普通高等教育“九五”国家级重点教材

21 世纪信息通信系列教材

通信原理

TONGXIN YUANLI

周炯槃　庞沁华
续大我　吴伟陵　编著

（下）

北京邮电大学出版社
www.buptpress.com

通 信 原 理

（下）

周炯槃　庞沁华　续大我　吴伟陵　编著

北 京 邮 电 大 学 出 版 社
北京·Beijing

内容简介

本书系统、深入地介绍了通信系统及通信网的基本原理及基本分析方法，是通信及信息专业的专业基础课教材。

全书共十一章，内容包括通信系统及通信网的基本概念、确定信号及随机过程、模拟通信系统、数字基带传输、数字频带传输、信源和信源编码、信道和信道容量、信道差错控制编码、正交编码与伪随机序列及其应用、通信网的基本原理。

本书概念清楚，取材新颖，书中除列举了大量例题，还附有习题及部分习题答案。

本书可作为高等学校通信工程、信息工程、电子工程和其他相近专业本科生的教材，也可供通信工程技术人员和科研人员作为参考书。

图书在版编目(CIP)数据

通信原理(下)/周炯槃，庞沁华等编著.—北京：北京邮电大学出版社，2002

ISBN 7-5635-0525-3

Ⅰ.通… Ⅱ.①周…②庞… Ⅲ.通信理论 Ⅳ.TN911

中国版本图书馆 CIP 数据核字(2002)第 047823 号

出 版 者：北京邮电大学出版社(北京市海淀区西土城路 10 号)

邮编：100876 (发行部)电话：62282185 传真：62283578

电子信箱：publish@bupt.edu.cn

经 销：各地新华书店

印 刷：北京源海印刷有限责任公司

印 数：8 001 — 11 000 册

开 本：787mm × 1 092mm 1/16

印 张：13.75

字 数：321 千字

版 次：2002 年 11 月第 1 版 2004 年 2 月第 3 次印刷

ISBN 7-5635-0525-3/TN·236 定 价：22.00 元

· 如有质量问题请与北京邮电大学出版社发行部联系 ·

序

通信乃是互通信息。

20世纪80年代以来,通信技术与计算机技术及微电子技术相互促进、迅猛发展。通信产业已成为当今世界生产力的重要领头产业。它极大地推动了各国经济的发展,使人类步入了信息社会。

通信技术日新月异的发展,使更多的人们期望掌握通信的基本原理。本书的宗旨是系统、深入地阐述通信系统和通信网的基本理论。

本书系四位作者多年来从事通信与信息学科本科生和研究生的教学与科研实践的总结。按认识论规律由概念的建立给出定量分析,且注重理论联系实际。在介绍基本理论的基础上,力求体现近年来国内外通信技术的发展。

为帮助读者掌握通信的基本原理及分析方法,提高运算能力,书中列举了许多例题,并附有大量习题及部分习题答案。习题中包含历年来学生作业、本科试题及考研试题的精选。

全书共十一章,第一章绪论,从通信发展简史及展望引导出通信的基本概念与通信系统及通信网的基本构成。第二章确定信号分析,第三章随机过程,这两章是分析通信系统的数学工具,如果读者已有先修基础,可将其作为复习内容。第四章模拟通信系统,阐述目前正在应用的各种模拟调制方式的基本原理及其性能分析。第五章至第十章阐述数字通信系统的基本理论。第五章数字信号的基带传输,第六章数字信号的频带传输,这两章是数字通信传输系统的基本理论。第七章信源和信源编码,第八章信道和信道容量,第九章信道差错控制编码,这三章是仙农信息论的基本理论。第十章阐述正交编码与伪随机序列的基本原理及其应用。第十一章介绍通信网的基本原理,阐述交换的基本原理及信令和协议的基本概念。全书分为上、下两册,一至六章为上册,七至十一章为下册。

本书可用作高等学校本科生通信工程专业、电子与信息专业以及其他相近专业的专业基础课的教科书。全书可在大学三年级分两学期授课。

本书也可作为通信工程技术人员的参考书。读者欲仅限于掌握模拟通信,则只需阅读前四章;若仅限于掌握数字通信,不必阅读第四章。

全书由周炯槃院士主编并编写第一、十一章,由续大我教授编写第二、三、八、十章,由庞沁华教授编写四、五、六章,由吴伟陵教授编写第七、九章,最后全书由周炯槃、庞沁华统编定稿。

鉴于首次正式出版,难免有不妥之处,敬请指正。

周炯槃

2002年8月于北京邮电大学

目　录

第七章　信源和信源编码

第八章　信　　道

第九章　信道编码

第十章　正交码与伪随机码

第十一章　通信网的基本知识

第七章　信源和信源编码

7.1　引　言

在现代通信中，信源和信道是组成通信系统的最基本单元。信源是产生信息的源，信道则是传送载荷信息的信号所通过的通道，信源与信宿之间的通信是通过信道来实现的。

度量通信的技术性能主要是从通信的数量与质量两方面来讨论的，一般数量指标用有效性度量，而质量指标用可靠性度量。前者主要与信源统计特性有关，而后者则主要决定于信道的统计特性。

以前，通信研究的重点是信道，主要研究的问题是通信的质量，即可靠性问题，这是非常必要的，但是它是不全面的。譬如：数字调制技术同时也要考虑有效性问题。从通信系统的优化观点来看，通信研究的另一个重点应是信源，它主要研究的问题是通信的数量，即有效性问题。只有同时研究通信的数量与质量、有效和可靠，同时研究信源和信道，才能使整个通信系统实现优化，达到既有效又可靠。可见，通信系统是信源与信道相配合的统一体，通信系统的优化应是寻求信源与信道之间最佳的统计匹配。

从信息论观点看，实际的信源若不经过信息处理，即信源编码，信源会存在大量的统计多余的成分，这一部分信息完全没有必要通过信道传送给接收端，因为它完全可以利用信源的统计特性在接收端恢复出来。信源编码的任务是在分析信源统计特性的基础上，设法通过信源的压缩编码去掉这些统计多余成分。这也就是为什么要在通信原理中特别加上信源编码这一章的主要原因。

在本章中，将扼要介绍信源统计特性描述、信源的信息度量，即信息熵以及信源编码概要，最后将介绍模拟信源的数字化 PCM 编码。

在信源统计特性描述方面，主要讨论信源的分类、信源统计模型与描述，重点是讨论最简单、最基本的离散单消息与离散无记忆信源。在信源的信息度量方面重点讨论熵、互信息以及信息率失真 $R(D)$函数等基本概念。在信源编码方面，主要讨论无失真与限失真编码定理概要；离散无失真单消息信源与无记忆信源的熵编码；离散限失真有记忆信源解除相关性的预测编码与域变换编码；最后讨论连续模拟信源 PCM 编码，它包括取样定理、最佳量化、压扩特性以及 A 律与 μ 律特性等。

7.2 信源统计特性描述

信源是产生信息的源头,从物理背景上看实际信源是多种多样的,最常见的有文字、语音、图像以及各类数据信源。这里为了分析与描述方便,可将各类实际信源抽象概括为两大类型:离散(或数字)信源和连续(或模拟)信源,其中文字、电报以及各类数据属于离散信源,而未经数字化的语音、图像则属于连续信源。

这里为了简化分析和突出问题实质,不妨假设信源中仅含有一个消息(符号),而这个消息是一个不确定量,比如它可以是二进制数中的"0"或"1",也可以是英文 26 个字母中的某一个字母,还可以是中文数千个单字中的某一个单字,称它为单消息符号信源。

下面,首先讨论最简单、最基本的由单个消息(符号)所组成的信源。它一般可以采用单消息的取值范围 X 以及消息取值 x_i 的概率 $P(x_i)$ 来共同描述,即

$$[X, P(x_i)]$$

也可写成

$$\begin{pmatrix} X \\ P(x_i) \end{pmatrix} = \begin{bmatrix} X = x_1 \cdots X = x_i \cdots X = x_n \\ P(x_1) \cdots P(x_i) \cdots P(x_n) \end{bmatrix} \tag{7.2.1}$$

例如,对于离散、单消息的二进制等概率信源

$$\begin{pmatrix} X \\ P(x_i) \end{pmatrix} = \begin{pmatrix} X=0, X=1 \\ P(0), P(1) \end{pmatrix} = \begin{bmatrix} 0, & 1 \\ \frac{1}{2}, & \frac{1}{2} \end{bmatrix} \tag{7.2.2}$$

同理,可给出单个连续变量信源

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} x \in (a, b) \\ p(x) \end{bmatrix} \tag{7.2.3}$$

其中 $p(x)$ 表示具体取连续值 x 的概率密度。

实际信源是由上述最基本的单个消息信源组合而成。离散时,它是一个消息序列,在数学描述上可写成一随机序列:$\boldsymbol{X} = (X_1 \cdots X_l \cdots X_L)$,这就是说表达离散、实际信源的随机序列 $\boldsymbol{X}$ 具有两个重要特征:在横轴方向上它是由 $l = 1, 2, \cdots, L$ 个单个消息(符号)x_l 构成,在纵轴方向上每个消息(符号)x_l 都是一个随机变量,它有 $i = 1, 2, \cdots, n$ 种取值可能;连续时,它是一个模拟消息,在数学描述上可写成一随机过程 $X(\omega, t)$,简记为 $X(t)$,其中 $\omega \in (-\infty, +\infty)$、$t \in (-T, T]$,对每个瞬间 $t = t_i$、$X(\omega, t_i)$ 是一个取值为 $\omega \in k' = (-\infty, \infty)$ 的连续随机变量。实际上,文字信源、数据信源以及数字化以后的语音与图像信源均可表达成离散消息序列信源。模拟语音与模拟图像等模拟信源均可表达成连续随机过程 $X(t)$ 信源。

前面已指出,只要满足限时、限数这类物理上可实现的基本条件,模拟信源可以离散化为离散消息序列信源来表达。因此对于实际信源的统计描述,这里仅讨论消息序列信源。

类似于对上述单个消息(符号)的描述方法,对于离散消息序列信源也可以采用类似方法。

假设消息序列信源由 L 个消息(符号)构成,且消息序列中每个消息(符号)取值集合(范围)是相同的,用 X 表示,则消息序列信源的取值集合可以表示为

$$X^L = X \times X \times \cdots \times X(\text{共计 } L \text{ 个}) \tag{7.2.4}$$

这时,信源输出是一个 L 维随机矢量 $\boldsymbol{X}$

$$\boldsymbol{X} = (X_1 \cdots X_l \cdots X_L) \tag{7.2.5}$$

随机矢量 $\boldsymbol{X}$ 的具体取值(样值)为 $\boldsymbol{x}$

$$\boldsymbol{x} = (x_1 \cdots x_l \cdots x_L) \tag{7.2.6}$$

样值 $\boldsymbol{x}$ 对应概率 $P_{\boldsymbol{X}}(\boldsymbol{x})$,简写成 $P(\boldsymbol{x})$ 为

$$P(\boldsymbol{x}) = P(x_1 \cdots x_l \cdots x_L) = P(x_1)P(x_2/x_1)P(x_3/x_2x_1)\cdots P(x_L/x_{L-1}\cdots x_1) \tag{7.2.7}$$

它为一个 L 维的联合概率。

根据信源是否允许失真,又可以将信源划分为无失真信源与限失真信源两类。对于无失真消息序列信源的描述可以采用信源的消息序列的取值集合 X^L 及其对应的概率 $P(\boldsymbol{x})$来共同描述:$[X^L, P(\boldsymbol{x})]$。它也可写成

$$\begin{pmatrix} X^L \\ P(\boldsymbol{x}) \end{pmatrix} = \begin{bmatrix} \boldsymbol{X} = x_1 \boldsymbol{X} = x_2 \cdots & \boldsymbol{X} = x_{n^L} \\ P(x_1), P(x_2) \cdots & P(x_{n^L}) \end{bmatrix} \tag{7.2.8}$$

其中,消息序列长度为 $l = 1, 2, \cdots, l, \cdots, L$,而每个消息又有 n 种可能的取值,即 $i = 1, 2, \cdots, n$,因此整个消息序列总共有 n^L 种取值。

对于离散序列信源还可以进一步划分为无记忆与有记忆两类,当序列中的前后消息相互统计独立时称为无记忆,否则称为有记忆。本节重点讨论简单的无记忆离散序列信源,这时

$$P(x_1 \cdots x_l \cdots x_L) = \prod_{l=1}^{L} P(x_l) = P^L \tag{7.2.9}$$

其中公式(7.2.9)中后面的等式仅在平稳无记忆信源条件下成立,这时序列中各消息统计特性与序列所处时间(位置)无关,故称它为平稳无记忆。实际通信中脉码调制(PCM)属于这类信源。下面为了简化分析,假设仅取 3 位码,即 $L = 3$,且设 $P(0) = P(1) = \frac{1}{2}$,则有

$$\begin{pmatrix} X^3 \\ P(\boldsymbol{x}) \end{pmatrix} = \begin{bmatrix} 000, & 001 & \cdots & 111 \\ P^3(0), & P^2(0)P(1), & \cdots, & P^3(1) \end{bmatrix} \begin{bmatrix} 000, 001 & \cdots & 111 \\ \frac{1}{8}, \frac{1}{8}, & \cdots, & \frac{1}{8} \end{bmatrix} \tag{7.2.10}$$

有记忆信源是指消息序列中前、后消息间不满足统计独立的条件,实际信源一般均属此类。但是描述这类信源比较困难,特别是序列消息间记忆长度 L 很大时,需要掌握全部记忆区域内 L 维的概率统计特性。在实际处理时,往往可以作进一步简化,特别是当消息序列中任一个消息仅与前面一个消息有直接统计关联,或者推而广之,消息序列中 K 个消息组成的一个消息状态仅与前面 K 个消息组成的前一消息状态有直接统计关联时,称该信源为一阶马氏链信源。进一步,若这类一阶马氏链信源又满足齐次与遍历的条件,这里齐次是指消息的条件转移概率随时间推移不变,即与所在时间位置无关;这里的遍历则是指当转移步数足够大时,序列的联合概率特性基本上与起始状态概率无关。在

这种特殊情况下描述与分析可以进一步大大减化。比如数字图像信源往往是可以采用这一模型作为分析的近似模型。

对于限定失真度为 D 的离散单个消息信源,其统计描述可以在公式(7.2.1)基础上加以推广

$$\{[X,P(x)],[X\times Y,d(x,y)]\} \tag{7.2.11}$$

其中,$d(x,y)$: $X\times Y\to[0,\infty)$,表示在单个消息信源 X 和单个消息信源 Y 的联合取值集合(范围)上定义的一个实值失真函数值。

7.3 信源的信息度量:信息熵 $H(X)$

上一节,讨论了信源的分类与统计描述,主要利用了信源的客观概率分布(含概率与概率密度)特性描述了信源。为了进一步深入定量地研究信源,仅限于上述一般化的描述是不够的。信源就其信息实质来说,具有不确定性,那么信息与不确定性是什么关系,而不确定性又如何利用客观概率来表达,这就是本节所要讨论的问题实质:它就是信源输出的消息(或符号)所包含信息的定量度量问题。

在正式讨论信息定量度量以前,先简要地介绍一下信息的基本概念。信息是一个既广泛而又抽象的哲学概念,至今无确切定义。其广泛性主要体现在宇宙间、自然界中充满了客观特性信息,人类生存每时每刻也离不开信息。其抽象性主要体现在它是一种内涵丰富的重要概念,是一个能为人人理解但又无法确切定义的抽象概念。物质、能量与信息是支撑现代社会及其发展的三大支柱,信息既不是物质,也不是能量,但又离不开物质与能量,它是人类认识客观世界的更高层次,掌握信息不仅能充分利用物质和能量,还会创造新的物质与能量形式。在这里不打算对广义信息的概念作进一步深入探讨,而将重点放在分析、理解通信领域中狭义信息的概念上。在通信中,信息是指信源的内涵。信源所表达的内容与含义,是信道待传送的内容与含义,它是一个抽象的哲学表达层次上的概念。在通信中至少可以从两个不同的层次(侧面)来进一步描述与刻画它。

通信中描述信息的第一个层次是在工程领域中经常采用的最为具体的物理表达层,该层次的代表是信号。信号是一个物理量,可描述、可测量、可显示。通信中待传送的信息就是以信号参量形式载荷在信号上,这些参量是信号的振幅、频率、相位乃至参量的有与无。所以就物理表达层来看,信息是信号所载荷的内容与含义。

通信中描述信息的第二个层次是在理论领域中常采用较为抽象的数学表达层,该层次的代表是消息(或符号),它将抽象待传送的信息从数学实质上加以分类:一类为离散型,即数字信源,用相应的随机变量 X、随机变量序列 $\boldsymbol{X}=(X_1X_2\cdots X_l\cdots X_L)$来描述;另一类为连续型,即模拟信源,可以用相应的随机过程 $X(t)$来描述。在这个层次上抽象的信息概念可以在数学层次上被描述为随机序列和随机过程,从而为信息定量度量打下坚实的基础。在这一层次中,信息是消息所描述和度量的对象。

信号、消息、信息三个表达层次是一个统一体,它们之间的关系可以看作是哲学上的内涵与外延的关系。这就是说,信息是信号与消息的内涵,即消息所要描述和度量的内容、信号所要载荷的内容;而信号则是信息在物理层上的外延,消息则是信息在数学层次

上的处延。这也就是说信号与消息是信息在物理与数学两个不同方面的表达形式。同一内涵的信息可以采用不同消息形式描述,也可以采用不同的信号形式来载荷;相反,不同内涵的信息也可以采用同一消息形式来描述,同一类型信号形式来载荷。可见,信息、消息与信号三个层次之间是一个既统一又辩证的关系。

信源输出的是消息,消息的内涵是信息。信息的最主要特征是具有不确定性。如何定量度量信息的不确定性?上一节已讨论过信源的统计特性可以采用概率及概率密度来描述,那么度量信息的不确定性与信源的概率与概率密度应是什么关系?这正是本节所要讨论的主题。

信息的定量化一直是人们长期追求的目标。早在 1928 年,信息论的先驱学者之一哈特莱(Hartley)首先研究并给出了一种具有 N^m 组合的非概率(实际上可看成等概率)信源的信息度量公式,即

$$I = \log N^m = m \log N \tag{7.3.1}$$

由于包含信息的消息或符号最主要的特征是不确定性,而不确定性主要来源于客观信源的概率统计上的不确定性,而 Hartley 信息度量公式可以看成在等概率信源条件下信息度量的一个特例。这一观点完全被后来信息论创始人 C. E. Shannon 所吸收。

本节,将首先从人们容易接受的直观概念出发,推导出信源的信息度量公式:信息熵的基本公式,它与 C. E. Shannon 从严格的数学上给出的结论是完全一致的,当然也可以引用熵的公理化结构来证明这一点,但由于篇幅所限这里从略。

从直观概念推导信息熵的公式,可以分为两步走:第一步首先求出当某一个具体的单个消息(符号)产生(出现)时,比如 $x = x_i$ 时的信息度量公式记为:$I[P(x_i)]$;第二步求单个消息(符号)信源的信息熵(信息度量),由于单消息(符号)信源有 $i = 1, 2, \cdots, n$ 种取值可能,因此要取统计平均即 $H(X) = E[I(P_i)]$。

通常,对单个消息信源,比如 $X = x_i$,它出现的概率 $P(x_i)$越小,它一出现必然使人越感意外,则由它所产生的信息量就越大,即 $P(x_i)\downarrow$,$I[P(x_i)]\uparrow$,且当 $P(x_i)\to 0$ 时,$I[P(x_i)]\to\infty$;反之,当 $P(x_i)\uparrow$,$I[P(x_i)]\downarrow$,且当 $P(x_i)\to 1$ 时,$I[P(x_i)]\to 0$。

可见,对于单个消息信源,某个消息 $X = x_i$ 所产生的信息 $I[P(x_i)]$应是其对应概率 $P(x_i)$的递降函数。另外,由两个不同的消息(两者间统计独立)所提供的信息应等于它们分别提供信息量之和,即信息应满足可加性(实际上若两者不满足统计独立,也应满足条件可加性)。显然,同时满足对概率递降性与可加性的函数应是下列对数函数,即

$$I[P(x_i)] = \log \frac{1}{P(x_i)} = -\log P(x_i) \tag{7.3.2}$$

通常称 $I[P(x_i)]$为信源单个离散消息 $X = x_i$ 时的非平均自信息量。

同理,可以定义信宿 $Y = p(y_j)$以及两个消息有统计关联时,条件非平均自信息量与两个消息的联合非平均自信息量分别如下

$$I[P(y_j)] = \log \frac{1}{P(y_j)} = -\log P(y_j) \tag{7.3.3}$$

注:log 根据需要可取以 2 为底或以 e,10 为底。

$$I[P(y_j/x_i)]=\log\frac{1}{P(y_j/x_i)}=-\log P(y_j/x_i) \tag{7.3.4}$$

$$I[P(x_i/y_j)]=\log\frac{1}{P(x_i/y_j)}=-\log P(x_i/y_j) \tag{7.3.5}$$

$$I[P(x_iy_j)]=\log\frac{1}{P(x_iy_j)}=-\log P(x_iy_j) \tag{7.3.6}$$

上面,我们从直观概念直接推导出当信源某一个单消息、条件单消息以及两个消息联合同时出现时的非平均自信息量的表达式。然而,一般离散信源,即使是单消息信源,也具有有限种取值的可能:即 $i=1,2,\cdots,n$, $j=1,2,\cdots,m$,因此,这时信源输出的信息量就应该是上述具体单个消息产生的非平均自信息量的概率统计平均值,显然它与信源本身的概率特性有关。因此,可以定义信源输出的信息量,即信息论创始人 C. E. Shannon 将其定义为信源信息熵的概念如下:

$$\begin{aligned}H(X)&=H[P(x_1)\cdots P(x_n)]\\&=E\{I[P(x_i)]\}\\&=E[-\log P(x_i)]\\&=-\sum_{i=1}^{n}P(x_i)\log P(x_i)\end{aligned} \tag{7.3.7}$$

其中“E”表示求概率统计平均值,即求数学期望值。C. E. Shannon 称 $H(X)$为信源的信息熵,简称为熵。可见,从数学上看,熵是信源消息概率 $P(x_i)$的对数函数 $\log P(x_i)$的统计平均值,故又称为是 $P(x_i)$的泛函数。它是定量描述信源的一个重要物理量。它是由 Shannon 于 1948 年首先给出的一个从概率统计角度来描述信源不确定性的一个客观物理量,是从信源整体角度上反映信源的不确定性度量。熵这个名词是 Shannon 从统计热力学借用过来的。不过在统计热力学中称为热熵,它是用来表达统计热力学中分子运动混乱程度的一个物理量。Shannon 将它引入通信中,用它描述信源平均不确定性,其含义是类似的。但是在热力学中已知任何孤立系统的演化热熵只会增加不会减少,然而在通信中信息熵只会减少不会增加,所以也有人称信息熵为负热熵。

信息熵的单位与非平均自信息量的单位一样都取决于所取对数的底。在通信中最常用的是以 2 为底,这时单位称为比特(bit);有时在理论分析和推导时采用以 e 为底比较方便,这时单位称为奈特(Nat);在工程运算时,有时采用以 10 为底较方便,这时单位称为笛特(Det)。它们之间可以引用对数换底公式进行互换。比如:

$$1\ \text{bit}=0.693\ \text{Nat}=0.301\ \text{Det}$$

上面,我们从直观概念直接推导出 Shannon 的信息熵公式(7.3.7)。范因斯坦(Feinstein)等人曾证明:当信息满足对概率 $P(x_i)$的递降性,以及可加性的条件下,公式(7.3.7)的信息熵表达式是唯一的,后来人们称这一证明为熵的公理化结构证明。

这里,有必要讨论一下信息熵与信息量之间的关系。前面已指出信息熵是表征信源本身统计特性的一个物理量,它是信源平均不确定性程度的一度量值,是从信源整体统计特性上刻画信源的一个客观物理量,是一个绝对量。而人们一般所说的信息量是针对接收者而言的,是相对的,它既可以看作是接收者从信源所获得不确定性的减少量,也可看

作是信源(发送者)给予信宿(接收者)的不确定性减少量。若发与收(源与宿)之间传送的信道无干扰,接收者(信宿)所获得的信息量在数量上就等于发送者(信源)所给出的信息量,即信源信息熵,但是两者在概念上是有区别的。若信道中存在噪声,则不仅两者概念上有区别而且数量上也不相等。可见,信息熵 $H(X)$ 也可以理解为信源输出的信息量,然而,通常所指的信息量都是指接收者从信源所获得的信息量,这也就是我们将在后面要进一步介绍的互信息,它是一个相对量。

类似于对信息熵 $H(X)$ 的定义,同理也可以进一步对信宿熵 $H(Y)$、条件熵 $H(Y/X)$、$H(X/Y)$ 联合熵 $H(X,Y)$ 作如下类似定义:

$$H(Y)=E\{I[P(y_j)]\}=E[-\log P(y_j)]=-\sum_{j=1}^{m}P(y_j)\log P(y_j) \tag{7.3.8}$$

$$H(Y/X)=E\{I[P(y_j/x_i)]\}=E[-P(y_j/x_i)]=-\sum_{i=1}^{n}\sum_{j=1}^{m}P(x_iy_j)\log P(y_j/x_i) \tag{7.3.9}$$

$$H(X/Y)=E\{I[P(x_i/y_j)]\}=E[-P(x_i/y_j)]=-\sum_{i=1}^{n}\sum_{j=1}^{m}P(x_iy_j)\log P(x_i/y_j) \tag{7.3.10}$$

$$H(X,Y)=E\{I[P(x_iy_j)]\}=E[-P(x_iy_j)]=-\sum_{i=1}^{n}\sum_{j=1}^{m}P(x_iy_j)\log P(x_iy_j) \tag{7.3.11}$$

它们之间,有如下主要性质:

(1)
$$H(X,Y)=H(X)+H(Y/X) \tag{7.3.12}$$
$$=H(Y)+H(X/Y) \tag{7.3.13}$$

(2)
$$H(X)\geqslant H(X/Y) \tag{7.3.14}$$
$$H(Y)\geqslant H(Y/X) \tag{7.3.15}$$

公式(7.3.14)、(7.3.15)又称为 Shannon 不等式。

下面,从定性角度简要地讨论一下熵的基本性质。这里以最简单的离散二进制信源为例,其信源熵函数可以用图 7.3.1 所示图形直观地表示:

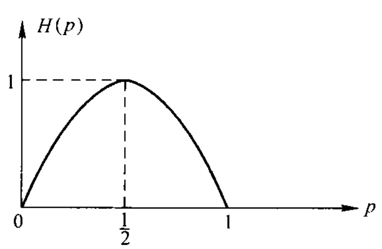

图 7.3.1　离散二进制信源熵函数

由图 7.3.1 可见,其熵函数除了在定义中引入的两个最基本条件(性质):对信源概率 $P(x_i)$ 满足递降性以及满足可加性以外,熵函数还应满足:非负性、对称性、上凸性、极值性、确定性等显而易见的性质。

在本节最后,我们讨论信源剩(冗)余度的概念,它是信源信息处理、信源编码的理论依据。它是信源统计分析中的一个非常重要的概念。由于实际信源几乎都是有记忆的,这也就是说信源输出的消息序列的各个消息之间存在着记忆,即统计关联,如果能首先定量地计算出这一统计关联引入的多余(冗余)度,就能充分利用它。下面,讨论多(冗)余度及其计数。前面讨论了熵的定义及性质,由熵的非负性、

Shannon 不等式以及熵的极值性，对于一个最简单二进制信源有下列基本不等式

$$0 \leqslant H(X/Y) \leqslant H(X) \leqslant \log_2 2 \tag{7.3.16}$$

其中 $\log_2 2$ 为离散二进制信源消息等概分布时的熵函数值(最大值、极值)。

下面，将公式(7.3.16)推广至离散多进制有记忆信源。

$$0 \leqslant H_\infty(X) = \lim_{L\to\infty}(X_L/X_{L-1}\cdots X_1) \leqslant \cdots \leqslant H_2(X_2/X_1) \leqslant H(X_1) \leqslant H_0(X) = \log N \tag{7.3.17}$$

式中，$l=1,2,\cdots,L$ 为信源记忆长度，$H_\infty(X)$为无限记忆长度时的信息熵，$H_2(X)$为一维记忆长度时的信息熵，$H_1(X)$为无记忆不等概率信息熵，$H_0(X)$为无记忆等概率信源最大熵。

由公式(7.3.17)可见，对于一般化的有记忆信源，最小的单个消息熵应为 $H_\infty(X)$，这就是说，从理论上看只需要在信道中传送 $H_\infty(X)$，在接收端则可利用信源统计关联的记忆特性，即无限维的全部概率特性，即可恢复出信源的全部信息。但是若人们不利用信源的统计特性与统计关联，信道中传送的则是信源的最大信息熵 $H_0(X)$，两者相比较，若不考虑、不利用信源的统计特性时，信道多传送的信息量为两者之差，即

$$H_0(X) - H_\infty(X) \tag{7.3.18}$$

为了定量地描述信源的有效性，定义下面两个概念

$$信源效率：\eta = \frac{H_\infty(X)}{H_0(X)} \tag{7.3.19}$$

$$信源相对剩余度：R = 1-\eta = \frac{H_0(X)-H_\infty(X)}{H_0(X)} = 1-\frac{H_\infty(X)}{H_0(X)} \tag{7.3.20}$$

公式(7.3.20)很重要，其中 $H_\infty(X)$表示考虑全部信源统计特性后信源的最小信息熵，它是信道传送的理论上最佳值。而 $H_0(X)$是不考虑信源统计特性，即认为信源消息均为等概率时信源的最大熵。公式(7.3.20)表示信源从统计特性上看的相对剩(冗)余度。一般说来，信源效率 η 越低，信源相对剩(冗)余度就越大，就越有必要对信源进行统计信息处理，采用信源编码、数据压缩也就越有必要。因此，可以认为信源的相对剩(冗)余度是信源编码和数据压缩的理论基础。

下面，举两个例子进一步说明信源编码和信源统计信息处理的必要性。首先以英文文字信源为例，讨论文字信源的剩(冗)余。根据英文中各个字母(含空格)出现的概率可列表 7.3.1。

表 7.3.1 英文字母出现概率统计表

字母	概率 $P(x_i)$	字母	概率 $P(x_i)$	字母	概率 $P(x_i)$
空格	0.2	S	0.052	Y,W	0.012
E	0.105	H	0.047	G	0.011
T	0.072	D	0.035	B	0.0105
O	0.0654	L	0.029	V	0.008
A	0.063	C	0.023	K	0.003
N	0.059	F,V	0.0225	X	0.002
I	0.055	M	0.021	J,Q	0.001
R	0.054	P	0.0175	Z	0.001

由上述表格，首先求得独立等概率情况下的 $H_0(X)$值：

$$H_0(X)=\log_2 27=4.76\ \text{bit} \tag{7.3.21}$$

再求独立不等概率情况下的 $H_1(X)$：

$$H_1(X)=-\sum_{i=1}^{27}P(x_i)\log_2 P(x_i)=4.03\ \text{bit} \tag{7.3.22}$$

还可进一步求得有一阶、二阶记忆下的 $H_2(X)$ 和 $H_3(X)$：

$$H_2(X)=3.32\ \text{bit} \tag{7.3.23}$$

$$H_3(X)=3.1\ \text{bit} \tag{7.3.24}$$

最后，利用统计推断方法求得 $H_\infty(X)$ 值。一般而言，由于采用不同的逼近方法和所取样本上的差异所推算的结果也有所不同，这里我们采用 Shannon 本人求得的推算值：

$$H_\infty(X)\approx 1.4\ \text{bit} \tag{7.3.25}$$

这样，利用公式(7.3.19)及(7.3.20)可分别求得 $\eta=0.29, R=0.71$。这一结论说明英文字母信源从理论上看，有 71%是多余的，即可以认为一本 100 页英文书，理论上看仅有 29 页是有效的，其余 71 页从统计角度看完全是多余的。也正是由于理论上存在着这一多余成分，引导了实际工作者对英文信源进行压缩编码的研究。

英文信源的分析也带动了各国对自己国家语言文字信源的分析，现将类似结果分别列于表 7.3.2。

表 7.3.2 不同文字信源剩余度估算

语种	H_0/bit	…	H_∞/bit	η	R
英文	4.7	…	1.4	0.29	0.71
法文	4.7	…	3	0.63	0.37
德文	4.7	…	1.08	0.23	0.77
西班牙文	4.7	…	1.97	0.42	0.58
中文	≈13(以 8 千字计)	…	4.1	0.315	0.685

下面，再给出语音信源剩(冗)余度的一个粗略估计：语音信源的编码大致可以分为波形编码、参量编码与混合编码三大类。这里，仅分析剩(冗)余最大的参量编码，即声码器的最大潜力。语音参量可以从不同角度给出，这里考虑潜力最大的从语音最基本参量单元音素出发，以英语为例，其音素大约有 $2^7(=128)\sim 2^8(=256)$ 个，若按人们通常讲话的速率，每秒钟大约平均发送 10 个音素，这时英语语音信源给出的信息率为

$$\text{上限：}I_1=\log_2(256)^{10}=80\ \text{bit/s} \tag{7.3.26}$$

$$\text{下限：}I_2=\log_2(128)^{10}=70\ \text{bit/s} \tag{7.3.27}$$

若按 PCM 常规数字化编码传送语音，其标准速率为 64 kbit/s，因此可求得

$$\eta_1=\frac{80}{64\ 000}=0.001\ 25 \tag{7.3.28}$$

$$\eta_2=\frac{70}{64\ 000}=0.001\ 1 \tag{7.3.29}$$

$$R_1=1-\eta_1=1-0.001\ 25=0.998\ 75 \tag{7.3.30}$$

$$R_2=1-\eta_2=1-0.001\ 1=0.998\ 9 \tag{7.3.31}$$

可见,语音参量编码潜力巨大。定义理论上最大压缩倍数如下:

$$K_1=\frac{64\ 000}{80}=800(\text{倍})$$

$$K_2=\frac{64\ 000}{70}=914(\text{倍})$$

7.4 互信息 $I(X;Y)$

上一节已指出:信息熵是信源输出符号所包含的信息量,而真正被接收者信源所获得的信息量则是互信息 $I(X;Y)$,互信息是与发送者 X,接收者 Y 双方都有关系的一个相对量,它既可理解为接收者信宿 Y 从发送者信源 X 中所获得的信息量,也可以理解为发送者信源 X 传送给接收者信宿 Y 的信息量,因此可记为 $I(X;Y)$。

从公式(7.3.14)、(7.3.15),即 Shannon 不等式,有

$$H(X)\geqslant H(X/Y)\geqslant 0 \tag{7.4.1}$$

$$H(Y)\geqslant H(Y/X)\geqslant 0 \tag{7.4.2}$$

移项后求得

$$H(X)-H(X/Y)\geqslant 0 \tag{7.4.3}$$

$$H(Y)-H(Y/X)\geqslant 0) \tag{7.4.4}$$

若令 X 为信源,Y 为信宿,则它们之间以互信息可直接定义为

$$\begin{aligned} I(X;Y)&=H(X)-H(X/Y)\\ &=E[-\log P(x_i)]-E[-\log P(x_i/y_j)]\\ &=E\left[-\log\frac{P(x_i)}{P(x_i/y_j)}\right]\\ &=E[i(x_i;y_j)] \end{aligned} \tag{7.4.5}$$

还可以等效定义为

$$\begin{aligned} I(X;Y)&=H(Y)-H(Y/X)\\ &=E[-\log P(y_j)]-E[-\log P(y_j/x_i)]\\ &=E\left[-\log\frac{P(y_j)}{P(y_j/x_i)}\right]\\ &=E[i(y_j;x_i)] \end{aligned} \tag{7.4.6}$$

其中,$i(x_i;y_j)$和 $i(y_j;x_i)$表示发、收某一对具体 x_i 与 y_j 值时的互信息,称它为信源与信宿之间的互信息密度,而互信息则可定义为互信息密度的统计平均值。这一定义与信息熵与信源的非平均信息量之间的定义完全相似。

对于互信息有下列主要基本数学性质:

(1)
$$\begin{aligned} I(X;Y)&=H(X)-H(X/Y)\\ &=H(Y)-H(Y/X)\\ &=H(X)+H(Y)-H(X,Y) \end{aligned} \tag{7.4.7}$$

(2)
$$I(X;Y)\geqslant 0$$

(3)
$$I(X;Y)\leqslant H(X)$$