots,s) \tag{3.38}$$

的约束下，对变量 $p(a_i/b_j)$ 求∩型凸函数 $I[p(a_i/b_j)]$ 的条件极大值，以及达到极大值的 $p(a_i/b_j)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)$

为此，作辅助函数

$$F[p(a_i/b_j),\lambda] = I[p(a_i/b_j)] + \lambda_j\left[\sum_{i=1}^{r} p(a_i/b_j) - 1\right] \quad (j=1,2,\cdots,s)$$

取函数 $F[p(a_i/b_j),\lambda]$ 对 $p(a_i/b_j)$ 的偏导数，并置之为零，得到稳定点方程

$$\frac{\partial}{\partial\, p(a_i/b_j)}F[p(a_i/b_j),\lambda]=0 \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.39}$$

把式(3.34)代入式(3.39)，有

$$\frac{p(a_i)p(b_j/a_i)}{p(a_i/b_j)}+\lambda_j=0 \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.40}$$

即有

$$p(a_i/b_j)=-\frac{p(a_i)p(b_j/a_i)}{\lambda_j} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.41}$$

由约束条件式(3.38)，得

$$\lambda_j=-\sum_{i=1}^{r} p(a_i)p(b_j/a_i) \quad (j=1,2,\cdots,s) \tag{3.42}$$

则可以得到

$$p^*(a_i/b_j)=\frac{p(a_i)p(b_j/a_i)}{\sum_{i=1}^{r} p(a_i)p(b_j/a_i)} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.43}$$

这表明，当采用“把后验概率 $P(X/Y)$：$\{p(a_i/b_j)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$ 看作变量，信源 X：$\{a_1,a_2,\cdots,a_r\}$ 的概率分布 $P(X)$：$\{p(a_i)(i=1,2,\cdots,r)\}$ 看作固定不变的量”这种近似处理的方法时，使平均交互信息量 $I[p(a_i/b_j)]$ 达到最大值，即信道容量 C 的后验概率 $p^*(a_i/b_j)$，就是信源 X：$\{a_1,a_2,\cdots,a_r\}$ 的概率分布 $P(X)$：$\{p(a_i)(i=1,2,\cdots,r)\}$ 时，给定信道 $P(Y/X)$：$\{p(b_j/a_i)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$ 的一般意义下的后验概率 $p(a_i/b_j)$。这是因为对给定信道 $P(Y/X)$：$\{p(b_j/a_i)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$ 来说，当输入信源 X：$\{a_1,a_2,\cdots,a_r\}$ 的概率分布 $P(X)$：$\{p(a_i)(i=1,2,\cdots,r)\}$ 固定不变时，其信道的平均交互信息量 $I(X;Y)=I[p(a_i),p(b_j/a_i)]$ 只有一个确定的值，其最大值也只可能就是这唯一的确定值。达到这唯一确定值的后验概率 $p^*(a_i/b_j)$ 当然只可能就是由(3.36)式所规定的一般意义下的后验概率 $p(a_i/b_j)$。所以，由(3.43)式可知，当变动后验概率 $p(a_i/b_j)$，而固定信源 X：$\{a_1,a_2,\cdots,a_r\}$ 的概率分布 $P(X)$：$\{p(a_i)(i=1,2,\cdots,r)\}$ 时，信道容量

$$C=\max_{p(a_i/b_j)}\{I[p(a_i),p(a_i/b_j)]\}=I[p(a_i),p^*(a_i/b_j)] \tag{3.44}$$

另一方面，同样可把信源 X：$\{a_1, a_2, \cdots, a_r\}$ 的概率分布 $P(X)$：$\{p(a_i)(i=1, 2, \cdots, r)\}$ 当作自变量，把本来要随之发生相应变化的应变量 $p(a_i/b_j)$ 近似的看作是固定不变的量。在采用这种近似处理时，由于 $p(b_j/a_i)$ 和 $p(a_i/b_j)$ 都是固定不变，则式(3.34)所示的平均交互信息量 $I(X;Y)$ 就可以看作是先验概率 $p(a_i)$ 的函数

$$I(X;Y)=I[p(a_i)]$$

由于 $I[p(a_i)]$ 是 $p(a_i)$ 的∩型凸函数，所以可在条件

$$\sum_{i=1}^{r} p(a_i) = 1 \tag{3.45}$$

的约束下，对变量 $p(a_i)$ 求函数 $I[p(a_i)]$ 的条件极大值，以及达到极大值的 $p^*(a_i)(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$

为此，作辅助函数

$$F[p(a_i), \lambda] = I[p(a_i)] + \lambda[\sum_{i=1}^{r} p(a_i) - 1] \tag{3.46}$$

取函数 $F[p(a_i), \lambda]$ 对 $p(a_i)$ 的偏导，并置之为零，得到稳定点方程

$$\frac{\partial}{\partial\ p(a_i)}F[p(a_i), \lambda]=0 \quad (i=1, 2, \cdots, r; j=1, 2, \cdots, s) \tag{3.47}$$

把式(3.34)代入式(3.47)，有

$$\begin{aligned}
\frac{\partial}{\partial\ p(a_i)}F[p(a_i), \lambda] &= \frac{\partial}{\partial\ p(a_i)}\Big\{-\sum_{i=1}^{r} p(a_i)\ln p(a_i) + \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_i)p(b_j/a_i)\ln p(a_i/b_j) \\
&\quad + \lambda\Big[\sum_{i=1}^{r} p(a_i) - 1\Big]\Big\} \\
&= -[\ln p(a_i) + 1] + \sum_{j=1}^{s} p(b_j/a_i)\ln p(a_i/b_j) + \lambda \\
&= -\ln p(a_i) + \sum_{j=1}^{s} p(b_j/a_i)\ln p(a_i/b_j) + \lambda - 1 \\
&= 0 \quad (i = 1, 2, \cdots, r)
\end{aligned} \tag{3.48}$$

即有

$$\ln p(a_i) = \sum_{j=1}^{s} p(b_j/a_i)\ln p(a_i/b_j) + \lambda - 1$$

即

$$p(a_i) = \exp\{\sum_{j=1}^{s} p(b_j/a_i)\ln p(a_i/b_j) + \lambda - 1\} \quad (i = 1, 2, \cdots, r) \tag{3.49}$$

由约束条件式(3.45)，得

$$\begin{aligned}
1 &= \sum_{i=1}^{r} \exp\{\sum_{j=1}^{s} p(b_j/a_i)\ln p(a_i/b_j) + \lambda - 1\} \\
&= \sum_{i=1}^{r} \exp\{\sum_{j=1}^{s} p(b_j/a_i)\ln p(a_i/b_j)\} \cdot e^{(\lambda-1)}
\end{aligned} \tag{3.50}$$

即得

$$e^{(\lambda-1)} = \frac{1}{\sum_{i=1}^{r} \exp\{\sum_{j=1}^{s} p(b_j/a_i)\ln p(a_i/b_j)\}}$$

则可以得到

$$p^*(a_i)=\frac{\exp\{\sum_{j=1}^{s}p(b_j/a_i)\ln p(a_i/b_j)\}}{\sum_{i=1}^{r}\exp\sum_{j=1}^{s}p(b_j/a_i)\ln p(a_i/b_j)}\quad(i=1,2,\cdots,r)\tag{3.51}$$

若令

$$E_i=\exp\{\sum_{j=1}^{s}p(b_j/a_i)\ln p(a_i/b_j)\}\quad(i=1,2,\cdots,r)$$

由式(3.51)，得

$$p^*(a_i)=\frac{E_i}{\sum_{i=1}^{r}E_i}\quad(i=1,2,\cdots,r)\tag{3.52}$$

则信道容量　$$C=\max_{p(a_i)}\{I[p(a_i),p(a_i/b_j)]\}=I[p^*(a_i),p(a_i/b_j)]\tag{3.53}$$

综上所述，式(3.44)是在信源 X：$\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X)$：$\{p(a_i)(i=1,2,\cdots,r)\}$固定不变，变动后验概率 $P(X/Y)$：$\{p(a_i/b_j)(i=1,2,\cdots,r;j=1,2,\cdots,s)\}$的假设前提下，传递概率为 $P(Y/X)$：$\{p(b_j/a_i)(i=1,2,\cdots,r;j=1,2,\cdots,s)\}$的给定信道的信道容量 C 的近似迭代公式。式(3.53)是在后验概率 $P(X/Y)$：$\{p(a_i/b_j)(i=1,2,\cdots,r;j=1,2,\cdots,s)\}$固定不变，变动信源 X：$\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X)$：$\{p(a_i)(i=1,2,\cdots,r)\}$的假设前提下，传递概率为 $P(Y/X)$：$\{p(b_j/a_i)(i=1,2,\cdots,r;j=1,2,\cdots,s)\}$的给定信道的信道容量 C 的近似迭代公式。实际上，由式(3.36)可知，对于传递概率固定为 $p(b_j/a_i)$的给定信道，在变动后验概率 $p(a_i/b_j)$时，先验概率 $p(a_i)$不可能固定不变；在变动先验概率 $p(a_i)$时，不可能后验概率 $p(a_i/b_j)$固定不变。迭代计算法就是用分别单独变动 $p(a_i/b_j)$和 $p(a_i)$的方法，逼近 $p(a_i/b_j)$和 $p(a_i)$同时变动的实际情况，求得信道容量 C 的近似值。

先假定一组 $p(a_i)(i=1,2,\cdots,r)$作为起始值，并记为 $p(a_i)^{(1)}(i=1,2,\cdots,r)$。把 $p(a_i)^{(1)}(i=1,2,\cdots,r)$作为固定值，变动 $p(a_i/b_j)(i=1,2,\cdots,r;j=1,2,\cdots,s)$。由式(3.43)求得后验概率

$$p^*(a_i/b_j)=\frac{p(a_i)^{(1)}p(b_j/a_i)}{\sum_{i=1}^{r}p(a_i)^{(1)}p(b_j/a_i)}=p(a_i/b_j)^{(1)}\quad(i=1,2,\cdots,r;j=1,2,\cdots,s)\tag{3.54}$$

由式(3.44)求得信道容量

$$C=\max_{p(a_i/b_j)}\{I[p(a_i)^{(1)},p(a_i/b_j)]\}=I[p(a_i)^{(1)},p(a_i/b_j)^{(1)}]=C(1,1)$$

再把由式(3.54)所得的 $p(a_i/b_j)^{(1)}$作为固定值，变动先验概率 $p(a_i)$，由式(3.52)求得使平均交互信息量 $I[p(a_i)]$达到最大值的输入信源 X：$\{a_1,a_2,\cdots,a_r\}$的概率分布

$$p^*(a_i)=\frac{\exp\left\{\sum_{j=1}^{s}p(b_j/a_i)\log p\,(a_i/b_j)^{(1)}\right\}}{\sum_{i=1}^{r}\exp\sum_{j=1}^{s}p(b_j/a_i)\log p\,(a_i/b_j)^{(1)}}$$

$$=\frac{E_i^{(1)}}{\sum_{i=1}^{r}E_i^{(1)}}=p\,(a_i)^{(2)}(i=1,\ 2,\ \cdots,\ r)$$

由式(3.53)求得信道容量

$$C=\max_{p(a_i)}\{I[p(a_i),\ p\,(a_i/b_j)^{(1)}]\}=I[p\,(a_i)^{(2)},\ p\,(a_i/b_j)^{(1)}]=C(2,\ 1)$$

以此类推，一般可有 $C(n,\ n)=\max\limits_{p(a_i/b_j)}\{I[p\,(a_i)^{(n)},\ p(a_i/b_j)]\}=I[p\,(a_i)^{(n)},\ p\,(a_i/b_j)^{(n)}]$

$$C(n+1,\ n)=\max_{p(a_i)}\{I[p(a_i),\ p\,(a_i/b_j)^{(n)}]\}=I[p\,(a_i)^{(n+1)},\ p\,(a_i/b_j)^{(n)}]$$

在实际计算中，逐段比较 $p\,(a_i)^{(n)}$ 和 $p\,(a_i)^{(n+1)}$；$p\,(a_i/b_j)^{(n)}$ 和 $p\,(a_i/b_j)^{(n+1)}$；$C(n,\ n)$ 和 $C(n+1,\ n)$ 的值。当 n 次迭代和 $(n+1)$ 次迭代的计算值的差，已小到可以允许的程度，就可认为达到了所求信道容量值 C。

3.9 平均交互信息量的不增性

在实际通信系统中，常需对信道传输的数据作适当处理，若把数据处理装置亦看作是一个信道，这就由两个信道串接，组成了一个串接信道。

设信道 I 和信道 II 串接，组成图 3.16 所示串接信道。信道 I 的输入符号集 X：$\{a_1,\ a_2,\ \cdots,\ a_r\}$ 输出符号集 Y：$\{b_1,\ b_2,\ \cdots,\ b_s\}$，输出符号集 Z：$\{c_1,\ c_2,\ \cdots,\ c_L\}$。又设，信道 I 的传递概率 $P(Y/X)$：$\{p(b_j/a_i)(i=1,\ 2,\ \cdots,\ s)\}$，信道 II 的传递概率 $P(Z/XY)$：$\{p(c_k/a_ib_j)(i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s;\ k=1,\ 2,\ \cdots,\ L)\}$。且假定 $p(a_ib_jc_k)>0(i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots s,\ k=1,\ 2,\ \cdots,\ L)$

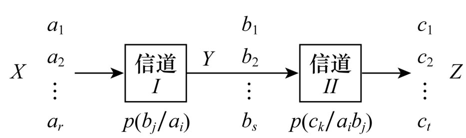


图 3.16 串接信道

由一个信道组成的通信系统只有输入、输出两个随机变量 X 和 Y。由两个信道串接组成的串接信道中，就有 X、Y、Z 三个随机变量，那么，由三个随机变量构成的随机变量序列 (XYZ)，在信息传输上又有什么新的特点和规律呢？

定理 3.5 设由两个信道串接构成随机变量序列 (XYZ)，则

$$I(XY;\ Z)\geq I(Y;\ Z) \tag{3.55}$$

当且仅当

$$P(Z/XY)=P(Z/Y) \tag{3.56}$$

即 (XYZ) 是 Markov 链时，才有

$$I(XY;Z)=I(Y;Z) \tag{3.57}$$

【证明】根据平均交互信息量的定义，有

$$I(XY;Z)=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)\log\frac{p(c_k/a_ib_j)}{p(c_k)} \tag{3.58}$$

而

$$I(Y;Z)=\sum_{j=1}^{s}\sum_{k=1}^{L}p(b_jc_k)\log\frac{p(c_k/b_j)}{p(c_k)}=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)\log\frac{p(c_k/b_j)}{p(c_k)} \tag{3.59}$$

考虑到$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)=1$，以及“底”大于 1 的对数是 ∩ 型凸函数，可有

$$\begin{aligned}I(Y;Z)-I(XY;Z)&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)\log\frac{p(c_k/b_j)}{p(c_k/a_ib_j)}\\&\leqslant\log\left\{\sum_{i}\sum_{j}\sum_{k}p(a_ib_jc_k)\log\frac{p(c_k/b_j)}{p(c_k/a_ib_j)}\right\}\\&=\log\left\{\sum_{i}\sum_{j}\sum_{k}p(a_ib_j)p(c_k/b_j)\right\}=\log1=0\end{aligned} \tag{3.60}$$

即证得

$$I(XY;Z)\geqslant I(Y;Z)$$

由式(3.56)有

$$\frac{p(c_k/b_j)}{p(c_k/a_ib_j)}=1\ (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,L) \tag{3.61}$$

则由式(3.60)有

$$\begin{aligned}I(Y;Z)-I(XY;Z)&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)\log\frac{p(c_k/b_j)}{p(c_k/a_ib_j)}\\&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)\log1=0\end{aligned} \tag{3.62}$$

即证得

$$I(XY;Z)=I(Y;Z)$$

定理 3.6　设由两个信道串接构成随机变量序列(XYZ)，则

$$I(XY;Z)\geqslant I(X;Z) \tag{3.63}$$

当且仅当

$$P(Z/XY)=P(Z/X) \tag{3.64}$$

即(YXZ)是 Markov 链时，才有

$$I(XY;Z)=I(X;Z) \tag{3.65}$$

【证明】　根据平均交互信息量的定义，有

$$I(XY;Z)=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)\log\frac{p(c_k/a_ib_j)}{p(c_k)}$$

而

$$I(X;Z)=\sum_{j=1}^{s}\sum_{k=1}^{L}p(b_jc_k)\log\frac{p(c_k/a_i)}{p(c_k)}=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)\log\frac{p(c_k/a_i)}{p(c_k)} \tag{3.66}$$

考虑到$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)=1$，以及“底”大于1的对数时∩型凸函数，可有

$$\begin{aligned}I(X;Z)-I(XY;Z)&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)\log\frac{p(c_k/a_i)}{p(c_k/a_ib_j)}\\&\leqslant\log\left\{\sum_{i}\sum_{j}\sum_{k}p(a_ib_jc_k)\log\frac{p(c_k/a_i)}{p(c_k/a_ib_j)}\right\}\\&=\log\left\{\sum_{i}\sum_{j}\sum_{k}p(a_ib_j)p(c_k/a_i)\right\}=\log1=0\end{aligned}\tag{3.67}$$

即证得

$$I(XY;Z)\geqslant I(X;Z)$$

由式(3.64)有

$$\frac{p(c_k/a_i)}{p(c_k/a_ib_j)}=1\ (i=1,2,\cdots,r;j=1,2,\cdots,s;k=1,2,\cdots,L)\tag{3.68}$$

则由式(3.67)有

$$\begin{aligned}I(X;Z)-I(XY;Z)&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)\log\frac{p(c_k/a_i)}{p(c_k/a_ib_j)}\\&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{L}p(a_ib_jc_k)\log1=0\end{aligned}\tag{3.69}$$

即证得

$$I(XY;Z)=I(X;Z)$$

定理3.7 设由两个信道串接构成的随机变量序列(XYZ)。当

$$P(Z/XY)=P(Z/Y)\tag{3.70}$$

即当(XYZ)是Markov链时，有

$$I(X;Z)\leqslant I(Y;Z)\tag{3.71}$$

当且仅当

$$P(Z/XY)=P(Z/X)\tag{3.72}$$

即(YXZ)亦是Markov链时，有

$$I(X;Z)=I(Y;Z)\tag{3.73}$$

【证明】由定理3.5可知当随机变量序列(XYZ)是Markov链时，有

$$I(XY;Z)=I(Y;Z)\tag{3.74}$$

由定理3.6可知，在随机变量序列(YXZ)不是Markov链的情况下，有

$$I(XY;Z)\geqslant I(X;Z)\tag{3.75}$$

由式(3.74)和式(3.75)可得，当(XYZ)是Markov链，而(YXZ)不是Markov链的情况下，有

$$I(X;Z)\leqslant I(Y;Z)$$

由定理3.6可知，当(YXZ)是Markov链时，有

$$I(XY;Z)=I(X;Z)\tag{3.76}$$

则由式(3.74)、式(3.75)证得，当(XYZ)和(YXZ)都是Markov链时，即

$$P(Z/XY)=P(Z/Y)$$
$$P(Z/YX)=P(Z/X)$$

时，有

$$I(X;Z)=I(Y;Z)$$

证毕。

对于图 3.16 所示串接信道，在工程上一般把随机变量序列 (XYZ) 看作 Markov 链，整个串接信道的传递概率频 $p(c_k/a_i)(i=1,2,\cdots,r;k=1,2,\cdots,L)$ 求得，即有

$$p(c_k/a_i)=\sum_{j=1}^{s}p(b_j/a_i)p(c_k/b_j)\quad(i=1,2,\cdots,r;k=1,2,\cdots,L)\tag{3.77}$$

这就是说，整个串接信道的信道矩阵 $[P]$（rxL 阶），等于信道 I 的信道矩阵 $[P_I]$（rxs 阶）与信道 II 的信道矩阵 $[P_{II}]$（sxL）阶的连乘，即

$$[P]=[P_I]\cdot[P_{II}]\tag{3.78}$$

若要求随机变量序列 (XYZ) 和 (YXZ) 都是 Markov 链，则要有

$$\begin{cases}P(Z/XY)=P(Z/Y)\\P(Z/YX)=P(Z/X)\end{cases}$$

即

$$\begin{cases}p(c_k/a_ib_j)=p(c_k/b_j)\\p(c_k/a_ib_j)=p(c_k/a_i)\end{cases}$$
$$(i=1,2,\cdots,r;j=1,2,\cdots,s;k=1,2,\cdots,L)\tag{3.79}$$

即必须有

$$p(c_k/b_j)=p(c_k/a_i)$$
$$(i=1,2,\cdots,r;j=1,2,\cdots,s;k=1,2,\cdots,L)\tag{3.80}$$

这就是说，整个串接信道的信道矩阵 $[P]$ 与信道 II 的信道矩阵 $[P_{II}]$ 要完全一样，即要有

$$[P]=[P_I]\cdot[P_{II}]=[P_{II}]\tag{3.81}$$

显然，信道 I 的矩阵 $[P_I]$ 是 (sxs) 阶单位矩阵，是式(3.81)能得到满足的一种情况，这时有

$$\begin{array}{c}\\a_1\\a_2\\\vdots\\a_s\end{array}\begin{array}{c}\begin{array}{cccc}b_1&b_2&\cdots&b_s\end{array}\\\begin{bmatrix}1&0&\cdots&0\\0&1&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&1\end{bmatrix}\end{array}\cdot\begin{array}{c}\\b_1\\b_2\\\vdots\\b_s\end{array}\begin{array}{c}\begin{array}{cccc}c_1&c_2&\cdots&c_L\end{array}\\\begin{bmatrix}p(c_1/b_1)&p(c_2/b_1)&\cdots&p(c_L/b_1)\\p(c_1/b_2)&p(c_2/b_2)&\cdots&p(c_L/b_2)\\\vdots&\vdots&&\vdots\\p(c_1/b_s)&p(c_2/b_s)&\cdots&p(c_L/b_s)\end{bmatrix}\end{array}$$

$$=\begin{array}{c}\\a_1\\a_2\\\vdots\\a_s\end{array}\begin{array}{c}\begin{array}{cccc}c_1&c_2&\cdots&c_L\end{array}\\\begin{bmatrix}p(c_1/b_1)&p(c_2/b_1)&\cdots&p(c_L/b_1)\\p(c_1/b_2)&p(c_2/b_2)&\cdots&p(c_L/b_2)\\\vdots&\vdots&&\vdots\\p(c_1/b_s)&p(c_2/b_s)&\cdots&p(c_L/b_s)\end{bmatrix}\end{array}\tag{3.82}$$

式(3.82)指明，当信道 I 是一一对应的确定关系的一般无噪信道时，随机变量序列(XYZ)与随机变量序列(XYZ)一样，亦是 Markov 链。

由以上分析可以得到这样一个结论：在一般情况下(XYZ)是 Markov 链，输出随机变量 Z 通过信道 II 一个信道，获取关于随机变量 Y 的信息量 $I(Y;Z)$，总是比输出随机变量 Z 通过信道 II、信道 I 两个信道，获取关于随机变量 X 的信息量 $I(X;Z)$ 大，这就是说，随机变量 Y 通过信道 I 到随机变量 X，总是要丢失一部分信息量(如图 3.17)。只有当随机变量序列(YXZ)亦是 Markov 链(如信道 I 是一一对应确定关系的一般无噪信道)时，随机变量 Y 到随机变量 X 才不会丢失信息量，因而从 Z 中获取关于 Y 的信息量 $I(Y;Z)$，才与从 Z 中获取关于 X 的信息量 $I(X;Z)$ 相等。

定理 3.8 设由两个信道串接构成随机变量序列(XYZ)，当

$$P(Z/XY)=P(Z/Y \tag{3.83}$$

即当(XYZ)是 Markov 链时，有

$$I(X;Y)I(X;Z) \tag{3.84}$$

当且仅当

$$P(X/YZ)=P(X/Z) \tag{3.85}$$

即(YZX)亦是 Markov 链时，才有

$$I(X;Y)=I(X;Z) \tag{3.86}$$

【证明】 由概率一般运算规则有

$$\begin{aligned} P(Z/XY)&=\frac{P(XYZ)}{P(XY)}=\frac{P(YZ)P(X/YZ)}{P(Y)P(X/Y)} \\ &=P(Z/Y)\cdot\frac{P(X/YZ)}{P(X/Y)} \end{aligned} \tag{3.87}$$

由式(3.83)，有

$$P(X/ZY)=P(X/Y) \tag{3.88}$$

这表明，当随机变量序列(XYZ)是 Markov 链时，随机变量序列(ZYX)亦是 Markov 链，根据定理 3.5，有

$$I(ZY;X)=I(Y;X) \tag{3.89}$$

在一般情况下，随机变量序列(YZX)不是 Markov 链，根据定理 3.6，有

$$I(ZY;X)I(Z;X) \tag{3.90}$$

由式(3.89)和式(3.90)，有

$$I(Y;X)I(Z;X) \tag{3.91}$$

根据平均交互信息量的交互性，可证得

$$I(X;Y)I(X;Z)$$

当随机变量序列(YZX)亦是 Markov 链时，根据定理 3.6，有

$$I(ZY;X)=I(Z;X) \tag{3.92}$$

由式(3.89)有

$$I(Y;X)=I(Z;X)$$

根据平均交互信息量的交互性，证得

$$I(X;\ Y)=I(X;\ Z)$$

这样定理就得到了证明。

对于图 3.16 所示串接信道，在一般情况下，随机变量序列(XYZ)都可看作是 Markov 链，即随机变量序列(ZYX)亦可看作 Markov 链。如要求随机变量序列(YZX)同时也是 Markov 链，则必须有

$$P(X/ZY)=P(X/Y)$$
$$P(X/YZ)=P(X/Z)$$

即有

$$\begin{cases} p(a_i/c_kb_j)=p(a_i/b_j) \\ p(a_i/c_kb_j)=p(a_i/c_k) \end{cases}$$
$$(i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s;\ k=1,\ 2,\ \cdots,\ L) \tag{3.93}$$

即必须有

$$p(a_i/b_j)=p(a_i/c_k)$$
$$(i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s;\ k=1,\ 2,\ \cdots,\ L) \tag{3.94}$$

而

$$p(a_i/b_j)=\frac{p(a_i)p(b_j/a_i)}{\sum_{i=1}^{r}p(a_i)p(b_j/a_i)} \tag{3.95}$$

$$p(a_i/c_k)=\frac{p(a_i)p(c_k/a_i)}{\sum_{i=1}^{r}p(a_i)p(c_k/a_i)} \tag{3.96}$$

由式(3.94)以及式(3.95)、式(3.96)可知，则必须有

$$p(b_j/a_i)=p(c_k/a_i)$$
$$(i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s;\ k=1,\ 2,\ \cdots,\ L) \tag{3.97}$$

这说明，如果要求随机变量序列(ZYX)是 Markov 链的条件下，随机变量序列(YZX)亦是 Markov 链，则必须要求图 3.16 串接信道的信道矩阵$[P]$与信道 I 的信道矩阵$[P_I]$完全相同，即必须有

$$[P]=[P_I]\cdot[P_{II}]=[P_I] \tag{3.98}$$

显然，信道 II 的信道矩阵$[P_{II}]$是$(s\times s)$阶单位矩阵，是式(3.98)能得到满足的一种情况，这时有

$$
\begin{array}{cc} & \begin{array}{cccc} b_1 & b_2 & \cdots & b_s \end{array} \\ \begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_r \end{array} & \begin{bmatrix} p(b_1/a_1) & p(b_2/a_1) & \cdots & p(b_s/a_1) \\ p(b_1/a_2) & p(b_1/a_2) & \cdots & p(b_s/a_2) \\ \vdots & \vdots & & \vdots \\ p(b_1/a_r) & p(b_1/a_r) & \cdots & p(b_s/a_r) \end{bmatrix} \end{array} \cdot \begin{array}{cc} & \begin{array}{cccc} c_1 & c_2 & \cdots & c_s \end{array} \\ \begin{array}{c} b_1 \\ b_2 \\ \vdots \\ b_s \end{array} & \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \tag{3.99}
$$

$$
= \begin{array}{cc} & \begin{array}{cccc} c_1 & c_2 & \cdots & c_s \end{array} \\ \begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_r \end{array} & \begin{bmatrix} p(b_1/a_1) & p(b_2/a_1) & \cdots & p(b_s/a_1) \\ p(b_1/a_2) & p(b_1/a_2) & \cdots & p(b_s/a_2) \\ \vdots & \vdots & & \vdots \\ p(b_1/a_r) & p(b_1/a_r) & \cdots & p(b_s/a_r) \end{bmatrix} \end{array}
$$

这表明，当信道 II 是一一对应的确定关系的一般无噪信道时，随机变量序列(YZX)与随机变量序列(ZYX)一样，亦是 Markov 链。

由以上分析可得到这样一个结论：对于图 3.16 串接信道，在一般情况下，由于随机变量序列(XYZ)可看作是 Markov 链，所以随机变量序列(ZYX)亦可看作是 Markov 链。随机变量 X 通过信道 I 获取关于随机变量 Y 的信息量 $I(X;Y)$，总比随机变量 X 通过信道 I 和信道 II 两个信道，获取关于随机变量 Z 的信息量 $I(X;Z)$大。这就是说，随机变量 Y 通过信道 II 到随机变量 Z，总要丢失一部分信息量(如图 3.18)。只有当随机变量序列(YZX)亦是 Markov 链(如信道 II 是一一对应确定关系的一般无噪信道时)，随机变量 Y 到随机变量 Z 才不会丢失信息量。因而从随机变量 X 中获取关于随机变量 Y 的信息量 $I(X;Y)$，才与从随机变量 X 中获取关于随机变量 Z 的信息量 $I(X;Z)$相等。

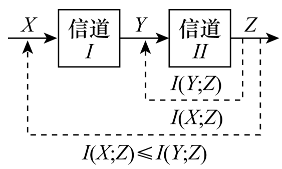


图 3.17　关注信道 II

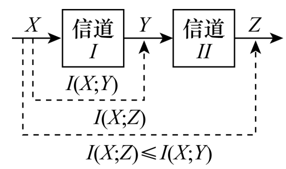


图 3.18　关注信道 I

【例 3.9】设信道 I 和信道 II 相接，组成图 3.19 串接信道。若随机变量序列(XYZ)是 Markov 链，试证明 $I(X;Z)=I(X;Y)$。

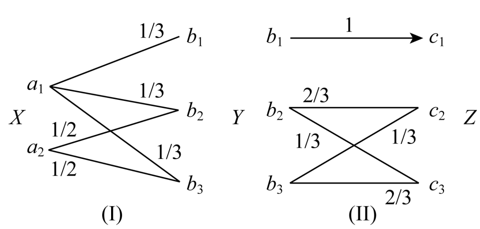


图 3.19　(XYZ)构成 Markov 链

解：因为随机变量序列(XYZ)是 Markov 链，由图 3.19 可知，串接信道的信道矩阵

$$[P]=\begin{array}{c} \\ \begin{matrix}a_1\\ a_2\end{matrix}\end{array}\begin{array}{c}\begin{matrix}b_1 & b_2 & b_3\end{matrix}\\ \begin{bmatrix}1/3 & 1/3 & 1/3\\ 0 & 1/2 & 1/2\end{bmatrix}\end{array}\cdot\begin{array}{c} \\ \begin{matrix}b_1\\ b_2\\ b_3\end{matrix}\end{array}\begin{array}{c}\begin{matrix}c_1 & c_2 & c_3\end{matrix}\\ \begin{bmatrix}1 & 0 & 0\\ 0 & 2/3 & 1/3\\ 0 & 1/3 & 2/3\end{bmatrix}\end{array}=\begin{array}{c} \\ \begin{matrix}a_1\\ a_2\end{matrix}\end{array}\begin{array}{c}\begin{matrix}c_1 & c_2 & c_3\end{matrix}\\ \begin{bmatrix}1/3 & 1/3 & 1/3\\ 0 & 1/2 & 1/2\end{bmatrix}\end{array}$$

由于串接信道的信道矩阵$[P]$与信道 I 的信道矩阵$[P_I]$完全一致，所以有

$$p(b_j/a_i)=p(c_k/a_i)\quad(i=1,\ 2;\ j=1,\ 2,\ 3;\ k=1,\ 2,\ 3)$$

根据式(3.97)、(3.96)、(3.95)、(3.94)、(3.93)，随机变量序列(YZX)亦是 Markov 链。根据定理 3.8，可证得

$$I(X;\ Y)=I(X;\ Z)$$

这一特例说明，在(XYZ)是 Markov 链的条件下，信道 II 是一一对应确定关系的一般无噪信道，是式(3.84)中等式(3.85)成立的必要条件，但不是充分条件。

推论　设有两个信道串接构成随机变量序列(XYZ)。若随机变量序列(XYZ)是 Markov 链，则有

$$I(X;\ Z)\leqslant\begin{cases}I(Y;\ Z)\\ I(X;\ Y)\end{cases}\tag{3.100}$$

【证明】因为随机变量序列(XYZ)是 Markov 链，则由定理 3.7 有

$$I(X;\ Z)\leqslant I(Y;\ Z)$$

又因为当(XYZ)是 Markov 链时，随机变量序列(ZYX)亦是 Markov 链，由定理 3.8，有 $I(X;\ Z)\leqslant I(X;\ Y)$

即证得，当随机变量序量(XYZ)是 Markov 链时，有

$$I(X;\ Z)\leqslant\begin{cases}I(Y;\ Z)\\ I(X;\ Y)\end{cases}$$

这样，推论就得到了证明。平均交互信息量 $I(X;\ Z)$与 $I(Y;\ Z)$和 $I(X;\ Y)$之间的关系综合表示为图 3.20 所示。

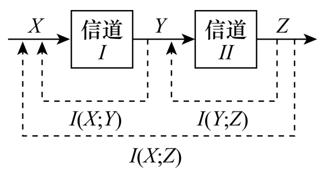


图 3.20　数据处理

推论指出，若以 Z 作为观察点，把信道 I 看作是一个数据处理装置(如图 3.21)，随机变量序列(XYZ)是 Markov 链。因随机变量 Y 经过处理后，变成随机变量 X，则从随机变量 Z 中获取关于随机变量 X 的平均交互信息量 $I(X;\ Z)$，不会超过数据处理前从随机变量 Z 中获取关于随机变量 Y 的平均交互信息量 $I(Y;\ Z)$，最多两者相等。这就是说，把随机变量 Y 变成随机变量 X 的数据处理过程，总是要丢失一部分信息。只有当随机变量序列(YXZ)亦是 Markov 链时(数据处理是一一对应确定关系时)，数据处理过程才不会丢失信息。

推论又指出，若以 X 作为观察点，把信道 II 看作是一个数据处理装置，如图 3.22 所示，随机变量序列(ZYX)是 Markov 链。因随机变量 Y 经过数据处理后，变成随机变量 Z，从随机变量 X 中获取关于随机变量 Z 的平均交互信息量 $I(X;Z)$，不会超过数据处理前从随机变量 X 中获取关于随机变量 Y 的平均交互信息量 $I(X;Y)$，最多两者相等。这就是说，把随机变量 Y 变成随机变量 Z 的数据处理过程，总是要丢失一部分信息。只有当随机变量序列(YZX)亦是 Markov 链时(数据处理是一一对应的确定关系时)，数据处理过程才不会丢失信息。

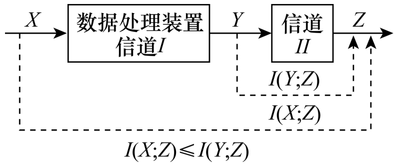


图 3.21　串接信道

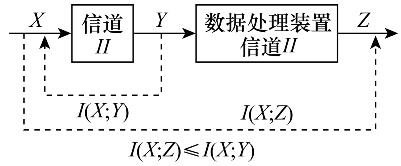


图 3.22　串接信道

综合推论所指出的两方面的结论，可以得到一个总的结论：任何无源数据处理过程都要丢失一部分信息量，最多不丢失信息量，但一定不会增加信息量。由定理 3.5、3.6、3.7、3.8 及其推论所阐明的平均交互信息量的不增性，一般通称为数据处理定理。

数据处理定理是信息传输的重要规律。把数据处理定理与前面讨论的平均交互信息量的极值性结合起来，就可导出信息传输和处理的一个完整的概念，如图 3.23(a)和(b)所示。对于信源 X 来说，经信息传输或信息处理，最终从随机变量 Q 中获取关于信源 X 的平均交互信息量 $I(X;Q)$，绝对不会超过信源 X 本身含有的平均信息量 $H(X)$，最多等于信源 X 本身含有的平均信息量 $H(X)$。信息所经过的传输信道或数据处理装置越多，丢失的信息量就可能越多。即有

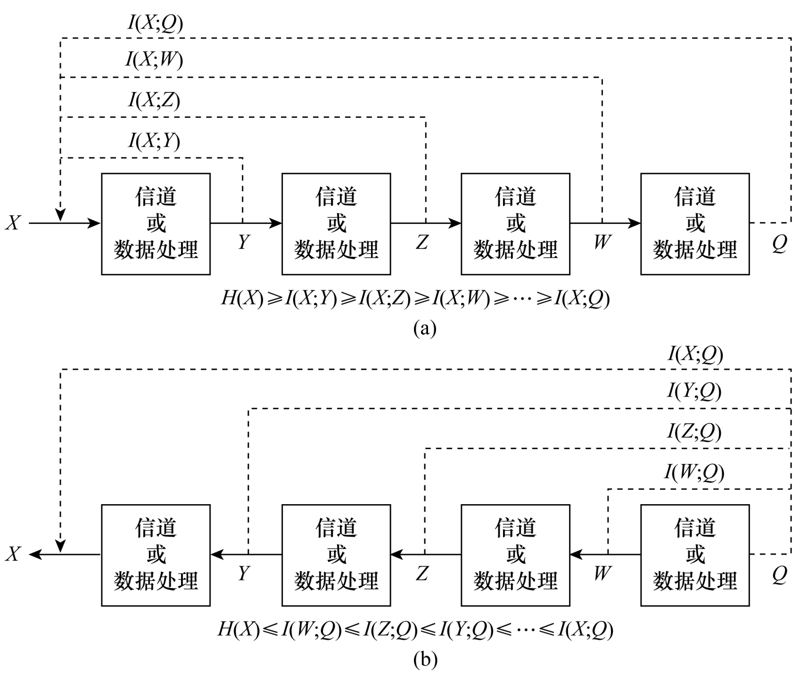


图 3.23(b)

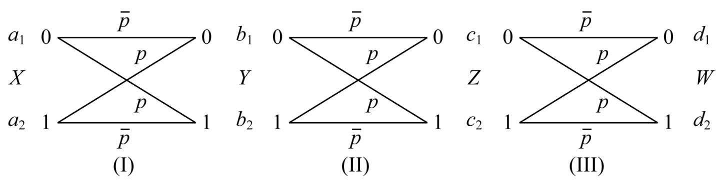
$$H(X) \geqslant I(X;Y) \geqslant I(X;Z) \geqslant I(X;W) \geqslant \cdots \geqslant I(X;Q) \tag{3.101}$$

对于随机变量 Q 来说，经信息传输或信息处理，最终从随机变量 X 中获取关于随机变量 Q 的平均交互信息量 $I(Q;X)$，绝不会超过随机变量 Q 本身含有的平均信息量 $H(Q)$，最多等于 $H(Q)$。信息所经过的传输信道或数据处理装置越多，丢失的信息量就可能越多。即有

$$H(Q) \geqslant I(W;Q) \geqslant I(Z;Q) \geqslant I(Y;Q) \geqslant \cdots \geqslant I(X;Q) \tag{3.102}$$

在传输或处理过程中，一旦在某一环节(信道或处理装置)丢失一部分信息，以后的系统不管怎样传输或处理，如不能接触到丢失信息环节的输入端(如多次测量)，就不能再恢复已丢失的信息。

【例 3.10】 设有三个二进制对称信道(BSC)，串接组成图 3.24 所示串接信道。随机变量序列($XYZW$)是 Markov 链。若信源 X：{0，1}是等概信源，试求平均交互信息量 $I(X;Y)$、$I(X;Z)$、$I(X;W)$，并比较它们的大小。



$(0<\bar{p}, p<1; \bar{p}+p=1; \bar{p}>p)$

图 3.24　三个信道串联

解：(1)因为信道 I 的矩阵为

$$(P_I) = \begin{array}{c} \\ 0 \\ 1 \end{array}\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix} \end{array}$$

信源 X 的概率分布为

$$P\{X=0\} = p(0) = \frac{1}{2};\ P\{X=1\} = p(1) = \frac{1}{2}$$

所以，随机变量 Y 的概率分布为

$$\begin{aligned} P\{Y=0\} &= P\{X=0\} \cdot P\{Y=0/X=0\} + P\{X=1\} \cdot P\{Y=0/X=1\} \\ &= p(0)\bar{p} + p(1)p = \frac{1}{2}(\bar{p}+p) = \frac{1}{2} \\ P\{Y=1\} &= P\{X=0\} \cdot P\{Y=1/X=0\} + P\{X=1\} \cdot P\{Y=1/X=1\} \\ &= p(0)p + p(1)\bar{p} = \frac{1}{2}(\bar{p}+p) = \frac{1}{2} \end{aligned}$$

则随机变量 Y 的熵

$$H(Y) = H\left(\frac{1}{2}, \frac{1}{2}\right) = 1\ (比特/符号)$$

由信道 I 的信道矩阵$[P_I]$，有

$$
\begin{aligned}
H(Y/X) &= -\sum_{i=1}^{2}\sum_{j=1}^{2} p(a_i)p(b_j/a_i)\log p(b_j/a_i) \\
&= -\{[p(0)\bar{p}\log\bar{p} + p(0)p\log p] + [p(1)p\log p + p(1)\bar{p}\log\bar{p}]\} \\
&= p(0)[-(\bar{p}\log\bar{p} + p\log p)] + p(1)[-(\bar{p}\log\bar{p} + p\log p)] \\
&= (p(0) + p(1))[-(\bar{p}\log\bar{p} + p\log p)] \\
&= -[\bar{p}\log\bar{p} + p\log p] = H(\bar{p},\ p) = H(p)
\end{aligned}
$$

所以，信道 I 的平均交互信息量

$$
\begin{aligned}
I(X;\ Y) &= H(Y) - H(Y/X) \\
&= 1 - H(p) \\
&= 1 - H\left[\frac{1}{2} - \frac{1}{2}(1-2p)\right] \quad (\text{比特/符号}) \qquad (3.103)
\end{aligned}
$$

（2）因为随机变量序列（$XYZW$）是 Markov 链，所以信道 I、信道 II 的串接信道（$X-Z$）的信道矩阵，等于信道 I 矩阵 $[P_I]$ 与信道 II 矩阵 $[P_{II}]$ 的连乘，即有

$$
[P]_{X-Z} = \begin{array}{c} \\ 0 \\ 1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix} \end{array} \cdot \begin{array}{c} \\ 0 \\ 1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix} \end{array} = \begin{array}{c} \\ 0 \\ 1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} \bar{p}^2+p^2 & 2\bar{p}p \\ 2\bar{p}p & \bar{p}^2+p^2 \end{bmatrix} \end{array}
$$

随机变量 Z 的概率分布为

$$
\begin{aligned}
P\{Z=0\} &= P\{X=0\} \cdot P\{Z=0/X=0\} + P\{X=1\} \cdot P\{Z=0/X=1\} \\
&= p(0)(\bar{p}^2+p^2) + p(1)2\bar{p}p = \frac{1}{2}\{(\bar{p}^2+p^2) + 2\bar{p}p\} = \frac{1}{2}
\end{aligned}
$$

$$
P\{Z=1\} = 1 - P\{Z=0\} = \frac{1}{2}
$$

则得随机变量 Z 的熵

$$
H(Z) = H\left(\frac{1}{2},\ \frac{1}{2}\right) = 1 \quad \text{比特/符号}
$$

由串接信道（$X-Z$）的信道矩阵 $[P]_{X-Z}$，有

$$
\begin{aligned}
H(Z/X) &= -\sum_{i=1}^{2}\sum_{k=1}^{2} p(a_i)p(c_k/a_i)\log p(c_k/a_i) \\
&= -\{[p(0)(\bar{p}^2+p^2)\log(\bar{p}^2+p^2) + p(0)(2\bar{p}p)\log(2\bar{p}p)] \\
&\quad + [p(1)(2\bar{p}p)\log(2\bar{p}p) + p(1)(\bar{p}^2+p^2)\log(\bar{p}^2+p^2)]\} \\
&= p(0)\{-[(\bar{p}^2+p^2)\log(\bar{p}^2+p^2) + (2\bar{p}p)\log(2\bar{p}p)]\} \\
&\quad + p(1)\{-[(2\bar{p}p)\log(2\bar{p}p) + (\bar{p}^2+p^2)\log(\bar{p}^2+p^2)]\} \\
&= [p(0)+p(1)]\{-[(\bar{p}^2+p^2)\log(\bar{p}^2+p^2) + (2\bar{p}p)\log(2\bar{p}p)]\} \\
&= H[2\bar{p}p] = H\left[\frac{1}{2} - \frac{1}{2}(1-2p)^2\right]
\end{aligned}
$$

则

$$I(X;\ Z)=H(Z)-H(Z/X)=1-H(2\bar{p}p)$$
$$=1-H\left[\frac{1}{2}-\frac{1}{2}(1-2p)^2\right]\quad（比特/符号）\tag{3.104}$$

同理可得

$$I(X;\ W)=1-H\left[\frac{1}{2}-\frac{1}{2}(1-2p)^3\right]\tag{3.105}$$

则可有

$$I(X;\ Y)\geqslant I(X;\ Z)\geqslant I(X;\ W)\tag{3.106}$$

（3）一般地，设有 N（大于 1 的正整数）个二进制对称信道（BSC），串接成如图 3.25 所示的串接信道。当信源 X 是等概分布时，同理可证：

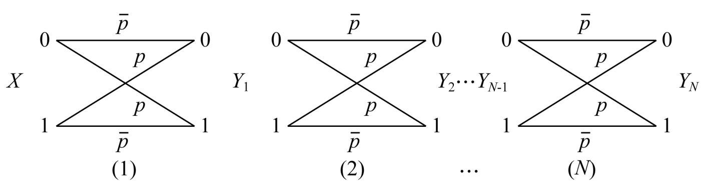


图 3.25　串接信道

$$I(X;\ Y_N)=1-H\left[\frac{1}{2}-\frac{1}{2}(1-2p)^N\right]\quad（比特/符号）\tag{3.107}$$
$$(N=1,\ 2,\ 3,\ 4,\ 5,\ \cdots)$$

则有

$$I(X;\ Y_{N-1})\geqslant I(X;\ Y_N)\qquad(N=2,\ 3,\ 4,\ 5,\ 6,\ \cdots)\tag{3.108}$$

（3.108）式是由 N 个二进制对称信道（BSC）组成的串接信道，当信源 X：$\{0,\ 1\}$ 为等概信源时，数据处理定理的具体表现。

实践：信道容量的迭代算法

【已知】信源符号个数 r、信宿符号个数 s、信道转移概率矩阵 $P=(p_{ji})^{r\times s}$。

【算法】

① 初始化信源分布 $p_i=\dfrac{1}{r}$，循环变量 $k=1$，门限 Δ，$C^{(0)}=-\infty$。

② $$\varphi_{ij}^{(k)}=\frac{p_i^{(k)}p_{ji}}{\sum\limits_{i=1}^{r}p_i^{(k)}p_{ji}}$$

③ $$p_i^{(k+1)}=\frac{\exp\left(\sum\limits_{j=1}^{s}p_{ji}\ln\varphi_{ij}^{(k)}\right)}{\sum\limits_{i=1}^{r}\exp\left(\sum\limits_{j=1}^{s}p_{ji}\ln\varphi_{ij}^{(k)}\right)}$$

④ $C^{(k+1)}=\log\left[\sum_{i=1}^{r}\exp\left(\sum_{j=1}^{s}p_{ji}\ln\varphi_{ij}^{(k)}\right)\right]$

⑤ 若$\dfrac{|C^{(k+1)}-C^{(k)}|}{C^{(k+1)}}>\Delta$，则 $k=k+1$，转第②步。

⑥ 输出 $\overline{P}^{*}=p_i^{(k+1)}$ 和 $C^{(k+1)}$，终止。

【要求】

（1）输入：任意的一个信道转移概率矩阵。信源符号个数、信宿符号个数和每个具体的转移概率在运行时从键盘输入。

（2）输出：最佳信源分布 $\overline{P}^{*}$，信道容量 C。

本 章 要 点

1. 离散信道模型

$$X\left\{\begin{matrix}a_1\\a_2\\\cdots\\a_r\end{matrix}\right.\longrightarrow P(b_j/a_i)\longrightarrow\left.\begin{matrix}b_1\\b_2\\\cdots\\b_s\end{matrix}\right\}Y$$

2. 信道的交互信息量

$$\begin{aligned}I(a_i;\ b_j)&=\log\frac{p(a_i/b_j)}{p(a_i)}\\&=\log\frac{p(a_ib_j)}{p(a_i)p(b_j)}\\&=\log\frac{p(b_i/a_j)}{p(b_i)}\end{aligned}$$

3. 条件交互信息量

$$\begin{aligned}I(a_i;\ b_j/c_k)&=\log\frac{p(a_i/b_jc_k)}{p(a_i/c_k)}\\&=\log\frac{1}{p(a_i/c_k)}-\log\frac{1}{p(a_i/b_jc_k)}\\&=I(a_i/c_k)-I(a_i/b_jc_k)\end{aligned}$$

4. 平均交互信息量

$$I(X;\ Y)=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)I(a_i;\ b_j)=I(Y;\ X)$$

5. 平均互信息的性质

非负性　$I(X;\ Y)\geqslant0$

极值性

$$I(X;\ Y)\leqslant H(X)$$

$$I(X;Y) \leqslant H(Y)$$

凸函数性

平均互信息量 $I(X;Y)$，在信道转移概率 $p(b_j/a_i)$ 给定条件下，是输入随机变量 X 的概率分布 $p(X)$：$\{p(a_i), i=1, 2, \cdots, r\}$ 的∩型凸函数。

平均交互信息量 $I(X;Y)$，在信源概率分布 $P(a_i)$ 给定的条件下，是信道转移概率 $p(Y/X)$：$\{p(b_j/a_i), i=1, 2, \cdots, r; j=1, 2, \cdots, s\}$ 的∩型凸函数。

6. 定义　信道容量

$$C = R_{\max} = \max_{p(x)}\{I(X;Y)\} \quad (\text{比特/符号})$$

$$C_t = R_{t\max} = \frac{1}{t}\max\{I(X;Y)\} \quad (\text{比特/秒})$$

7. 离散对称信道容量

$$\begin{aligned} C &= \max_{P(x)}\{I(X;Y)\} = \max_{P(x)}\{H(Y) - H(Y/X)\} \\ &= \max_{P(x)}\{H(Y)\} - H(p_1', p_2', \cdots, p_s') \\ &= \log s - H(p_1', p_2', \cdots, p_s') = \log s - H(P\text{ 的行矢量}) \end{aligned}$$

8. 准对称离散信道容量

$$C = H(Y) - H(p_1', p_2', \cdots, p_s') = \log r - \sum_{k=1}^{n} N_k \log M_k - H(p_1', p_2', \cdots, p_s')$$

9. 数据处理定理

信息所经过的传输信道或数据处理装置越多，丢失的信息量就可能越多。即有

$$H(Q) \geqslant I(W;Q) \geqslant I(Z;Q) \geqslant I(Y;Q) \geqslant \cdots \geqslant I(X;Q)$$

习　题

3.1　设信源 $[X \cdot P]$：$\begin{cases} X: & a_1 & a_2 \\ P(X): & 0.7 & 0.3 \end{cases}$ 通过一信道，信道的输出随机变量 Y 的符号集 Y：$\{b_1, b_2\}$，信道矩阵

$$[P] = \begin{array}{c} \\ a_1 \\ a_2 \end{array}\!\!\begin{array}{c} \begin{array}{cc} b_1 & b_2 \end{array} \\ \begin{bmatrix} 5/6 & 1/6 \\ 1/4 & 3/4 \end{bmatrix} \end{array}$$

试求：

(1) 信源 X 中的符号 a_1 和 a_2 分别含有的自信量；

(2) 收到消息 $Y=b_1$，$Y=b_2$ 后，获得关于 a_1、a_2 的交互信息量：$I(a_1;b_1)$、$I(a_1;b_2)$、$I(a_2;b_1)$、$I(a_2;b_2)$；

(3) 信源 X 和信宿 Y 的信息熵；

(4) 信道疑义度 $H(X/Y)$ 和噪声熵 $H(Y/X)$；

(5) 接收到消息 Y 后获得的平均交互信息量 $I(X;Y)$。

3.2　某二进制对称信道，其信道矩阵是

$$[P]=\begin{array}{c|cc} & 0 & 1 \\ 0 & 0.98 & 0.02 \\ 1 & 0.02 & 0.98 \end{array}$$

设该信道以 1500 个二进制符号/秒的速度传输输入符号。现有一消息序列共有 14000 个二进制符号，并设在这消息中 $p(0)=p(1)=\frac{1}{2}$。问从信息传输的角度来考虑，10 秒钟内能否将这消息序列无失真地传送完？

3.3　有两个二元随机变量 X 和 Y，它们的联合概率为：$P[X=0,\ Y=0]=\frac{1}{8}$；$P[X=0,\ Y=1]=\frac{3}{8}$；$P[X=1,\ Y=1]=\frac{1}{8}$；$P[X=1,\ Y=0]=\frac{3}{8}$。定义另一随机变量 $Z=XY$，试计算：

(1) $H(X)$，$H(Y)$，$H(Z)$，$H(XZ)$，$H(YZ)$，$H(XYZ)$；

(2) $H(X/Y)$，$H(Y/X)$，$H(X/Z)$，$H(Z/X)$，$H(Y/Z)$，$H(Z/Y)$，$H(X/YZ)$，$H(Y/XZ)$，$H(Z/XY)$；

(3) $I(X;Y)$，$I(X;Z)$，$I(Y;Z)$，$I(X;Y/Z)$，$I(Y;Z/X)$，$I(X;Z/Y)$。

3.4　已知信源 X 的信源空间为

$$[X\cdot P]:\begin{cases} X: & a_1 \quad a_2 \quad a_3 \quad a_4 \\ P(X): & 0.1 \quad 0.3 \quad 0.2 \quad 0.4 \end{cases}$$

某信道的信道矩阵为

$$\begin{array}{c|cccc} & b_1 & b_2 & b_3 & b_4 \\ a_1 & 0.2 & 0.3 & 0.1 & 0.4 \\ a_2 & 0.6 & 0.2 & 0.1 & 0.1 \\ a_3 & 0.5 & 0.2 & 0.1 & 0.2 \\ a_4 & 0.1 & 0.3 & 0.4 & 0.2 \end{array}$$

试求：

(1)“输入 a_3，输出 b_2”的概率；

(2)“输出 b_4”的概率；

(3)“收到 b_3 的条件下，推测输入 a_2”的概率。

3.5　已知从符号 B 中获取关于符号 A 的信息量是 1 比特，当符号 A 的先验概率 $P(A)$ 为下列各值时，分别计算收到 B 后推测 A 的后验概率应是多少？

(1) $P(A)=10^{-2}$；

(2) $P(A)=\frac{1}{32}$；

(3) $P(A)=0.5$。

3.6　某信源发出 8 种消息，它们的先验概率以及相应的码字如表 3.2 所示。以 a_4 为

例，试求：

表 3.2　题 3.6

消息	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
概率	1/4	1/4	1/8	1/8	1/16	1/16	1/16	1/16
码字	000	001	010	011	100	101	110	111

（1）在 $W_4=011$ 中，接到第一个码符号“0”后获得关于 a_4 的信息量 $I(a_4;\ 0)$；

（2）在收到“0”的前提下，从第二个码符号“1”中获取关于 a_4 的信息量 $I(a_4;\ 1/0)$；

（3）在收到“01”的前提下，从第三个码符号“1”中获取关于 a_4 的信息量 $I(a_4;\ 1/01)$；

（4）从码字 $W_4=011$ 中获取关于 a_4 的信息量 $I(a_4;\ 011)$。

3.7　证明要使信道的疑义度 $H(X/Y)=0$ 的充分必要条件是信道矩阵 $[P]$ 中每列有一个，也只有一个非零元素。

3.8　把 n 个二进制对称信道串接起来，每个二进制对称信道的错误传递概率为 $p(0<p<1)$，试证明：整个串接信道的错误传递概率 $p_n=\dfrac{1}{2}[1-(1-2p)^n]$。再证明：当 $n\to\infty$ 时，$\lim\limits_{n\to\infty} I(X_0;\ X_n)=0$。信道串接如图 P3.1 所示。

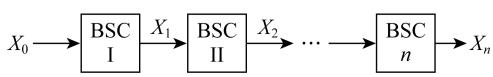


图 P3.1　题 3.8

3.9　若有二个串接的离散信道，它们的信道矩阵都是

$$[P]=\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

并设 I 信道输入随机变量 X 的符号集 X：$\{a_1,\ a_2,\ a_3,\ a_4\}$，且 $p(a_1)=p(a_2)=p(a_3)=p(a_4)=\dfrac{1}{4}$。又设随机变量序列 (XYZ) 是马氏链，如图 P3.2 所示。试求：$I(X;\ Z)$ 和 $I(X;\ Y)$，并比较它们的大小。

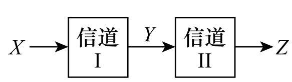


图 P3.2　题 3.9

3.10　试求以下各信道矩阵代表的信道的信道容量：

（1）
$$
[P_1]=\begin{array}{c} \\ a_1\\ a_2\\ a_3\\ a_4 \end{array}\begin{array}{c} \begin{array}{cccc} b_1 & b_2 & b_3 & b_4 \end{array}\\ \begin{bmatrix} 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}
$$

（2）
$$
[P_2]=\begin{array}{c} \\ a_1\\ a_2\\ a_3\\ a_4\\ a_5\\ a_6 \end{array}\begin{array}{c} \begin{array}{ccc} b_1 & b_2 & b_3 \end{array}\\ \begin{bmatrix} 1 & 0 & 0\\ 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 1 \end{bmatrix} \end{array}
$$

（3）
$$
[P_3]=\begin{array}{c} \\ a_1\\ a_2\\ a_3 \end{array}\begin{array}{c} \begin{array}{cccccccccc} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 & b_9 & b_{10} \end{array}\\ \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0.3 & 0.7 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0.2 & 0.1 & 0.3 \end{bmatrix} \end{array}
$$

3.11　设二进制对称信道的信道矩阵为

$$
[P]=\begin{array}{c} \\ 0\\ 1 \end{array}\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array}\\ \begin{bmatrix} 3/4 & 1/4\\ 1/4 & 3/4 \end{bmatrix} \end{array}
$$

（1）若 $p(0)=\dfrac{2}{3}$，$p(1)=\dfrac{1}{3}$，求 $H(X)$，$H(X/Y)$，$H(Y/X)$ 和 $I(X;Y)$；

（2）求该信道的信道容量及其达到信道容量的输入概率分布。

3.12　设某信道的信道矩阵为

$$
[P]=\begin{array}{c} \\ a_1\\ a_2\\ a_3\\ a_4\\ a_5 \end{array}\begin{array}{c} \begin{array}{ccccc} a_1 & a_2 & a_3 & a_4 & a_5 \end{array}\\ \begin{bmatrix} 0.6 & 0.1 & 0.1 & 0.1 & 0.1\\ 0.1 & 0.6 & 0.1 & 0.1 & 0.1\\ 0.1 & 0.1 & 0.6 & 0.1 & 0.1\\ 0.1 & 0.1 & 0.1 & 0.6 & 0.1\\ 0.1 & 0.1 & 0.1 & 0.1 & 0.6 \end{bmatrix} \end{array}
$$

试求：

（1）该信道的信道容量 C；

（2）$I(a_3;Y)$；

（3）$I(a_5;Y)$。

3.13　设某信道的信道矩阵为

$$
[P]=\begin{array}{c} \\ a_1 \\ a_2 \end{array}\begin{array}{c} \begin{array}{cccc} b_1 & b_2 & b_3 & b_4 \end{array} \\ \begin{bmatrix} 1/3 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{bmatrix} \end{array}
$$

试求：

（1）该信道的信道容量 C；

（2）$I(a_1;Y)$；

（3）$I(a_2;Y)$。

3.14　求下列二个信道的信道容量，并加以比较（其中 $0<p$，$q<1$，$p+q=1$）。

（1）$[P_1]=\begin{bmatrix} p-\delta & q-\delta & 2\delta \\ q-\delta & p-\delta & 2\delta \end{bmatrix}$；

（2）$[P_2]=\begin{bmatrix} 2\delta & 0 & p-\delta & q-\delta \\ 0 & 2\delta & q-\delta & p-\delta \end{bmatrix}$。

3.15　设某信道的信道矩阵为

$$
[P]=\begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-\varepsilon & \varepsilon \\ 0 & \varepsilon & 1-\varepsilon \end{bmatrix} \end{array}
$$

试求：

（1）信道容量 C；

（2）达到信道容量 C 时的输入概率分布；

（3）当 $\varepsilon=0$，$\varepsilon=\dfrac{1}{2}$ 时的信道容量 C。