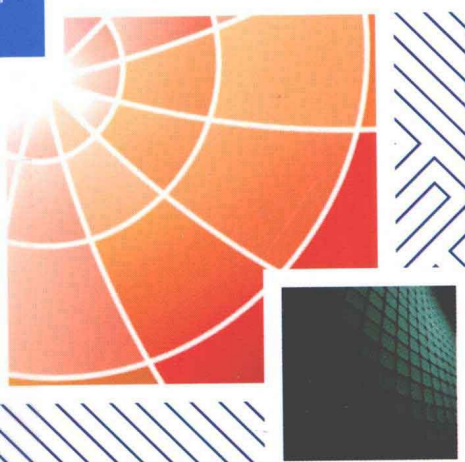
工业和信息化普通高等教育
“十二五”规划教材立项项目

21世纪高等院校信息与通信工程规划教材
21st Century University Planned Textbooks of Information and Communication Engineering

信息论基础与编码

王军选 田小平 曹红梅 编著

Elements of Information Theory and Coding



人民邮电出版社
POSTS & TELECOM PRESS

高校系列

工业和信息化普通高等教育
“十二五”规划教材立项项目

21世纪高等院校信息与通信工程规划教材
21st Century University Planned Textbooks of Information and Communication Engineering

信息论基础与编码

王军选 田小平 曹红梅 编著

Elements of Information Theory and Coding

人民邮电出版社
北京

高校系列

图书在版编目（CIP）数据

信息论基础与编码 / 王军选，田小平，曹红梅编著
— 北京 ：人民邮电出版社，2011.9
21世纪高等院校信息与通信工程规划教材
ISBN 978-7-115-25860-1

Ⅰ. ①信… Ⅱ. ①王… ②田… ③曹… Ⅲ. ①信息论
—高等学校—教材②信源编码—高等学校—教材 Ⅳ.
①TN911.2

中国版本图书馆CIP数据核字(2011)第154386号

内 容 提 要

本书系统地介绍了信息论基础与编码的主要内容，以及在无线通信系统中的主要应用。全书共9章，在介绍了有关信息度量的基础上，重点讨论了无失真信源编码、限失真信源编码、信道容量、信道编码和密码学的理论知识。全书从简单的理论入手，结合大量的例题描述信息论与编码的原理和应用，其中，原理的叙述力求突出概念和思路，尽量免去深奥的纯数学推导，与具体的应用相结合。在各章还附有相应的习题，便于学生加深理解。

本书可作为高等院校信息工程、通信工程以及电子信息类学生的教材，也可供低年级研究生或工程技术人员阅读参考。

21世纪高等院校信息与通信工程规划教材

信息论基础与编码

◆ 编　　著　王军选　田小平　曹红梅
责任编辑　贾　楠

◆ 人民邮电出版社出版发行　　北京市崇文区夕照寺街14号
邮编　100061　　电子邮件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
北京艺辉印刷有限公司印刷

◆ 开本：787×1092　1/16
印张：16.75　　2011年9月第1版
字数：406千字　　2011年9月北京第1次印刷

ISBN 978-7-115-25860-1

定价：32.80元

读者服务热线：(010)67170985　印装质量热线：(010)67129223
反盗版热线：(010)67171154
广告经营许可证：京崇工商广字第0021号

前言

“信息论基础与编码”是通信工程、电子信息工程类专业的基础课程。信息论与编码在光通信、无线通信领域中应用广泛。本书系统地介绍了信息论的信源编码、信道编码、密码编码、信道容量等方面的知识和内容。

信息论基础和编码的理论性内容较多，目前相关的教材其内容都有不同的偏重。根据读者不同的需求，信息论基础和编码也可以从不同的层次上去学习和理解：理论分析、技术层次和工程应用。目前全国几乎所有的工科、综合类高校都开设了通信工程或电子信息类专业，由于学生基础、培养侧重点以及办学层次的差异，对信息论基础与编码的学习要求各不相同。另一方面，对于工程技术人员而言，工作性质不尽相同，对信息论基础与编码知识要求的深度与广度必然也是不同的。正是这种需求的多样性，产生了对教材多层次、多样性的要求，这也是编写这本书的原因。

很多有关信息论基础、信息编码、通信原理、数字系统等教材中均有一些信息论以及编码部分的内容，这些内容对于大多数非专业研究人员来说显得过于深奥。因此，本书从简单的理论入手，结合大量的例题，从工程角度来描述信息论与编码的原理与应用，其中，原理的叙述力求突出概念和思路，尽量免去深奥的纯数学推导；设计与应用则尽量具体化，采用实例分析。

本书适合作为高等学校电子信息类高年级学生的教材，也可供低年级研究生和工程技术人员阅读参考。

本书编写分工如下：王军选编写第 1 章、第 3 章、第 6 章、第 7 章、第 8 章，田小平编写第 2 章、第 4 章、第 5 章，曹红梅编写第 9 章。张普等参加了书中部分内容的录入。全书由王军选统稿。在本书编写过程中，得到了很多同事的帮助和鼓励，在此表示感谢。

限于作者的水平，书中难免有不妥和谬误之处，敬请读者指正。

作　者

2011 年 5 月

目　录

第 1 章 概 论

内容简介

本章主要介绍信息论的基本概念、研究对象和内容，以及信息论的发展等。通过本章的学习可以掌握信息论的基本概念，为后面的学习打下坚实的基础。

1.1 信息的基本概念

研究信息论，首先要明白什么是信息，本节的内容主要是几个常用的关于信息的定义以及信息的基本性质。

1.1.1 信息的定义

现代社会已经进入一个新的时代——信息时代。人类的社会生活离不开信息，社会实践活动不仅需要对周围世界的情况能做出正确的反应，而且还要与周围的人群进行沟通。因此，人类不仅时刻需要从外界获取信息，而且还要和其他人交流信息。

什么是信息？至今无确切定义，但它是一种人人皆知的抽象概念，是一种不言自明的概念。信息在日常生活中常被认为是“消息”、“情报”、“信号”等，然而信息和它们既有联系也有明显的区别。消息一般是形式，没有具体的数学含义；情报的含义较窄，不像信息那么广泛；信号则是通信系统中消息的载体。

就狭义而言，在通信中对信息的表达分为 3 个层次：信号、消息、信息。

信号：是信息的物理表达层，是 3 个层次中最具体的层次。它是一个电参量、物理量，是一个载荷信息的运载工具，可测量、可描述，如正弦信号、脉冲信号等。

消息：（或称为符号），是信息的数学表达层，是信息的载体，它虽不是一个物理量，但是可以定量地加以描述，它是具体物理信号的进一步数学抽象，可将具体物理信号抽象为两大类型。

① 离散（数字）消息，是一组未知量，可用随机序列来描述：$\boldsymbol{U}=(U_1 \cdots U_l \cdots U_L)$。

② 连续（模拟）消息，也是未知量，它可用随机过程来描述：$U(t, \omega)$。

信息：它是更高层次上的抽象，是信号与消息的更高表达层次。

3 个层次中，信号最具体，信息最抽象。它们三者之间的关系是哲学上的内涵与外延的关系。

这就是说，信息可以认为是具体的物理信号、数学描述的消息的内涵。

而信号则是抽象信息在物理层表达的外延；消息则是抽象信息在数学层表达的外延。同一信息，可以采用不同的信号形式（如文字、语言、图像等）来载荷；同一信息，也可以采用不同的数学表达形式（如离散或连续）来定量描述。同样，同一信号形式，如“0”与“1”，可以表达不同形式的信息，比如无与有、断与通、低与高（电平）等。

在信息的具体定义上，从不同的侧面、角度、层次，有不同的定义。1928 年，哈特莱（R. V. L Hartley）提出，“发信者所发出的信息，就是他在通信符号表中选择符号的具体方式”，该定义只考虑通信符号的具体选择方式，没有涉及信息的价值和具体内容，也没有考虑各种不同选择方式的概率统计特性；维纳（N. Wiener）也曾指出，“信息是信息，不是物质，也不是能量”，将“信息”上升到最基本的概念的位置，后来他又指出“信息是人们在适应外部世界和控制外部世界的过程中，同外部世界进行交换的内容的名称”；1948 年，香农（C.E.Shannon）从研究通信系统传输的实质出发，对信息做了科学的定义，并进行了定性和定量的描述：**信息是事物运动状态或存在方式的不确定性的描述。**

1.1.2 信息的性质

信息虽然没有一个明确的定义，但是却具有两个明显的特征：广泛性与抽象性。

广泛性可从以下 3 个方面来理解：①信息的客观性，动物、大树、沙粒、水滴甚至大海等所有客观存在的事物都具有自己的信息；②人类的生存离不开对信息的处理，从原始社会的狩猎、耕种等生活开始，人类不停地感知、接收信息，人的神经系统在不停地传递信息，人的大脑则在不停地处理与决策信息，人与人之间又在不停地交流信息，人活在世上的百分之百时间都在自觉与不自觉地与信息打交道；③信息的积累，信息可以通过书本、技能等知识进行积累，人类依靠知识改造自然、适应自然，靠知识促进社会的发展与进步。

信息的抽象性体现在信息是组成客观世界并促进社会发展的最基本的要素之一。一般来讲，物质世界主要由物质、能量、信息等要素构成，其中，物质是基础；能量是物质运动的形式，是改造客观世界的主要动力；信息依附于物质和能量，但又不同于物质和能量。没有信息就不能更好地利用物质和能量，人类利用信息和知识改造物质，创造新物质，提高能量利用效率。

根据上面的叙述，信息应该具有如下性质：

① 信息是可以识别的；
② 信息的载体是可以转换的；
③ 信息是可以存储的；
④ 信息是可以传递的；
⑤ 信息是可以加工的；
⑥ 信息是可以共享的；
⑦ 信息是可以测量的。

1.2 信息论研究的对象和内容

信息论经过几十年的发展，已经有系统的理论和体系，研究的领域包括通信、电子、控制、管理等行业。特别是在通信领域，受益于信息论的指导，各种通信新技术层出不穷，带

来了人类社会的巨大变化。

1.2.1 信息论研究的对象

信息论是关于信息的本质和传输规律的科学理论。信息论应用概率论、随机过程、数理统计以及近世代数、矩阵理论等方法来研究信息的计量、发送、传递、交换、接收和储存的一般规律。下面就以图 1-1 所示的通信系统为例，说明香农信息论研究的对象。

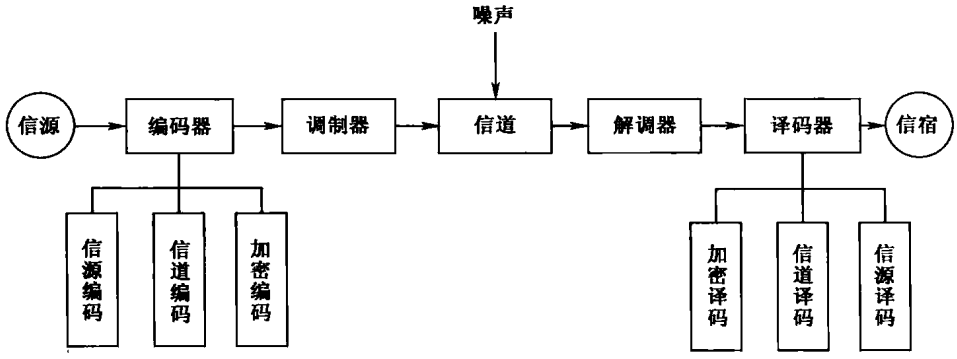


图 1-1 通信系统框图

在图 1-1 所示的通信系统中，形式上传输的是消息，但实质上传输的是信息。消息只是表达信息的载体。因此，在通信中被利用的（亦即携带信息的）载体是不重要的，而重要的是信息。通信系统主要分成 5 个部分。**信源**是产生消息和消息序列的来源，消息可以是离散的，也可以是连续的（数据、文字、语言、图像），通常信源的消息序列是随机发生的，因此要用随机变量来描述。**编码器**包括信源编码器、信道编码器和密码编码器，主要是提高通信系统的有效性、可靠性和安全性。**调制器**是将编码器输出的数字序列变换为振幅、频率或相位受到调制控制的信号，以适合在信道中进行较长距离的传输。**信道**是信号由发送端传输到接收端的媒介，常用的信道包括明线、电缆、高频无线信道、微波通道、光纤通道等，甚至包括磁盘等存储介质。**噪声**（干扰）是对传输信道或存储媒介构成影响的来源的总称，包括热噪声、电磁干扰等。噪声和干扰往往具有随机性，所以信道的特征也可以用概率空间来描述，根据统计特性，分为两类：①加性干扰（噪声），它是由外界原因产生的随机干扰，它与信道中传送的信号的统计特性无关，因而信道的输出是输入和干扰的叠加；②乘性干扰，信道的输出信号可看成是输入信号和一个时变参量相乘的结果。**解调器**是从载波中提取信号，是调制的逆过程，将调制的信号转换为携带信息的消息序列。**译码器**包括进行加密译码、信道译码以及信源译码，是编码器的逆过程，将消息序列转换成适合接收者接收的信息形式。**信宿**是信息的接受者。

通信的目的就是消除或部分消除不确定性，从而获得信息。信息论就是通过对系统中消息的传输和处理的研究来找出信息传输和处理的共同规律，具体到通信系统，就是研究通信系统的有效性、安全性、可靠性等环节。

1.2.2 信息论的基本定义

一般概率空间用$[X, P]$来表示。在离散情况下，X的样本空间可写成$[a_1, a_2, \cdots, a_q]$，样本空间的大小为q。样本空间中选择任意元素a_i的概率表示为$p(a_i)$。在离散情况下，概率空间

可描述为

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ p(a_1) & p(a_2) & \cdots & p(a_q) \end{bmatrix} \tag{1-1}$$

其中，$p(a_i)$就是选择符号 a_i 作为消息的概率，称为**先验概率**，并且满足$\sum_{i=1}^{q} p(a_i)=1$。在接收端，是否选择这个消息（符号）a_i的不确定性是与 a_i的先验概率成反比的，即对 a_i的不确定性可表示为先验概率 $p(a_i)$的倒数的某一函数。

自信息定义：对概率空间 X，其消息（符号）a_i的自信息定义为

$$I(a_i)=\log\frac{1}{p(a_i)}=-\log p(a_i) \tag{1-2}$$

在接收端，收到的消息集合为 Y。由于信道中存在干扰，假设接收端收到的消息（符号）为 b_j（$j=1,2,\cdots,r$），其中 r 可能与 q 相同，也可能不同，于是把条件概率 $p(a_i|b_j)$称为**后验概率**，它表示接收端收到消息（符号）b_j后，发送端发的是 a_i的概率。接收端收到 b_j后，发送端发送的是否是 a_i尚存在的不确定性应是后验概率的函数，即$\log\frac{1}{p(a_i|b_j)}$。于是，收信者在收到消息（符号）b_j 后，已经消除的不确定性为：先验的不确定性减去尚存在的不确定性，这就是收信者**获得的信息量**。

互信息定义：对概率空间 X 和 Y，X 的消息（符号）a_i和 Y 的消息（符号）b_j之间的互信息定义为

$$I(a_i;b_j)=\log\frac{1}{p(a_i)}-\log\frac{1}{p(a_i|b_j)} \tag{1-3}$$

当对数的底取 2 时，$I(a_i)$的单位为比特（bit）；当对数的底取 e 时，$I(a_i)$的单位为奈特（nat）；当对数的底取 10 时，$I(a_i)$的单位为哈特（hart）。

1.2.3 信息论研究的内容

目前，对信息论的研究内容一般有 3 种理解，如图 1-2 所示。

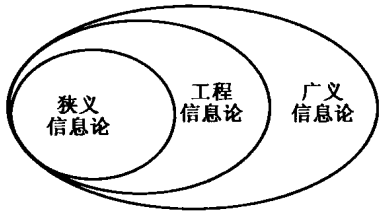


图 1-2 信息论的 3 种形式

狭义信息论（又称香农信息论）：主要通过数学描述与定量分析，研究通信系统从信源到信宿的全过程，包括信息的测度、信道容量、信源和信道编码理论等问题，强调通过编码和译码使收、发两端联合最优化，并且以定理的形式证明极限的存在，这部分内容是信息论的基础理论。狭义信息论是以存在性研究为主体，又称它为数学信息论。

工程信息论（又称一般信息论、通信理论）：主要是研究信息传输和处理问题，除了香农理论外，还包括噪声理论、信号滤波和预测、统计检测和估计理论、调制理论、信息处理理论等，比如信源编、译码理论及其设计构造方法，信道编、译码理论及其设计构造方法，最佳调制与解调理论与实现，最佳检测、估值与最佳接收理论与实现等。

广义信息论：广义信息论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、语言学等。

1.3 信息论的发展

自从19世纪20～30年代法拉第发现电磁感应以来，人类社会就开始了无线通信的探索。1832年莫尔斯建立了电报系统，1895马可尼发明了无线通信，1922年卡松提出边带理论，指明信号在调制（编码）与传送过程中与频谱宽度的关系，1922年哈特莱发表《信息传输》的文章，首先提出消息是代码、符号而不是信息内容本身，使信息与消息区分开来，并提出用消息可能数目的对数来度量消息中所含有的信息量，为信息论的创立提供了思路。美国统计学家费希尔从古典统计理论角度研究了信息理论，苏联数学家哥尔莫戈洛夫也对信息论做过研究。控制论创始人维纳建立了维纳滤波理论和信号预测理论，也提出了信息量的统计数学公式。这些研究均为信息论的建立打下了坚实的基础。

1948年贝尔研究所的香农在题为《通信的数学理论》的论文中系统地提出了关于信息的论述，创立了较为系统的信息论学科。维纳提出的关于度量信息量的数学公式开辟了信息论的广泛应用前景。1951年信息论学科获得美国无线电工程学会承认，为以后的迅速发展打下基础。20世纪50～60年代是信息论发展的关键时期，信源编码和信道编码得到了巨大的发展，产生了大量的研究成果，比如霍夫曼编码、费诺编码以及信道编码中的卷积码、LDPC编码等，同时，该时期也是信息论向其他各门学科渗透的时期。到20世纪70年代，由于数字计算机的广泛应用，通信系统的能力也有了很大提高。信息的概念和方法已广泛渗透到各个科学领域，它迫切要求突破香农信息论（狭义信息论）的范围，以便使它能成为人类各种活动中所碰到的信息问题的基础理论，从而推动其他许多新兴学科进一步发展。20世纪90年代以来，Turbo码编码理论、多天线理论等技术的提出，使得无线通信迅速发展，带来了人类社会的巨大变化。

信息论从诞生到今天已有60多年了，现已成为一门独立的学科。香农理论的思想、方法，甚至某些结论已渗透到统计数学、计算机科学、物理学、哲学、科学方法论等其他学科中。

同时，信息论还和其他学科结合产生了许多交叉学科，比如在经济、生物等方面已产生了“信息经济学”、“信息生物学”等边缘学科；在和量子力学理论结合后，产生了量子信息论、量子编码理论、量子计算理论等。

第2章 信源与信息熵

内容简介

信源是输出信息的源，信源产生的信息是可以度量的。本章介绍了如何来描述信源的数学模型、信源的分类、互信息、信息熵的定义以及性质等；同时介绍了离散信源、马尔可夫信源以及连续信源的信息熵及性质。

2.1 信源的数学模型及分类

作为 Shannon 信息论研究的对象——信息，被假设为由一系列的随机变量所代表。它们往往用随机出现的符号来表示，我们称输出这些符号集的源为信息源，或者说信息源是发出消息的源，简称信源。

由信号源输出的随机符号，如果其取值于某一连续区间，则称此信息源为连续信源，如语言、图像、图形等；如果其取值于某一离散集合，则称此信息源为离散信源，如文字、数字、数据等符号。

离散信源可进一步分类如下。

① 离散无记忆信源，包括发出单个符号的无记忆信源和发出符号序列的无记忆信源两种。

发出单个符号的无记忆信源每次只发出一个符号代表一个消息，而发出符号序列的无记忆信源每次发出一组含两个以上符号的符号序列代表一个消息。

离散无记忆信源所发出的各个符号是相互独立的，发出的符号序列中的各个符号之间没有统计关联性。

② 离散有记忆信源，包括发出符号序列的有记忆信源和发出符号序列的马尔可夫信源两种。

和离散无记忆信源不同，离散有记忆信源所发出的各个符号的概率是有关联的，这种概率关联性可用两种方式来表示：一种是用信源发出的一个符号序列的联合概率来反映有记忆信源的特征，即发出符号序列的有记忆信源；一般情况下，当记忆长度很长甚至无限长时，表述有记忆信源要比无记忆信源困难得多，实际问题中，往往限制记忆长度，也就是说，某一符号出现的概率只与前面一个或有限个符号有关，与更前面的符号无关，这类信源可以用信源发出符号序列内各个符号之间的条件概率来反映有记忆信源的特征，即发出符号序列的

马尔可夫信源。

假定某一离散信源发出的各个符号消息的集合为

$$X=\{x_1,x_2,\cdots,x_n\}$$

各个符号的先验概率为

$$P=\{p(x_1),p(x_2),\cdots,p(x_n)\}$$

通常将它们写在一起

$$\begin{bmatrix}X\\P\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\\ p(x_1) & p(x_2) & \cdots & p(x_n)\end{bmatrix} \tag{2-1}$$

称为概率空间，其中 $p(x_i)\geqslant 0$，$\sum_{i=1}^{n}p(x_i)=1$。

最简单的有记忆信源是 N=2 的情况，此时信源为：$X=X_1X_2$，其概率空间为

$$\begin{bmatrix}X\\P\end{bmatrix}=\begin{bmatrix}x_1x_1 & x_1x_2 & \cdots & x_nx_n\\ p(x_1x_1) & p(x_1x_2) & \cdots & p(x_nx_n)\end{bmatrix} \tag{2-2}$$

对于无记忆信源来说，其联合概率为

$$p(x_1x_2\cdots x_n)=p(x_1)p(x_2)\cdots p(x_n)$$

若其满足平稳性

$$p(x_1x_2\cdots x_n)=p(x_1)p(x_2)\cdots p(x_n)=p^n$$

2.2 离散信源熵和互信息

离散信源的自信息和互信息是理解信息论的基础，本节主要介绍离散信源的信息熵以及互信息的概念和性质，并通过例子分析信源熵的计算。

2.2.1 信息量

在一切有意义的通信中，虽然消息的传递意味着信息的传递，但对于接收者来说，某些消息比另外一些消息却含有更多的信息。例如，若一方告诉另一方一件非常可能发生的事件“今年冬天的气候要比去年冬天的更冷一些”，比起告诉另一方一件很不可能发生的事件“今年冬天的气候将与去年夏天的一样热”来说，前一消息包含的信息量显然要比后者少些。因为前一事件很可能发生，不足为奇，但后一事件却极难发生，听后使人惊奇！这表明：消息确实有量值的意义。

可以看出：对接收者来说，事件越不可能发生，越是使人感到意外和惊奇，信息量就越大；事件越是有可能发生，信息量就越小。

概率论告诉我们，事件的不确定程度，可用其出现的概率来描述。事件出现的可能性愈小，则其概率就越小；反之，则其概率就越大。

由此我们可以得到：消息中的信息量与消息发生的概率紧密相关，消息出现的概率越小，则消息中包含的信息量就越大。

如果消息是必然的（概率为 1），则它传递的信息量应为 0。如果事件是不可能的（概率

为 0），则它将有无穷的信息量。

如果我们得到的不是由一个事件构成，而是由若干个独立事件构成的消息，那么，这时我们得到的总的信息量就是若干个独立事件的信息量的总和。

综上可以看出，为了计算信息量，消息中所含信息量 I 与消息出现的概率 $p(x)$间的关系式应当反映如下规律。

① 消息中所含信息量 I 是出现该消息的概率 $p(x)$的函数，即

$$I = I[p(x)] \tag{2-3}$$

② 消息出现的概率越小，它所含的信息量越大；消息出现的概率越大，它所含的信息量越小；且当 $p(x)=1$ 时，$I=0$。

③ 若干个互相独立的事件构成的消息，其所含信息量等于各独立事件的信息量之和，即

$$I[p(x_1)p(x_2)\cdots] = I[p(x_1)] + I[p(x_2)] + \cdots \tag{2-4}$$

不难看出，若 I 与 $p(x)$之间的关系式为

$$I = \log_a \frac{1}{p(x)} = -\log_a p(x) \tag{2-5}$$

就可满足上述要求。

上式中的 I 称为随机事件的自信息量。它的单位与所采用的对数的底有关。若取 $a=2$，则自信息量的单位为 bit；若取 $a=e$，则其单位为 nat；若取 $a=10$，则其单位为 det。

对于一个以等概率出现的二进制码元(0, 1)，它所包含的信息量为

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

若有一个 m 位的二进制数，其自信息量为

$$I = -\log_2 \frac{1}{2^m}\text{bit} = \log_2 2^m \text{bit} = m\ \text{bit}$$

即需要 m bit 的信息来指明这样的二进制数。

若有两个消息 x_i，y_j同时出现，则其自信息量定义为

$$I(x_i y_j) = -\log_a p(x_i y_j)$$

$p(x_iy_j)$为联合概率。

若 x_i与 y_j相互独立，即 $p(x_i y_j) = p(x_i)p(y_j)$，则有

$$I(x_i y_j) = I(x_i) + I(y_j)$$

若 x_i与 y_j的出现不是相互独立的，而是有联系的，此时，要用条件概率 $p(x_i \mid y_j)$来表示，即在事件 y_j出现的条件下，事件 x_i发生的条件概率，其条件自信息量定义为

$$I(x_i \mid y_j) = -\log_a p(x_i \mid y_j)$$

例题 2-1 英文字母中“a”出现的概率为 0.063，“c”出现的概率为 0.023，“e”出现的概率为 0.105，分别计算它们的自信息量。

解 由自信息量的定义式（2-5），有

$$I(a) = -\log_2 0.063\text{bit} = 3.96\text{bit}$$

$$I(c) = -\log_2 0.023\text{bit} = 5.44\text{bit}$$

$$I(e) = -\log_2 0.105\text{bit} = 3.25\text{bit}$$

例题 2-2 将二信息分别编码为 A 和 B 进行传送，在接收端，A 被误收作 B 的概率为 0.02；而 B 被误收作 A 的概率为 0.01，A 与 B 传送的频繁程度为 2∶1。若接收端收到的是 A，计算原发信息是 A 的条件自信息量。

解 设 U_0 表示发送 A，U_1 表示发送 B；V_0 表示接收 A，V_1 表示接收 B。

由题意知：$p(U_0)=\frac{2}{3}$，$p(U_1)=\frac{1}{3}$，$p(V_1|U_0)=0.02$，$p(V_0|U_0)=0.98$，$p(V_0|U_1)=0.01$，$p(V_1|U_1)=0.99$。则接收到 A 时，原发信息是 A 的条件概率为

$$\begin{aligned} p(U_0|V_0) &= \frac{p(U_0V_0)}{p(V_0)} = \frac{p(V_0|U_0)p(U_0)}{p(V_0)} \\ &= \frac{p(V_0|U_0)p(U_0)}{p(V_0|U_0)p(U_0)+p(V_0|U_1)p(U_1)} \\ &= \frac{0.98\times\frac{2}{3}}{0.98\times\frac{2}{3}+0.01\times\frac{1}{3}} \\ &= \frac{196}{197} \end{aligned}$$

相应的条件自信息量为

$$I[p(U_0|V_0)] = -\log_2\frac{196}{197} = 0.0734\text{ bit}$$

2.2.2 离散信源熵

假设离散信息源是一个由 n 个符号组成的集合，称为符号集。符号集中每一个符号 x_i 在消息中是按一定的概率 $p(x_i)$独立出现，其概率空间为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$$

且有 $\sum_{i=1}^{n} p(x_i)=1$，则 $x_1, x_2, \cdots, x_n$ 所包含的信息量分别为

$$-\log_2 p(x_1), -\log_2 p(x_2), \cdots, -\log_2 p(x_n)$$

于是，每个符号所含信息量的统计平均值，即平均信息量为

$$\begin{aligned} H(X) &= p(x_1)[-\log_2 p(x_1)] + p(x_2)[-\log_2 p(x_2)] + \cdots + p(x_n)[-\log_2 p(x_n)] \\ &= -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \end{aligned}$$

由于 H 同热力学中的熵形式相似，故通常又称它为信息源的熵，简称信源熵，其单位为 bit/符号。

前面定义的自信息量 $I(x_i)$是表征信源中各个符号的不确定性。由于信源中各个符号的概率分布（先验概率）不同，因而各个符号的自信息量就不相同，所以，自信息量不能作为信

源总体的信息量。而平均信息量 $H(X)$或信源熵 $H(X)$是从平均意义上来表征信源的总体特征，因此可以表征信源的平均不确定性。

定义信源的平均不确定性 $H(X)$为信源中各个符号不确定性的数学期望，即

$$H(X)=E[I(x)]=\sum_i p(x_i)I(x_i)=-\sum_i p(x_i)\log_2 p(x_i) \tag{2-6}$$

称作信息熵，简称熵。

由于 $I(X)$是非负值的量，所以熵 $H(X)$也是非负值的，只有在 $p(x)=0$ 和 $p(x)=1$ 时，熵 $H(X)$才为零[$p(x)=0$ 时，规定 $0\log 0=0$，其合理性由极限 $\lim\limits_{x\to 0^+} x\log x=0$ 得证]。

例题 2-3 一幅 500×600 的图像，每个像素的灰度等级为 10，若为均匀分布，计算平均每幅图像能提供的信息量。

解 能够组成的图像有 $n=10^{3\times 10^5}$ 个，有

$$H(X)=-\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)=-\log_2 10^{-(3\times 10^5)}=9.96\times 10^5 \text{ (bit/图像)}$$

例题 2-4 某信息源由 4 个符号 0,1,2,3 组成，它们出现的概率分别为：1/8、1/2、1/4、1/8，且每个符号的出现都是相互独立的。试计算某条消息“2010201302130012032101003210100231020020103120321001202 10”的信息量。

解 在此条消息中，符号 0 出现了 23 次，符号 1 出现了 14 次，符号 2 出现了 13 次，符号 3 出现了 7 次，消息共有 57 个符号。

其中出现符号 0 的信息量为：$23\log_2 8=69$bit。

其中出现符号 1 的信息量为：$14\log_2 2=14$bit。

其中出现符号 2 的信息量为：$13\log_2 4=26$bit。

其中出现符号 3 的信息量为：$7\log_2 8=21$bit。

因此，该消息的信息量为：I=69+14+26+21=130bit。平均（算术平均）每个符号的信息量应为：$\overline{I}=\dfrac{I}{57}=2.28\text{bit/符号}$。

若用信源熵的概念进行计算，有

$$H=\left(-\frac{1}{8}\log_2\frac{1}{8}\log_2-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{4}\log_2\frac{1}{4}-\frac{1}{8}\log_2\frac{1}{8}\right)\text{bit/符号}=1.75\text{bit/符号}$$

那么，该条消息所含的信息量为：$I=57\times 2.28\text{bit}\approx 129.96\text{bit}$。

可以看到，用算术平均和信源熵两种计算所得的结果略有差别，其根本原因在于它们平均处理的方法不同。按算术平均的方法，结果可能存在误差，这种误差将随着消息中符号数的增加而减小。

例题 2-5 一个由字母 A、B、C、D 组成的字，对于传输的每一个字母用二进制脉冲编码，A:00，B:01，C:10，D:11，每个脉冲的宽度为 5ms。

（1）若每个字母是等概率出现时，计算传输的平均信息速率。

（2）若每个字母出现的概率分别为：$p_A=0.2$，$p_B=0.25$，$p_C=0.25$，$p_D=0.3$，计算传输的平均信息速率。

解 信息传输速率也称为信息速率或传信率，它被定义为每秒传递的信息量，单位是比特/秒，记为 bit/s，它是衡量通信系统的一个主要性能指标。