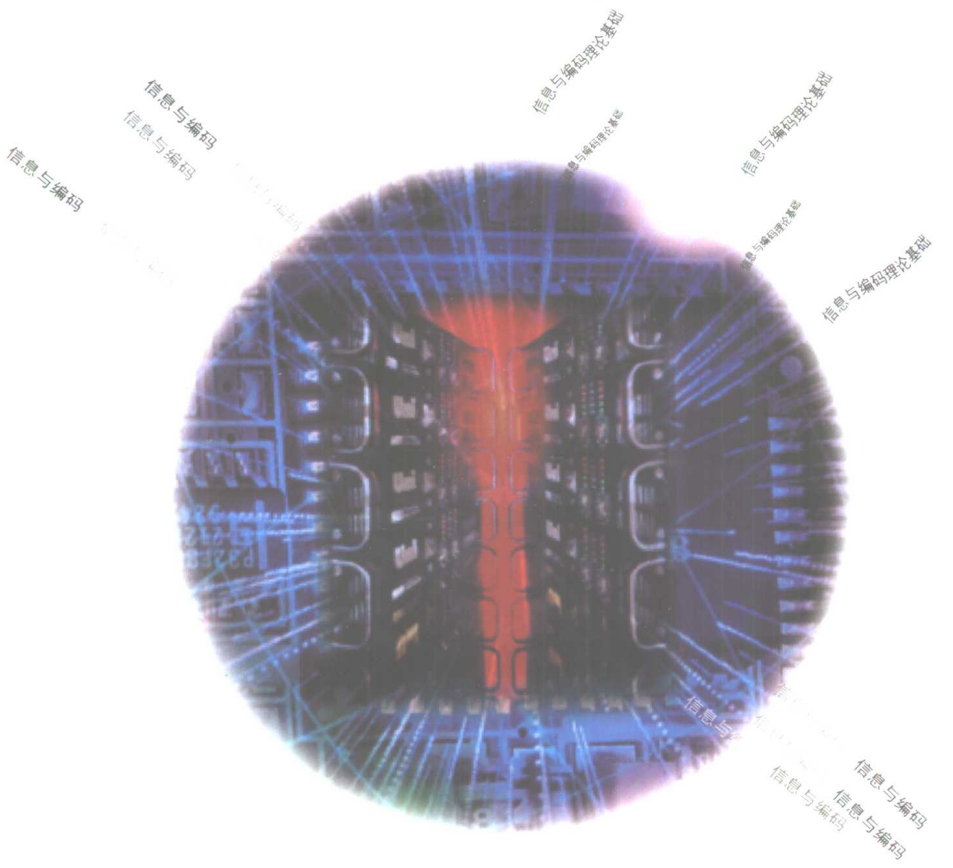
上海普通高校“九五”重点教材

信息与编码理论基础

万旺根 余小清 编著



上海大学出版社

上海普通高校“九五”重点教材
世界银行贷款资助项目
上海市教育委员会组编

信息与编码理论基础

万旺根　余小清　编著

上海大学出版社
·上 海·

内 容 简 介

本书在阐述信息论基本概念的基础上，结合当今信息产业与信息社会的实际内涵，详细论述了信源无失真编码理论、信道特性及信道容量、信源有失真编码理论，阐明了纠错编码的基本概念、线性分组码及循环码的编译原理，并配以适当的例题。

全书深入浅出、通俗易懂，对数学推导作了必要的简化，并着重解释其物理含义，具有一定的可读性和实用性。书中包含了概率论和通信的基本知识，内容可以自成一体，读者可以在无需具备太多的其他预备知识的条件下，较容易地读懂本书。

本书可作为通信与信息类、计算机类、电子类等学科的大学本科高年级以及研究生的教材，也适合于一般的工程技术人员及其他对信息科学感兴趣的人员学习使用。

图书在版编目 (CIP) 数据

信息与编码理论基础 / 万旺根，余小清编著. —上海：上海大学出版社，2000.12
ISBN 7-81058-254-2

Ⅰ.信… Ⅱ.①万…②余… Ⅲ.①信息论②信源编码-通信理论 Ⅳ.TN911.2

中国版本图书馆 CIP 数据核字 (2000) 第 59025 号

上海大学出版社出版发行
(上海市延长路 149 号 邮政编码：200072)
复旦大学印刷厂印刷 各地新华书店经销
开本：787×1092 1/16 印张：8 字数：179 千字
2000 年 12 月第 1 版 2000 年 12 月第 1 次印刷
印数：1～2100
定价：13.50 元

前　言

信息论作为一门经典理论，从1948年发展至今，已取得了不少新的研究成果，在信息处理、通信等研究领域起着越来越重要的作用，对人们认识和研究信息的主要特征，了解信息的基本概念，掌握信息处理的基本方法，以及各种信息传输手段等提供了很好的理论依据。目前，作为我国六大支柱产业之一的信息产业，在我们的日常生活和社会活动中发挥着越来越重要的影响，信息处理的研究成果将直接改善和改变人们的生活方式和社会活动方式。信息与编码理论作为信息处理的基础理论对信息处理的研究起着非常重要的指导作用，它虽然没有给出具体的编解码方法，但给出了这些方法所应遵循的理论框架及其性能极限值。

本书作者早期作为西北电讯工程学院的研究生时，信息论方面的知识主要来源于王育民教授等编著的《信息论与编码》，进入上海大学任教以后，对研究生及高年级本科生讲授了四年共十遍信息论与编码课程，其间主要参考了吴伯修教授编著的《信息论与编码》，在纠错编码方面则更多地参考了王新梅教授等编著的《纠错码——原理与方法》。

在香港科技大学进行合作研究期间，则主要参阅了Robert B.Ash教授编著的"*Information Theory*"以及Richard E.Blahut教授编著的"*Principles and Practice of Information Theory*"等许多英文原版书，同时在与该校区子廉博士在信息论方面的许多颇有意义的讨论中也受益匪浅。因此，在以上过程中，逐步形成了本书目前的格局。本书并不试图全面介绍信息论的内容，而是择其主要的基础内容，并做合理安排后，呈献给读者，因此作者的目的是力求书中内容简单易懂，使读者可以较容易地掌握信息理论主要的和基础的知识。在编写该书的过程中，同时也融入了一些作者的教学体会。

作者要特别感谢上海交通大学颜国正教授和上海大学曹家麟教授对本书出版所给予的支持和鼓励，并感谢他们对本书所提出的宝贵意见。作者同时还要感谢他们的四位研究生在文字排版方面所做的大量工作。本书的出版得到了世界银行以及上海大学的资助，在此表示感谢。

由于时间仓促，水平有限，书中内容难免出错，望读者多加指正。

作　者

1999年12月于上海

目　录

第一章 引 论

本章主要介绍信息的定义及其基本特征，引出信源、信宿、信道、信源编码和译码与信道编码和译码的基本概念，给出信息与编码理论所要研究的主要问题。

§1.1 基 本 概 念

在我们现实生活当中，经常会遇到这样几个名词，即：信息、消息和信号，在我们正式开始学习信息与编码理论之前，搞清楚这样几个基本概念对学好这门课程是非常有益的。

一、信息

信息是任何随机事件发生后所包含的内容。这里要强调一点，就是信息仅仅与随机事件的发生相关，非随机事件的发生不包含任何信息。从这一点出发，我们可以预感到，信息量的大小与随机事件发生的概率有直接的关系，出现概率小的随机事件一旦发生，它所包含的信息量越大，而出现概率大的随机事件一旦发生，它所包含的信息量越小。

显然，信息量的大小与随机事件出现的概率呈反比，而且是对数反比关系，信息量的定义式将在第二章中给出。

二、消息

消息是信息的载体，是包含信息的语言、文字、图像等。上面已经谈到，信息只与随机事件的发生有关，这也是非常自然的，因为每时每刻在世界的每个地方，都会有各种事情发生，这些事情的发生一定是随机的，如果不是随机的而是确定的，那么通信也就失去了意义。在世界各地的人要想知道其他地方发生事情的内容，只能从各种各样的消息中得到，这些消息可以是广播中的语言、报纸上的文字、电视中的图像或互联网中的文字与图像等等。

这里要强调一点，不包含任何信息的语言、文字或图像不能构成消息。

三、信号

信号是消息的物理体现。就拿我们人类的语言来说，当语言从人们的嘴唇发出时，用仪器可以在嘴唇附近检测到声信号，这种声信号经过麦克风的转换，可以变成电信号。这里提到的声信号和电信号都是我们所指的信号，但在本课程中，我们主要是指电信号。

从以上的讨论中可以看到，信息、消息和信号之间有着密切的关系。信息是一切通信系统所要传递的内容；而消息作为信息的载体只能是一种“高级”载体；信号作为消息的物理体现，是信息的一种“低级”载体。作为系统设计人员，我们所接触的只是信号，而这种信号最终要变成消息的形式才能被大众接受。

四、信息的主要特征

信息作为通信系统所要传递的主要内容，它有以下几个特征：

(1) 信息在收到之前，它的内容是未知的；

(2) 信息可以产生、消失、被携带、储存及处理；

(3) 信息可以使认识主体逐渐了解和认识某一未知事物；

(4) 信息可以度量，信息量有多少之分。

从信息的以上几个主要特征中，可以进一步发现，信息包含在未知事件当中。信息可以从无到有，在以信息为主要特征的现代社会当中，如何有效地存储信息、处理与传递信息成为我们所要研究的主要问题。常常，获取信息的目的是为了逐渐了解某一未知事物，从信息的不断积累当中，可以逐渐发现一些未知事物的运动规律，从而达到完全了解与认识这一事物的目的，因此，获取信息是人类了解自然、认识自然的唯一途径。从人的感知角度上讲，信息显然有大小之分，一个很少发生的事情，当它突然发生的时候，给人的冲击是巨大的，这说明该事件的发生所包含的信息量巨大；反之，经常发生的事情给人的印象很淡。信息有量的大小之分，这一点是信息与编码理论研究的基础。

五、什么是信息论(信息与编码理论)

信息论是在信息可以量度的基础之上，应用统计数学的方法来研究如何有效地与可靠地传递信息的一门科学。其研究重点集中在方法的有效性与可靠性上。

§1.2 数字通信系统模型

数字通信系统所包括的范围很广，从现在的市话通信系统、数字蜂窝系统、计算机通信系统到雷达系统、遥控遥测系统等都是数字通信系统。简言之，数字通信系统是以数字信号的形式来传递信息的一种通信系统，可以归结为如图 1-1 所示的模型。

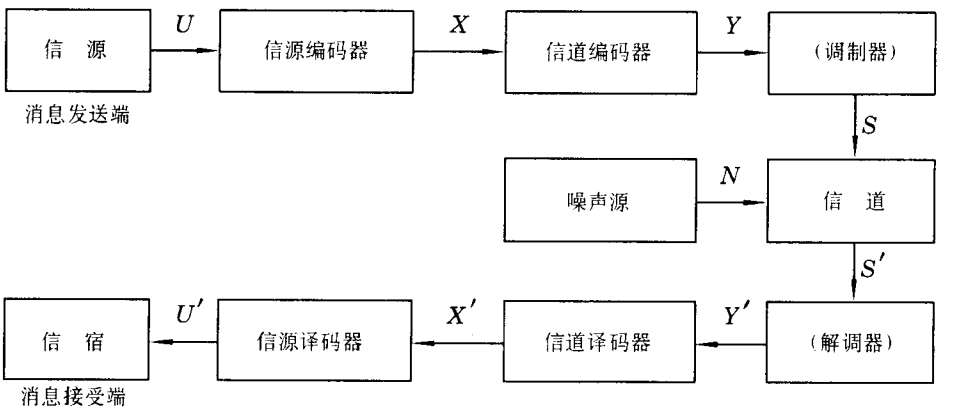


图 1-1 数字通信系统模型

在以上的数字通信系统模型中，各部分的定义及其功能可以归纳如下。

(1) 信源。信源是产生消息的源，它可以是人或物。它所发出的消息可以是语言、文字

或图像等，但其内容一定是不确定的(随机的)。信源可以是离散的(离散信源)，如发出文字消息的信源；也可以是连续的(连续信源)，如发出语言消息的信源。研究信源的重点集中在其所发出的消息的统计特性及其产生信息的速率上。

(2) 信源编码器。信源编码器的作用是将信源发出的消息变换成二进制或多进制数字序列，同时要求用来表示每条消息的数字序列的平均长度越小越好，这样，消息在传递过程中占用信道的时间相对减少，从而可以提高消息传递的有效性。因此，研究信源编码器的主要问题是如何提高编码的有效性。

(3) 信道编码器。消息经过信源编码器之后，已经被变换成了数字序列，并去除了原来消息中的许多冗余度以获得更高的有效性。这种码(信源码)在信道的传输过程中，一般不具备抗干扰能力，因此，为了保证传输的可靠性，往往人为地在信源码中加入一些纠错码，形成信道码，使消息在传递过程中具有一定的抗干扰能力，这就是信道编码器的最基本功能。因此，信道编码器研究的主要问题是如何提高消息传递的可靠性。

(4) 调制器。由信道编码器输出的数字序列(数字信号)一般被称为基带信号，这种基带信号一般都要经过较长距离的传输才能到达接收端。为了进行长距离传输，一般都要将这种基带信号调制到某一高频信号上去，以利于宽带信道的多路复用，如长途电话系统和数字蜂窝系统等；或是将它变换成音频模拟信号，以利于在窄带模拟信道上传输，如计算机通信系统等。

通常情况下，数字通信系统中的调制器都是必需的，但如果通信系统的两端，即消息发送端与消息接收端相隔不是很远，则有时可以采用基带信号的直接传输方式进行通信。这时，系统中的调制器就可以不要了。我们称有调制器的情况为数字信号的频带传输方式，而没有调制器的情况为数字信号的基带传输方式。

(5) 信道。信道是信息传输的通道，又称信息传输媒质。实际信道分为有线信道和无线信道，如双绞线、同轴电缆、光缆等均为有线信道，而微波、红外等可以用来作为无线信道。

信道研究的主要问题是它的统计特性和传信能力，即信道容量。

(6) 噪声源。噪声源包含了所有的外部噪声、通信设备内部噪声以及信道本身由于物理条件的变化而产生的信道随机噪声等。对噪声统计特性的研究有利于减弱噪声对通信系统的影响，提高通信质量。

(7) 解调器、信道译码器和信源译码器。处于数字通信系统接收端的解调器、信道译码器和信源译码器与发送端的调制器、信道编码器和信源编码器的作用正好相反。解调器的作用是将信道中传过来的高频已调信号转换成数字基带信号；信道译码器的作用是将数字基带信号(信道码)变换成信源码，在此变换过程中，要去掉用来纠错的多余码字，并纠正因传输所造成的错误；信源译码器的作用是将信源码变换成信宿可以接收的消息，同时要使该消息尽可能地不失真。

因此，解调器、信道译码器、信源译码器的最终作用就是要从受干扰的信号中，尽可能精确地恢复信源输出的消息，最大限度地提取信源输出消息中的信息并传给信宿。

(8) 信宿。信宿是消息的接收者，它可以是人或物。信宿与信源是完全相对的两个概念，它们虽处于两个不同的地点，一个收信而另一个发信，这只是单向通信的情况。实际系统中，大多数是双向通信，因此信宿与信源的身份是在不断地变化的。

图 1-1 也可以用图 1-2 的简化模型来表示。图 1-2 中的编码器包括信源编码器和信道

编码器；译码器包括信道译码器和信源译码器；编码信道包括调制器、解调器和实际信道，它又被称为广义信道。编码信道的输入输出均为二进制或多进制数字序列。

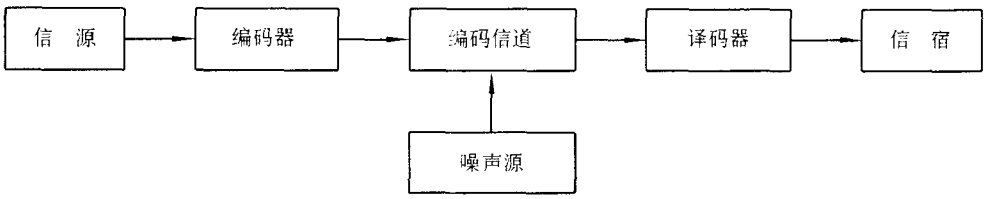


图 1-2　数字通信系统简化模型

§1.3　信息论研究的主要问题

从§1.1 节中可以看到，本书对信息的定义是比较狭窄的，而实际上信息所涉及的范围非常广泛，如语义信息、生物遗传信息、经济信息、管理信息等等都是信息的范畴。这些信息的研究涉及到语言学、生物遗传学、社会经济学、管理科学等更广泛的学科领域乃至边缘学科领域，这些内容已超出了本书信息论所要讨论的范围，属于“广义信息论”研究的范畴。与此相对应，本书内容被称作“狭义信息论”或“经典信息论”，也可称作“香农信息论”。

作为一门经典理论，香农信息论所要研究的主要问题是通信系统的有效性与可靠性问题。针对系统有效性问题，它研究在给定信源及信源编码有一定失真的条件下，信源编码的最低速率是多少，或者说，在给定信源编码速率的条件下，信源编码的最小失真是多少。这是信息论在信源编码方面所要研究的理论问题，与之对应的实际问题是寻找切实可行的和有效的信源编码及译码方法。信息论为我们寻找这种方法提供了理论依据和有价值的改进方向。

香农信息论所要研究的另一个主要问题是信道编码问题，即在保证信息传输可靠性或传输错误概率小于某一给定值的条件下，如何最有效地利用信道的传信能力。与之对应的实际问题是寻找切实可行的信道编译码方法。

可以说，香农信息论是从通信系统的最佳化角度来研究信息的传递和处理问题的。它的最大特点是将概率统计的观点和方法引入到通信理论研究中，揭示了通信系统中传输的对象是信息，并对信息给出了科学和定量的描述，指出通信系统设计的中心问题是在噪声干扰下系统如何有效而可靠地传递信息，实现这一目标的途径是编码，并且从理论上证明了编码方法可以达到的最佳性能极限。

学习香农信息论对于正确理解信息和信息理论，以及进一步发展信息论是非常必要和有益的。

第二章　信息量与熵

对信息的研究是在信息可以度量的基础之上进行的，因此，本章将首先给出信息量的定义，介绍离散消息的自信息量和互信息量、以及离散消息集合的平均自信息量和平均互信息量，同时引出不确定度与平均不确定度（熵）的概念。

§2.1　离散消息的自信息量

一、自信息量

离散消息的自信息量又称为非平均自信息量，它是指离散信源符号集合 X 中某一个符号 x_i 作为一条消息发出时对外提供的信息量，具体定义式如下：

$$I(x_i)=-\log_a p(x_i)=\log_a\left[\frac{1}{p(x_i)}\right] \tag{2.1.1}$$

式中：$p(x_i)$ 为 x_i 出现的概率。

式 (2.1.1) 中的自信息量单位取决于对数底 a 的取值，具体单位见表 2.1。

表2.1　自信息量单位

对数底a的取值	$I(x_i)$ 的单位
2	bit（比特）
e	nat（奈特）
10	hartley（哈特莱）

通过单位换算可以得到：1 nat = 1.443 bit；1 hartley = 3.322 bit。

以 nat 和 bit 的换算关系举例如下。

例 2.1　离散消息 x_i 所含的自信息量可以分别表示为：$-\log_e p(x_i)$ nat 或 $-\log_2 p(x_i)$ bit，显然，这两个信息量是相等的，所以有

$$-\log_e p(x_i)\ \text{nat}=-\log_2 p(x_i)\ \text{bit}$$

或写成

$$1\ \text{nat}=\frac{\log_2 p(x_i)}{\log_e p(x_i)}\ \text{bit}=\frac{\log_2 p(x_i)}{\frac{\log_2 p(x_i)}{\log_2 e}}\ \text{bit}=\log_2 e\ \text{bit}=1.443\ \text{bit}$$

所以有

$$1\ \text{nat}=1.443\ \text{bit}$$

为了进一步理解自信息量的概念，再来看下面的例子。

例 2.2 设信源只含有两个符号“0”和“1”，且它们以消息形式向外发送时均以等概率出现，求它们各自的自信息量。

解 因为

$$p(0)=p(1)=0.5$$

所以由式(2.1.1)可得

$$I(0)=I(1)=-\log_2 0.5=\log_2 2=1 \text{ bit}$$

由上例可以看到，二进制码以等概率出现时，每个码元所含的信息量是 1 bit。

二、不确定度

不确定度是指一个随机事件未发生之前所具有的不确定程度。显然，一个随机事件发生的概率越小，它的不确定度越大，那么当它发生时对外提供的自信息量也越大。因此，不确定度与自信息量是等量的关系，自信息量可以用来解除不确定度，消息只有被接收以后才有信息量，否则只有不确定度，因此不确定度是任何随机事件本身所具有的特性。

例 2.3 某地某月份的气象资料如表 2.2 所列，求相应事件的不确定度。

表2.2 气象资料

x_i	x_1(晴)	x_2(阴)	x_3(雨)	x_4(雪)
$p(x_i)$	1/2	1/4	1/8	1/8

解 因为，不确定度与自信息量是等量的，所以有

$$d(x_1)=I(x_1)=-\log_2 p(x_1)=1 \text{ bit},\ d(x_2)=I(x_2)=-\log_2 p(x_2)=2 \text{ bit}$$

$$d(x_3)=I(x_3)=-\log_2 p(x_3)=3 \text{ bit},\ d(x_4)=I(x_4)=-\log_2 p(x_4)=3 \text{ bit}$$

三、条件自信息量

条件自信息量是指一个随机事件 x_i 在另外一个随机事件 y_j 已发生的情况下再发生时所提供的信息量。其定义式如下：

$$I(x_i|y_j)=-\log_a p(x_i|y_j)=\log_a\left[\frac{1}{p(x_i|y_j)}\right] \tag{2.1.2}$$

式中：$p(x_i|y_j)$ 是在已知 y_j 的条件下，x_i 出现的条件概率。

当 x_i 与 y_j 相互独立时，有 $p(x_i|y_j)=p(x_i)$，此时，x_i 的条件自信息量就等于自信息量，即 $I(x_i|y_j)=I(x_i)$。

一般情况下，有 $p(x_i|y_j)>p(x_i)$，由式 (2.1.2) 可知，$I(x_i|y_j)<I(x_i)$。这是因为当两个随机事件 x_i 与 y_j 不相互独立时，一个事件 y_j 发生后，另一个事件 x_i 的不确定度减小，它发生以后所提供的信息量自然也随之减小。

极端情况下，当 $x_i=y_j$ 时，因为 $p(x_i|y_j)=1$，所以 $I(x_i|y_j)=0$，即此时由于事件 y_j 的发生而使事件 x_i 的不确定度变为 0，它已不含任何信息量。

§2.2 离散消息的互信息量

一、互信息量

离散消息的互信息量又被称为非平均互信息量。它的定义式如下：

$$I(x_i;y_j)=\log_a\left[\frac{p(x_i|y_j)}{p(x_i)}\right] \tag{2.2.1}$$

将上式右边展开，可得

$$\begin{aligned}I(x_i;y_j)&=-\log_a p(x_i)-[-\log_a p(x_i|y_j)]\\&=I(x_i)-I(x_i|y_j)\end{aligned}$$

即 x_i 与 y_j 之间的互信息量等于 x_i 的自信息量与其条件自信息量之差。

如果把 x_i 看成是信源发出的符号消息，y_j 看成是经过信道传输后信宿收到的符号消息，那么，由于信宿已经收到符号 y_j，如果 y_j 与 x_i 相关的话，则 x_i 的条件自信息量 $I(x_i|y_j)$ 必然要比它的自信息量 $I(x_i)$ 有所下降，这一下降的信息量就是信道从信源传到信宿的信息量。

当 $I(x_i|y_j)=0$ 时，说明信宿收到 y_j 后，信源符号 x_i 已经没有任何信息量，这也说明信道已经将有关 x_i 的全部信息量从信源传到了信宿，所以此时有 $I(x_i;y_j)=I(x_i)$。

当 x_i 与 y_j 完全独立时，说明信宿收到的符号 y_j 不包含任何有关 x_i 的内容，因此说信道没有传递任何有关 x_i 的信息，此时由于 x_i 与 y_j 完全独立，所以 $I(x_i|y_j)=I(x_i)$，即 $I(x_i;y_j)=0$。

从以上的讨论中可以看到，互信息量实际上就是信道传递的信息量，因此，互信息量的大小也反映了一个信道的传信能力。

二、互信息量的基本性质

性质 1 如果 x_i 与 y_j 相互独立，则互信息量 $I(x_i;y_j)=0$。

证明 因为 x_i 与 y_j 相互独立，所以有

$$p(x_i|y_j)=\frac{p(x_iy_j)}{p(y_j)}=\frac{p(x_i)p(y_j)}{p(y_j)}=p(x_i)$$

由互信息的定义式 (2.2.1) 可得

$$I(x_i;y_j)=\log_a\left[\frac{p(x_i|y_j)}{p(x_i)}\right]=\log_a\left[\frac{p(x_i)}{p(x_i)}\right]=0$$

这一性质从理论上证明了，如果信宿收到的符号消息 y_j 不包含任何有关 x_i 的信息时，则信道传递的信息量为 0。

性质 2 互信息量不可能大于自信息量，即 $I(x_i;y_j)\leqslant I(x_i)$。

证明 因为

$$p(x_i|y_j)\leqslant 1$$

所以

$$\frac{p(x_i|y_j)}{p(x_i)}\leqslant\frac{1}{p(x_i)}$$

$$\log_a\left[\frac{p(x_i|y_j)}{p(x_i)}\right]\leqslant\log_a\left[\frac{1}{p(x_i)}\right]$$

由互信息和自信息的定义式，可得

$$I(x_i;y_j)\leqslant I(x_i)$$

这一性质说明，信道上传递的信息量不可能超过信源所提供的信息量。

例 2.4 设信源发出四条消息，它们分别是

$$x_1=\text{“晴天”},\quad x_2=\text{“阴天”},\quad x_3=\text{“雨天”},\quad x_4=\text{“雪天”}$$

各自的出现概率分别为

$$p(x_1)=1/2,\quad p(x_2)=1/4,\quad p(x_3)=1/8,\quad p(x_4)=1/8$$

如果信宿收到的消息是 $y_1=$“不是晴天”，试求 y_1 分别与 x_1、x_2、x_3 和 x_4 之间的互信息量。

解 由已知条件可知，“晴天”出现的概率为 1/2，即 $p(x_1)=1/2$，显然，“不是晴天”的出现概率也应该是 1/2，即 $p(y_1)=1-p(x_1)=1/2$。

y_1 分别与 x_1,x_2,x_3 和 x_4 的联合概率为

$$p(x_1y_1)=0$$
$$p(x_2y_1)=p(x_2)=1/4$$
$$p(x_3y_1)=p(x_3)=1/8$$
$$p(x_4y_1)=p(x_4)=1/8$$

对以上联合概率稍加解释：因为消息 x_1 与 y_1 为两个不相容事件，不可能同时发生，因此它们的联合概率为 0；而消息 x_2、x_3、x_4 则完全包含在消息 y_1 中，因此它们之间的联合概率分别就是它们各自的出现概率。

根据以上联合概率可以求得相应的后验概率

$$p(x_1|y_1)=\frac{p(x_1y_1)}{p(y_1)}=2\,p(x_1y_1)=0$$

$$p(x_2|y_1)=\frac{p(x_2y_1)}{p(y_1)}=2\,p(x_2y_1)=\frac{1}{2}$$

$$p(x_3|y_1)=\frac{p(x_3y_1)}{p(y_1)}=2\,p(x_3y_1)=\frac{1}{4}$$

$$p(x_4|y_1)=\frac{p(x_4y_1)}{p(y_1)}=2\,P(x_4y_1)=\frac{1}{4}$$

根据互信息的定义，可得

$$I(x_1;y_1)=0\text{ bit}\quad(\text{定义 }\log_a0=0)$$

$$I(x_2;y_1)=\log_2\left[\frac{p(x_2|y_1)}{p(x_2)}\right]=\log_2\left[\frac{(1/2)}{(1/4)}\right]=1\text{ bit}$$

$$I(x_3;y_1)=\log_2\left[\frac{p(x_3|y_1)}{p(x_3)}\right]=\log_2\left[\frac{(1/4)}{(1/8)}\right]=1\text{ bit}$$

$$I(x_4;y_1)=\log_2\left[\frac{p(x_4|y_1)}{p(x_4)}\right]=\log_2\left[\frac{(1/4)}{(1/8)}\right]=1\text{ bit}$$

对以上结果稍做解释：由于 x_1 与 y_1 为两个互不相容事件，它们之间不存在任何互信息，因

此互信息量为0，而 y_1 与 x_2,x_3,x_4 之间的互信息量为1 bit，说明信宿收到符号消息 y_1 后，信道分别传递了有关 x_2、x_3、x_4 的一个比特的信息量。换句话说，在信宿收到符号消息 y_1 后，信源符号消息 x_2、x_3、x_4 的不确定度下降了一个比特，它们所能提供的信息量分别从原来（信宿未收到 y_1 之前）的2 bit、3 bit、3 bit 下降到（信宿收到 y_1 以后）1 bit、2 bit、2 bit。因此可以说，信道传递的互信息量是用来解除信源符号的不确定度的，当信道传递的互信息量等于信源符号提供的信息量时，则信源符号的不确定度完全被解除，信宿完全接收到了信源提供的信息量。

§2.3 离散消息集合的平均不确定度——熵

一、熵

对于单个符号消息，已经知道了自信息量与不确定度两个基本概念。然而，在大多数情况下，更关心的是离散信源符号集合的平均信息量问题，即信源中平均每个符号对外所能提供的信息量问题，以及信源符号集合的平均不确定度问题。

我们将信源符号集合的平均不确定度定义为熵，即信源符号消息在未接收到之前，它的平均不确定程度，这是信源的固有属性。只有在信源符号消息被接收之后，才有平均信息量存在，因此，平均信息量可以用来解除信源的平均不确定度，它与信源熵是等量的，但意义有所不同。

为此，将离散信源的熵定义为

$$H(X)=\sum_X p(x)I(x)=-\sum_X p(x)\log_a p(x) \tag{2.3.1}$$

式中：X 为离散信源符号集合；$H(X)$ 的单位取决于对数底 a 的取值，通常情况下取 $a=2$，这时，$H(X)$ 的单位为 bit/ 符号。

若离散信源 X 中只有 M 个符号，则上式又可以表示成

$$H(X)=-\sum_{i=1}^{M} p(x_i)\log_a p(x_i) \tag{2.3.2}$$

为了进一步理解熵的概念，我们来看以下两例。

例 2.5 设离散信源含26个英文字母，且每个字母均以等概率出现，求信源熵。

解 已知信源概率分布为

$$p(x_i)=\frac{1}{26},\quad i=1\sim 26$$

由信源熵定义式 (2.3.2) 可得

$$H(X)=-\sum_{i=1}^{26} p(x_i)\log_2 p(x_i)=-\sum_{i=1}^{26}\left(\frac{1}{26}\right)\log_2\left(\frac{1}{26}\right)=\log_2 26=4.7\ \text{bit/ 符号}$$

即信源每个符号的平均不确定度为4.7 bit，换句话说，信源平均每个符号所能提供的信息

量为 4.7 bit。

例 2.6 设信源 X 只有两个符号 x_1、x_2，各符号的出现概率分别为 $p(x_1)=q$，$p(x_2)=1-q$，求信源熵 $H(X)$。

解 根据信源熵的定义式 (2.3.2) 可得

$$H(X)=-\sum_{i=1}^{2}p(x_i)\log_2 p(x_i)=-q\log_2 q-(1-q)\log_2(1-q)\quad \text{bit/ 符号} \tag{2.3.3}$$

根据上式，可以画出信源熵随参数 q 的变化曲线，如图 2-1 所示。

(1) 当 $q=1/2$ 时，有 $p(x_1)=p(x_2)=1/2$，将 q 代入式 (2.3.3) 可得

$$H(X)=-\frac{1}{2}\log_2\left(\frac{1}{2}\right)-\frac{1}{2}\log_2\left(\frac{1}{2}\right)$$
$$=\log_2 2=1\ \text{bit/ 符号}$$

显然，当 $q=1/2$ 时，信源符号为等概率分布，从图 2-1 中可以看到，此时信源熵达到最大值，这并不是偶然的。对任何离散信源，当其中的符号等概率分布时，信源熵均达到最大值，即此时信源对外提供的平均信息量最大。

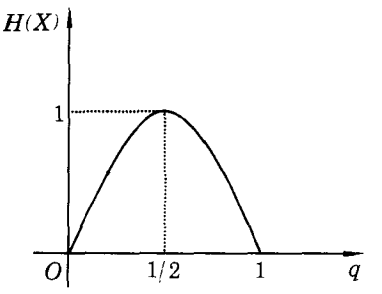


图 2-1 信源熵随参数 q 的变化曲线

(2) 当 $q=0$ 时，有 $p(x_1)=0$，$p(x_2)=1$，将 q 代入式 (2.3.3) 得

$$H(X)=-\lim_{q\to 0}(q\log_2 q)-\log_2 1=-\lim_{q\to 0}(q\log_2 q)=0\ \text{bit/ 符号}$$

同样，当 $q=1$ 时，有 $p(x_1)=1$，$p(x_2)=0$，可以证明此时也有 $H(X)=0$ bit/ 符号。

可以看到，无论是 $q=0$ 还是 $q=1$，信源中的一个符号总是出现，而另一个符号总是不出现，即信源中消息的出现是完全确定的，这已经不是随机变量了，所以这样的信源已经不可能提供任何信息量。

二、条件熵

信源符号集合 X 相对于信源符号集合 Y 的条件熵定义为

$$H(X|Y)\overset{\text{def}}{=}\sum_{XY}p(xy)I(x|y)=-\sum_{XY}p(xy)\log_a p(x|y) \tag{2.3.4}$$

即条件熵等于条件自信息量在 XY 上的数学期望值。

$H(X|Y)$ 表示信宿收到 Y 后，X 仍然具有的平均不确定度，通常称 $H(X|Y)$ 为信源疑义度。

如果 X 和 Y 相互独立，根据条件自信息的性质有 $I(x|y)=I(x)$，则

$$H(X|Y)=\sum_{XY}p(xy)I(x|y)=\sum_{XY}p(x)p(y)I(x)=\sum_{X}p(x)I(x)\sum_{Y}p(y)$$

$$=\sum_{X}p(x)I(x)=H(X)$$

即此时的条件熵就等于信源熵。X 与 Y 独立，说明信宿收到的符号集合 Y 中不含有任何有关符号集合 X 的信息，此时信源疑义度最大，即信宿收到 Y 后，信源此时的不确定度最大。

如果 $X=Y$，由于 $I(x|y)=0$，所以 $H(X|Y)=0$。说明信宿在收到 Y 后，信源被完全确定，它不再有任何不确定度。

另一种条件熵

$$H(Y|X)\overset{\text{def}}{=}\sum_{XY}p(xy)I(y|x)=-\sum_{XY}p(xy)\log_a p(y|x) \tag{2.3.6}$$

称为噪声熵。

显然，当 X 与 Y 独立时，$H(Y|X)=H(Y)$，此时 Y 不包含任何有关 X 的信息，Y 全部是信道中的噪声，因此噪声熵由此而得名。当 $Y=X$ 时，噪声熵为 0，即此时信道中没有噪声。通常情况下，噪声熵的值在 0 与 $H(Y)$ 之间，因此噪声熵的大小反映了信道上噪声的大小。

三、联合熵

联合熵是定义在符号集合 XY 上的自信息量 $I(xy)=-\log_a p(xy)$ 的数学期望值

$$H(XY)\overset{\text{def}}{=}\sum_{XY}p(xy)I(xy)=-\sum_{XY}p(xy)\log_a p(xy) \tag{2.3.7}$$

实际计算时常用下式，即

$$H(XY)=H(X)+H(Y|X) \tag{2.3.8}$$

上式的合理性证明如下。

证明 因为

$$\begin{aligned}I(xy)&=-\log_a p(xy)=-\log_a[p(x)p(y|x)]\\&=-\log_a p(x)-\log_a p(y|x)=I(x)+I(y|x)\end{aligned}$$

即有

$$I(xy)=I(x)+I(y|x) \tag{2.3.9}$$

所以由式 (2.3.7) 和 (2.3.9)，有

$$\begin{aligned}H(XY)&=\sum_{XY}p(xy)I(xy)=\sum_{XY}p(xy)[I(x)+I(y|x)]\\&=\sum_{XY}p(xy)I(x)+\sum_{XY}p(xy)I(y|x)=\sum_{XY}p(x)p(y|x)I(x)+\sum_{XY}p(xy)I(y|x)\\&=\sum_{X}p(x)I(x)\sum_{Y}p(y|x)+\sum_{XY}p(xy)I(y|x)=\sum_{X}p(x)I(x)+\sum_{XY}p(xy)I(y|x)\\&=H(X)+H(Y|X)\end{aligned}$$

同样可以证明

$$H(XY)=H(Y)+H(X|Y) \tag{2.3.10}$$

在式 (2.3.8) 中，当 X 与 Y 独立时，有

$$H(XY)=H(X)+H(Y) \tag{2.3.11}$$

这是因为 X 与 Y 独立时，有 $I(xy)=I(x)+I(y|x)=I(x)+I(y)$，同时有 $p(xy)=p(x)p(y)$，将此两式代入联合熵的定义式 (2.3.7) 中，有