信息论与编码

Information Theory & Coding

姜 丹 编著

中国科学技术大学出版社

信息论与编码

姜 丹 编著

中国科学技术大学出版社
2001·合肥

内 容 简 介

本书系统论述香农信息论的基本理论，介绍编码的基本方法. 全书共分九章. 内容包括：信息的定义、信息论的基本思路；单符号离散信源与信道、信息熵、互信息、信道容量、数据处理定理、加权熵；多符号离散信源与信道、极限熵、独立并列信道的信道容量；连续信源与信道、相对熵、高斯白噪声加性信道的信道容量；无失真信源编码定理、霍夫曼编码方法；抗干扰信道编码定理、线性分组码、汉明码与扩展汉明码；限失真信源编码定理、信息率失真函数、平均失真度、数据压缩原理、信息价值、广义信息率失真函数；网络信息理论等.

本书可作为高等院校、科研院所相关专业的本科生、研究生的教材或教学参考书，也可供从事信息理论、信息技术和信息科学的教学、科研和工程技术人员参考.

图书在版编目(CIP)数据

信息论与编码/姜丹 编著. —合肥：中国科学技术大学出版社，2001.8

ISBN 7-312-01260-4

Ⅰ. 信… Ⅱ. 姜… Ⅲ. 信息论-编码技术-高等学校-教材 Ⅳ. G201

中国版本图书馆 CIP 数据核字(2001)第 03459 号

中国科学技术大学出版社出版发行

(安徽省合肥市金寨路 96 号，邮编：230026)

中国科学技术大学印刷厂印刷

全国新华书店经销

开本：787×1092/16　　印张：38　　字数：950 千

2001 年 8 月第 1 版　　2001 年 8 月第 1 次印刷

印数：1—5000 册

ISBN 7-312-01260-4/TN・44　　定价：45.00 元

前　言

随着科学技术，特别是信息技术的发展，信息理论在通信领域中发挥越来越重要的作用，显示出它是解决通信领域中有关问题的有力工具的本色. 同时，由于信息理论解决问题的思路和方法的独特、新颖和有效，在当今的信息时代，信息理论已渗透到其它相关的自然科学，甚至社会科学领域，与电子技术、自动控制、计算机网络以及管理科学、生物医学工程、遗传工程、人工智能、心理学等学科密切结合，显示出它的勃勃生机和不可估量的发展前景. 信息论是信息科学中最成熟、最完整、最系统的一个重要组成部分，它是信息科学的发展起源与基石.

本书以香农(C. E. shannon)信息论为基础，论述近代信息理论的基本概念和主要结论.

作者鉴于20余年的教学经验，为了有利于读者正确认识通信领域中信息的定义和本质，理解信息论解决问题的思路和方法，在“引言”中归纳提炼出香农信息论的三大理论支柱. 为了有利于读者建立关于信息流通的完整系统概念，把信息论的基础理论部分由传统的“信源一条线”、“信道一条线”的“纵向结构”，改变成由“单符号离散通信系统”(第一章、第二章)、“多符号离散通信系统”(第三章)、“单维连续通信系统”(第四章)、“多维连续通信系统”(第五章)等四个“横向教学板块”组成的“横向结构”. 遵照由易到难、由浅入深、循序渐进的教学规律，顺序安排教学内容. 信息论是一门既具有严密的演绎系统，又具有丰富生动的物理含义的科学理论. 为了有利于帮助读者排除学习信息论过程中经常遇到的数学分析和正确理解物理概念两方面的困难，结合有关内容，系统而简明地介绍必要的数学基础知识. 对信息论的重要结论，尽量给出详细的数学推演过程和证明的不同途径和方法，并通俗形象地阐明其物理概念.

本书按照理论联系实际的原则，从数学上系统严密地证明和论述了“无失真信源编码定理”、“抗干扰信道编码定理”、“限失真信源编码定理”等信息论中的三大定理，介绍“霍夫曼(Huffman)码”、“线性分组码”等实际编码方法. 使读者既能看到实现有效而可靠的通信系统的光明前景，又能掌握某些实现有效、可靠通信的实际编码方法. 本书不仅介绍“信息率失真函数”导致“限失真信源编码定理”的演绎过程，而且论述“信息率失真函数”在“数据压缩”、“信息价值”等方面的应用. 凝练“信息率失真函数”的数学精髓，构建“广义信息率失真函数”，估算通信系统有关指标界限. “网络信息理论”这一章，以“多用户信道”的容量界限为重点，初步讨论、介绍网络信息传输的有关特性，以供读者探究当今正在蓬勃兴起的互联网络通信理论的一个初步基础知识.

本书既适用于信息论初学者，也有助于已具信息论初步知识的读者在更高层次上深入研究和应用信息理论. 它可作为高等院校、科研院所的研究生、高年级本科生的教材或教学参考书. 本书也可供从事信息理论、信息技术和信息科学的教学、科研以及工程技术人员参考.

热忱希望广大读者对书中的错误和不当之处予以批评指正.

姜　丹

2000年11月于北京

引　言

信息论是人们在长期通信实践活动中，由通信技术与概率论、随机过程、数理统计等学科相结合而逐步发展起来的一门新兴交叉学科. 美国科学家香农(C. E. Shannon)于 1948 年发表的著名论文《通信的数学理论》，奠定了信息论的理论基础.

随着科学技术，特别是信息技术的迅猛发展，在当今信息时代中，“信息”这个词汇逐步被人们所接受，并得到越来越广泛的应用. 当人们收到由电话、传真、广播、电视等媒体传播的消息后，往往说成获得了“信息”. 一般以数字、数据、图表、曲线等形式出现的计算机通信、运算和处理所需要的条件、内容和结果，人们也习惯称之为“输入信息”、“输出信息”、“补充信息”、“反馈信息”等等. 人们通常也把由眼、耳、鼻、口等感觉器官直接感觉到的颜色、声音、气味、味道、气温、湿度说成是感知到了某种外界环境“信息”. 确实，在人类社会中，信息的传递与交换，是时时处处都发生着的事情. 无数看不见的信息溪流，昼夜川流不息，聚成汹涌澎湃的信息洪涛，汇成浩瀚无垠、波澜壮阔的信息海洋. 人类就生活在这样一个信息的海洋之中.

通信是人类活动中最为普遍的现象之一. 在频繁的信息传递和交换中，人们总是希望有效、可靠地传递信息. 那么，自然而然地会提出这样一个问题，用什么标准来衡量通信的有效程度和可靠程度？怎样判断通信方法的优劣？显然，解决这些问题的关键在于解决信息的度量问题. 例如，某人收到一封来信，谈的是同学最近的工作、学习情况. 同时又收到一封家信，谈的是家人的健康情况. 显然他从这两封信中都获得了信息. 但若要问：他从哪一封信中获得了更多的信息？也许，按某种想当然的感觉，他会给出某种模糊的回答，如“家信中得到了更多的信息”. 这个结论可靠吗？就算这个结论不错，如进一步问：“家信中含有的信息比同学来信中含有的信息多了多少？”一般来说，很难回答这个问题.

为什么很难回答以上的问题呢？其主要原因在于对信息的本质缺乏明确的认识，经验性地把“信息”与“消息”混为一谈.

众所周知，消息是用文字、符号、数据、语言、音符、图片、图像等能被人们的感觉器官所感知的形式，对客观物质运动和主观思维活动状态的一种表述. 不同的消息，不仅有不同的形式，而且含有不同的内容和不同的效用. 例如，“中国男子体操队获 27 届奥运会团体冠军”这条消息. 在形式上，可把它看作是从汉字表中挑选 17 个字的一种选择. 从语义上分析，这条消息就相当复杂. 是“中国”，而不是“美国”、“俄罗斯”等其它国家. 是“男子”，而不是“女子”. 是“体操队”，而不是“足球队”、“篮球队”、“排球队”等其它代表队. 是“27 届”，而不是“26 届”、“25 届”等其它届. 是“奥运会”，而不是“全运会”、“亚运会”等其它运动会. 是“团体”，而不是“个人”、“双人”等其它项目. 是“冠军”，而不是“亚军”、“季军”. 从语用上来说，这条消息对中国人和俄罗斯等外国人来说所引起的反响程度显然是大不相同的. 要解决信息的度量问题，必然要应用数学工具，进行数量的运算. 我们知道，数学是刻画物质运动形式的工具，用数学对消息的形式进行刻画，不存在法则上的困难. 但如何运用数学工具刻画消息含有的语义乃至语用，至今仍然是一个巨大的难题. 所以，如把信息与消息混为一谈，把形式、语义、语用三个因素交织在一起，综合地解决信息度量问题，必然面临头绪纷繁，无从下手的僵局.

美国科学家香农(C. E. Shannon)针对人类通信活动的特点，精辟地提出了“形式化假说”、“非决定论”、“不确定性”等三个论点，以新颖的思想和方法，打破了这个僵局，跨出了用数学方法定量

描述信息的关键一步，开创了通信领域信息理论新局面.

(一)形式化假说

通过对通信活动的基本功能的观察分析，香农指出，“通信的基本问题是在消息的接收端精确地或近似地复制发送端所挑选的消息.通常消息是有意义的，即是说，它按某种关系与某些物质或概念的实体联系着.通信的语义方面的问题与工程问题是没有关系的.”这就是说，通信的任务只是在接收端把发送端发出的消息从形式上复制出来，通信工程并不须要对复制出来的消息的语义作任何处理和判断.对消息的语义内容的处理和判断，是接收者自己的事，不是通信工程师本身的任务，与通信工程师无关.至于消息的效用问题，更应该是接收者自己的感受问题，与传送消息的通信系统无关.例如电视屏幕上出现一则消息，有的观众看了兴高采烈；有的观众看了满腔愤怒，甚至把电视机从楼上掷到楼下；有的观众看了漠不关心，毫无反应.不论不同的观众有什么不同的效用反应，对电视通信工程来说，已经完成了它本身的任务.这就是香农对通信活动的“形式化”假说.

这种通信工程的“形式化”假说，大胆地去掉了消息的语义，语用因素，巧妙地保留了能用数学描述的形式因素，使用数学工具定量度量信息成为可能，打开了信息理论进入科学殿堂的大门.

(二)非决定论

经过对通信活动的对象和过程的分析研究，香农指出，“重要的是，一个实际的消息，总是从可能发生的消息集合中选择出来的.因此，系统必须设计得对每一种选择都能工作，而不是只适合工作于某一种选择.因为，各种消息的选择是随机的，设计者事先无法知道什么时候会选择什么消息来传送.”这就是说，一切有通信意义的消息的发生都是随机的，是事先无法预料的.消息传递过程中遇到的噪声干扰也是随机的，通信系统的工程设计者也是无法事先预料的.面对公众的通信系统，不是针对某一特定的通信对象设计的，什么样的用户，什么时候使用，传递什么样的消息都是无法始料的.显然，根据通信工程系统的这些特点，必须应用概率论、随机过程、数理统计等数学工具，从大量不可预料的随机消息(包括噪声)中，寻求其统计规律，作为通信工程师设计通信系统的依据，用非决定论观点揭示信息的本质.这就是香农看待通信活动的“非决定论”观点.

这种“非决定论”观点是对通信活动的总的认识观，它从原则上解决了用什么样的数学工具解决信息度量问题.

(三)不确定性

通过对通信活动的机制和本质的分析研究，香农一针见血地指出，“人们只在两种情况下有通信的需要.其一，是自己有某种形式的消息要告知对方，而估计对方“不知道”这个消息；其二，是自己有某种“疑问”，要询问对方，而估计对方能作出一定的解答.这里的所谓“不知道”、“疑问”，就是通信前对某事件可能发生的若干种结果不能作出明确的判断，存在某种知识上的“不确定性”.通信后，通过消息的传递，由原先的“不知道”到“知道”，或由“知之不多”到“知之甚多”；原先的“疑问”得到了解答，或部分解答，由原先的“疑问”到“明白”，或部分“明白”.这就是说，通信后，消除了或部分消除了通信前存在的不确定性.所以，通信的作用就是通过消息的传递，使接收者从收到的消息中获取了一定的信息，消除了原先存在的某些不确定性.这样，我们就有理由明确地说，“信息就是用来消除不确定性的东西”，通信后接收者获取的信息，在数量上等于通信前后不确定性的消除量.这就是香农从“不确定性”观点出发，给“信息”下的明确的定义.

我们知道，“不确定性”是与“可能性”相联系的.“可能性”的大小在数学上可以用概率的大小来表示.概率大即表示出现的“可能性”大；概率小即表示出现的“可能性”小.而“可能性”大就意味着“不确定性”小；“可能性”小就意味着“不确定性”大.这样，“不确定性”就可与消息发生的概率联系起来.例如，“中国女子乒乓球队夺取亚运会冠军”这条消息，根据中国女子乒乓球队历来的表现，夺

取亚运会冠军的概率很大，即可能性很大，也就意味着“不确定性”很小. 这个消息一旦发生，消除的不确定性也很小. 收信者从这条消息中获取的信息量也很小. 相反，“中国男子足球队夺取世界杯赛冠军”这条消息，根据中国男子足球队历来的表现，夺取世界杯赛冠军的概率很小，即“可能性”很小，也就意味着不确定性很大. 若有朝一日这个消息真的发生了，消除的“不确定性”很大，收信者从这条消息中获取的信息量也很大，甚至惊喜万分、欢呼跳跃. 由此可见，“不确定性”与消息发生的概率有内在联系，它应该是消息发生概率的某一函数. 根据香农关于信息的定义，通信后收信者从消息中获取的信息，从数量上等于通信前后“不确定性”的消除. 当然，通信后获取的信息量也应该是消息发生概率的某一函数. 这样，从理论原则上完全解决了信息的度量问题.

香农从“不确定性”观点出发对信息的明确定义告诉我们，“信息”与“消息”两者之间既有联系，又有区别，两者不应混为一谈. “消息”是表达“信息”的形式，是载荷“信息”的客体；“信息”是“消息”统计特性的参量，是“消息”的抽象本质. 不同形式的“消息”，可能有相同数量的“信息”，相同形式的“消息”，可能有不同数量的“信息”. 信息论的研究对象不是具体的消息，而是抽象于各种不同形式的“消息”的“信息”. 信息论是一门高度抽象和概括的学科.

信息论是起源于通信领域，描述通信活动规律的一门学科. 它是广义信息科学中最完整、最系统、最成熟的一个重要组成部分，是信息科学继续发展的起点和基石.

目　次

第一章　单符号离散信源

通信系统一般由信源、信道和信宿三部分组成(图 1.1).“信源”就是信息的源泉.信息不是消息本身,但它又包含在消息之中.信源是由含有信息的消息组成的集合.若信源是由有限或无限可

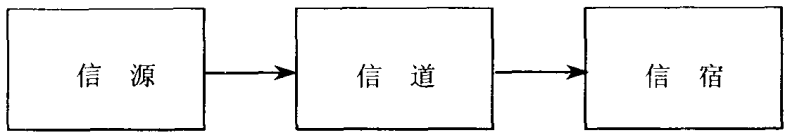

图 1.1

列个取值离散的符号(如文字、字母、数字等)组成的离散集合,则这种信源称为离散信源.又若一个符号就代表一个完整的消息,则这种离散信源又称为单符号离散信源.单符号离散信源是最简单的离散信源.

第一节　信源的数学模型

信源要含有一定的信息,必须具有随机性,以一定的概率发出各种不同的符号.单符号离散信源是具有一定概率分布的离散符号的集合.基于对信源的这种认识,我们可用一个离散随机变量的可能取值,表示信源可能发出的不同符号;用离散随机变量的概率分布,表示信源发出不同符号可能性的大小.总之,我们可用一个离散随机变量来代表一个单符号离散信源.

例如,掷一个六面质地均匀的骰子,每次出现朝上一面的点数是随机的.如把出现朝上一面的点数作为这个随机试验的结果,并把试验的结果看作信源的输出消息,无疑,这个随机试验可看作是一个信源.这个信源输出有限种离散数字,其组成的集合为 $A:\{1,2,3,4,5,6\}$,而且每一个数字代表一个完整的消息.所以,这个信源是单符号离散信源.我们可用离散随机变量 X 来表示这个单符号离散信源:X 的可能取值就是信源可能发出的各种不同符号,其状态空间就是信源可能发出的各种不同符号组成的集合 $A:\{1,2,3,4,5,6\}$;X 的概率分布,就是信源发出各种不同符号的先验概率,其概率空间就是信源发出各种不同符号的先验概率组成的概率空间 $P:\left\{P(X=1)=\frac{1}{6},P(X=2)=\frac{1}{6},\cdots,P(X=6)=\frac{1}{6}\right\}$.所以,这个单符号离散信源的数学模型可完整地表示为

$$[X\cdot P]:\begin{cases} X: & 1 & 2 & 3 & 4 & 5 & 6 \\ P(X): & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{cases}$$

我们把$[X\cdot P]$称为信源 X 的“信源空间”.信源 X 输出的符号只可能是集合 $A:\{1,2,3,4,5,6\}$中的任何一种,不可能是集合 A 以外的其它任何符号.信源 X 的概率空间是一个完备集,即有

$$P(X=1)+P(X=2)+\cdots+P(X=6)=1$$

在这个典型实例的启发下,我们可构建一般单符号离散信源的数学模型.若某信源可能发出 r 种不同的符号 $a_1,a_2,\cdots,a_r$,相应的先验概率分别是 $p(a_1),p(a_2),\cdots,p(a_r)$.我们用随机变量 X 表示这个信源,其信源空间可表示为

$$[X\cdot P]:\begin{cases} X: & a_1 & a_2 & \cdots & a_r \\ P(X): & p(a_1) & p(a_2) & \cdots & p(a_r) \end{cases} \tag{1.1}$$

其中

$$0 \leqslant p(a_i) \leqslant 1 \qquad (i = 1,2,\cdots,r)$$
$$\sum_{i=1}^{r} p(a_i) = 1 \tag{1.2}$$

不同信源对应不同的信源空间. 如信源给定,这就意味着相应的信源空间已经确定. 反之,如信源空间已经确定,这就意味着相应的信源已经给定. 用信源空间表示信源的数学模型的必要前提,就是信源可能发出的各种不同符号的概率先验可知,或事先可测定的. 测定信源的概率空间是构建信源空间的关键. 例如,在一个箱子中,有红、黄、蓝、白四种不同颜色的彩球,它们的大小、质量和重量完全一样. 若从这个箱子中任意摸取出一个球,并把球的颜色当作试验的结果. 显然,这个随机试验就可看作是一个单符号离散信源,信源的输出符号集就是四种不同的颜色 A:{红,黄,蓝,白}. 构建这个信源的信源空间的关键在于测定出现各种不同颜色的概率. 在这个问题中,我们可把各种不同颜色的彩球的出现频率,近似地看作其出现的概率. 假如,箱子中共有 32 个球,其中:红球 16 个;黄球 8 个;蓝球和白球各 4 个. 则可得各种彩球出现频率,即各种彩球出现的先验概率分别为

$$\text{出现红球的概率} \quad P(\text{红}) = \frac{16}{32} = \frac{1}{2};$$
$$\text{出现黄球的概率} \quad P(\text{黄}) = \frac{8}{32} = \frac{1}{4};$$
$$\text{出现蓝球的概率} \quad P(\text{蓝}) = \frac{4}{32} = \frac{1}{8};$$
$$\text{出现白球的概率} \quad P(\text{白}) = \frac{4}{32} = \frac{1}{8}.$$

若用随机变量 X 表示这个信源,其信源空间为

$$[X \cdot P]:\begin{cases} X: & \text{红} & \text{黄} & \text{蓝} & \text{白} \\ P(X): & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \end{cases}$$

用一个离散随机变量 X 代表一个单符号离散信源,这就是我们用数学描述信源的基本出发点. 随机变量 X 的状态空间和概率空间,是构建信源空间$[X \cdot P]$的两个基本要素,而概率空间是决定性要素. 概率可测是香农信息论的基本前提.

第二节　信源符号的自信量

由信息的定义可知,在通信过程中,收信者所获取的信息量,在数量上等于通信前后不确定性的消除. 例如,一位教授作学术报告,由于扩音设备不好,在报告的声音中夹杂悦耳的音乐声. 学生听到的是报告声音的一种变型. 听了这场报告后,虽然知道了报告的大致内容,但某些具体细节没有听清,仍然存在一定的不确定性. 学生从这场报告中所获取的信息量,应该是听报告前对报告内容的不确定性,减去听报告后仍然存在的不确定性所得之差,即不确定性的消除.

若信源发某符号 a_i,由于信道中噪声的随机干扰,收信者收到的是 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,则有

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

$$-[\text{收到 } b_j \text{ 后，收信者对信源发 } a_i \text{ 仍然存在的不确定性}]$$

$$=\text{收信者收到 } b_j \text{ 前、后，对信源发 } a_i \text{ 的不确定性的消除} \tag{1.3}$$

当信道中没有噪声的随机干扰(无噪信道)时，信源发出的符号 a_i 可以不受任何干扰传递给收信者，收信者收到的 b_j 就是 a_i 本身. 由于收信者确切无误地收到了信源发出的符号 a_i，当然就完全消除了对信源发符号 a_i 的不确定性，即

$$[\text{收到 } b_j \text{ 后，收信者对信源发 } a_i \text{ 仍然存在的不确定性}]=0 \tag{1.4}$$

这时，(1.3)就可改写为

$$I(a_i;a_i)=[\text{收到 } a_i \text{ 前，收信者对信源发 } a_i \text{ 的不确定性}] \tag{1.5}$$

式(1.5)中的 $I(a_i;a_i)$ 表示收到 a_i 后，收信者从 a_i 中获取关于 a_i 的信息量. 当然，这就是信源符号 a_i 所含有的全部信息量. 我们把 $I(a_i;a_i)$ 称为信源符号 a_i 的自信息量，并用 $I(a_i)$ 表示. 由(1.5)式即可有

$$I(a_i)=[\text{收到 } a_i \text{ 前，收信者对信源发 } a_i \text{ 的不确定性}] \tag{1.6}$$

(1.6)式表明，信源符号 a_i 的自信息量 $I(a_i)$ 的度量问题，已转变为信源发符号 a_i 的不确定性的度量问题. 我们知道，不确定性是与可能性相联系的，而可能性又可由概率的大小来表示. 可以断言，自信息量 $I(a_i)$ 一定是信源发符号 a_i 的先验概率 $p(a_i)$ 的某一函数，即

$$I(a_i)=f[p(a_i)] \qquad (i=1,2,\cdots,r) \tag{1.7}$$

从客观事实和人们的习惯概念出发，函数 $I(a_i)=f[p(a_i)](i=1,2,\cdots,r)$ 必须满足以下四个公理性条件：

(1)若有两条消息：一条是“中国男子足球队获取世界杯冠军”. 根据中国男子足球队的历来表现，获取世界杯赛的冠军的概率很小，即可能性很小. 从习惯概念出发，认为“中国男子足球队获取世界杯冠军”这条消息的不确定性很大. 这一事件一旦发生，人们就会获取很大的信息量，甚至会出现震动全国、万众欢呼的动人场面；另一条消息是“中国女子乒乓球队获取亚运会冠军”. 根据中国女子乒乓球队的历来表现，中国女子乒乓球队获取亚运会冠军的概率很大，即可能性很大. 从习惯概念出发，认为从“中国女子乒乓球队获取亚运会冠军”这条消息中获取的信息量要小得多.

这个大家都承认的公理，可以这样来表述. 如信源符号 a_i 和 a_j 的先验概率分别为 $p(a_i)$ 和 $p(a_j)$，且 $0<p(a_i),p(a_j)<1$. 若 $p(a_i)>p(a_j)$，则

$$I(a_i)=f[p(a_i)]<I(a_j)=f[p(a_j)] \tag{1.8}$$

即函数 $I(a_i)=f[p(a_i)]$ 是先验概率 $p(a_i)$ 的单调递减函数.

(2)众所周知，“太阳从西边升起”的概率等于零，是不可能事件. 从习惯概念出发，“太阳从西边升起”这条消息的不确定性应是无穷大. 这个事件一旦发生，将会天翻地覆，人们获取无穷大的信息量. 这就是说，如信源符号 a_i 的先验概率 $p(a_i)=0$，则

$$I(a_i)=f[p(a_i)]\longrightarrow\infty \tag{1.9}$$

(3)人们同样公认，“太阳从东边升起”的概率等于 1，是确定事件. 从习惯概念出发，“太阳从东边升起”这条消息不存在任何不确定性. 如果你把这条消息告诉别人，凡听到这条消息的人都会认为你讲的是废话，不会得到任何信息量. 这就是说，如信源符号 a_i 的先验概率 $p(a_i)=1$，则

$$I(a_i)=f[p(a_i)]=0 \tag{1.10}$$

(4)人们的习惯概念认为，两个统计独立事件的联合信息量，应等于它们各自信息量之和. 比如，一般可把“天安门广场有人在照像”与“中国科学院研究生院上信息论课”看作为相互统计独立、互不相关的两件事. 若我们同时得知“天安门广场有人照像”和“中国科学院研究生院上信息论课”这两条消息，从这两条消息中得到的联合信息量，应等于这两条消息各自信息量之和.

这个公理条件可这样来表述. 设有两个统计独立的信源 X 和 Y. 信源 X 的符号 a_i 的先验概率为 $p(a_i)$;信源 Y 的符号 b_j 的先验概率为 $p(b_j)$. 符号 a_i 和 b_j 组成的联合消息 (a_ib_j) 的先验概率是联合概率 $p(a_ib_j)$. 则

$$\begin{aligned} I(a_ib_j) &= f[p(a_ib_j)] \\ &= I(a_i) + I(b_j) \end{aligned} \tag{1.11}$$

从数学上可以证明,满足(1.8)、(1.9)、(1.10)、(1.11)四个公理条件的函数 $I(a_i)$,是符号 a_i 的先验概念 $p(a_i)$ 的倒数的对数,即

$$\begin{aligned} I(a_i) &= \log \frac{1}{p(a_i)} \\ &= -\log p(a_i) \quad (i = 1,2,\cdots,r) \end{aligned} \tag{1.12}$$

很容易证明,(1.12)式满足公理条件(1.8)、(1.9)、(1.10). 同样,因为信源 X 和 Y 统计独立,所以

$$p(a_ib_j) = p(a_i)p(b_j)$$

则由(1.12)式可得

$$\begin{aligned} I(a_ib_j) &= \log \frac{1}{p(a_ib_j)} = \log \frac{1}{p(a_i)p(b_j)} \\ &= \log \frac{1}{p(a_i)} + \log \frac{1}{p(b_j)} = I(a_i) + I(b_j) \end{aligned}$$

这说明,(1.12)式亦满足公理条件(1.11).

由(1.12)式表达的函数形式,满足公理条件(1.8)、(1.9)、(1.10)、(1.11),它可作为信源

$$[X \cdot P]: \begin{cases} X: & a_1 & a_2 & \cdots & a_r \\ P(X): & p(a_1) & p(a_2) & \cdots & p(a_r) \end{cases}$$

发符号 $a_i(i=1,2,\cdots,r)$ 的不确定性,即符号 $a_i(i=1,2,\cdots,r)$ 含有的自信息量的度量函数. 我们称(1.12)式为"信息函数". 由(1.12)式可知,信源发符号 $a_i(i=1,2,\cdots,r)$ 的不确定性,即符号 $a_i(i=1,2,\cdots,r)$ 的自信息量 $I(a_i)(i=1,2,\cdots,r)$ 由符号 $a_i(i=1,2,\cdots,r)$ 的先验概率 $p(a_i)(i=1,2,\cdots,r)$ 唯一确定. 从这个意义上说,我们又把由(1.12)式度量的信息称为"概率信息".

由(1.12)式可知,自信息量 $I(a_i)$ 的单位取决于对数的"底". 若以"2"为底,则所得自信息量的单位为"比特"(bit—binary unit),即

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

若以"e"为底,则所得自信息量的单位为"奈特"(nat—nature unit),即

$$\begin{aligned} I(a_i) &= \log_e \frac{1}{p(a_i)} \\ &= \ln \frac{1}{p(a_i)} \quad \text{奈特} \end{aligned}$$

若以"10"为底,则所得自信息量的单位为"哈特(Hart—Hartley),即

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

若以正整数"r"为底,则所得自信息量的单位为"r 进制信息单位",即

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

根据对数的“换底公式”，不同信息单位之间可进行换算. 本书在后续章节中，如不加说明，一般采用以“2”为底的对数. 为了书写方便起见，把底数“2”略去不写，以“log”代表“$\log_2$”.

信息函数(1.12)式的导出，在人类历史上第一次解决了信息的度量问题，是信息理论发展史上的里程碑，是信息论逐步发展成为一门成熟的学科的基石.

【例 1.1】 假设一次掷两个各自均匀、互相不可区分又互不相关的骰子. 如事件(A)、(B)、(C)分别表示：(A)仅有一个骰子是 3；(B)至少有一个骰子是 4；(C)骰子上点数的总和是偶数. 试计算事件(A)、(B)、(C)发生后所提供的信息量.

解 两个骰子朝上一面点数的组合总数

$$N = 6 \times 6 = 36$$

(A)事件的样本数　　$n_A = 2 \times 5 = 10$

(B)事件的样本数　　$n_B = 5 \times 2 + 1 = 11$

(C)事件的样本数　　$n_C = 6 \times 3 = 18$

我们用随机事件出现频率近似地看作随机事件出现的概率，则事件(A)、(B)、(C)出现的概率分别是：

$$P(A) = \frac{n_A}{N} = \frac{10}{36} = \frac{5}{18};$$

$$P(B) = \frac{n_B}{N} = \frac{11}{36};$$

$$P(C) = \frac{n_C}{N} = \frac{18}{36} = \frac{1}{2}.$$

由(1.12)式得随机事件(A)、(B)、(C)出现后提供的信息量分别是：

$$I(A) = \log \frac{1}{P(A)} = 1.8480 \quad \text{比特}$$

$$I(B) = \log \frac{1}{P(B)} = 1.7105 \quad \text{比特}$$

$$I(C) = \log \frac{1}{P(C)} = 1 \quad \text{比特}$$

由 $I(C)$我们看到，1 比特信息量就是两个互不相容的等概事件之一发生时所提供的信息量，有时亦称为“是否信息”. 例如，你的家人根据你历来的习惯，估计你“中午回家吃饭”与“中午不回家吃饭”的可能性相同. 若某天由于某种原因，你回家吃中午饭，这样，你就给你家人带去了 1 比特信息量.

【例 1.2】 设有 n 个球，每个球都能以同样的概率$\frac{1}{N}$落到 N 个格子的每一个格子中，其中 $N \geqslant n$. 假定(A)表示某指定的 n 个格子中各落入一个球；(B)表示任何 n 个格子中各落入一个球. 试计算事件(A)、(B)发生后各自提供的信息量.

解 由于每个球可落入 N 个格子中的任一个，所以 n 个球在 N 个格子中的分布相当于从 N 个元素中选取 n 个进行有重复的排列，总共有 N^n 种可能分布.

(A)事件的样本点数等于 n 个球在指定的 n 个格子中的全排列数 $n!$. 若把随机事件出现的频率近似地看作随机事件出现的概率，则(A)的概率为

$$P(A) = \frac{n!}{N^n}$$

由(1.12)式得(A)发生后提供的信息量为

$$I(A) = \log \frac{1}{P(A)}$$

$$=\log\frac{N^n}{n!}$$
$$=n\log N-\log n! \quad \text{比特}$$

对于随机事件(B)来说，由于可以从 N 个格子中任选 n 个格子，这种选法共有 C_N^n 种. 对于每种选定的 n 个格子，样本数与(A)的样本数相同. 所以，随以事件(B)的总样本数为 $C_N^n \cdot n!$. 同样，如把随机事件的出现频率近似地当作随机事件出现的概率，则(B)的概率为

$$P(B)=\frac{C_N^n \cdot n!}{N^n}$$
$$=\frac{N!}{N^n(N-n)!}$$

由(1.12)式得(B)发生后提供的信息量为

$$I(B)=\log\frac{1}{P(B)}=\log\frac{N^n(N-n)!}{N!}$$
$$=n\log N+\log(N-n)!-\log N! \quad \text{比特}$$

显然，由于 $P(B)>P(A)$，所以随机事件(B)出现后提供的信息量 $I(B)$，小于随机事件(A)提供的信息量 $I(A)$.

第三节　信源的信息熵

虽然，信息函数 $I(a_i)$破天荒地使信息度量成为可能，是信息度量的有力工具，但在信息度量方面仍然存在某些不足. 首先，信源发符号 a_i 不是确定事件，是以 $p(a_i)$为概率的随机事件，相应的自信息量 $I(a_i)$也是一个以 $p(a_i)$为概率的随机性的量. 显然，用一个随机性的量来度量信息是不方便的. 其次，信息函数 $I(a_i)$只能表示信源发某一特定的具体符号 a_i 所提供的信息量. 不同的符号，有不同的自信息量. 所以它不足以作为整个信源的总体信息测度. 据此，在信息函数 $I(a_i)$的基础上，构架一个确定的量，作为信源的总体信息测度，就成为我们面临的一个重要课题.

显然，能作为信源总体信息测度的确定的量，应是信源 X 可能发出的各种不同符号 $a_i(i=1,2,\cdots,r)$含有的自信息量 $I(a_i)(i=1,2,\cdots,r)$，在信源的概率空间$\{p(a_1),p(a_2),\cdots,p(a_r)\}$中的统计平均值. 为了指明是信源 X 的信息测度，我们把这个统计平均值记为 $H(X)$，即有

$$\begin{aligned}H(X)&=p(a_1)I(a_1)+p(a_2)I(a_2)+\cdots+p(a_r)I(a_r)\\&=-p(a_1)\log p(a_1)-p(a_2)\log p(a_2)-\cdots-p(a_r)\log p(a_r)\\&=-\sum_{i=1}^{r}p(a_i)\log p(a_i) \quad \text{比特 / 信源符号}\end{aligned} \tag{1.13}$$

我们称 $H(X)$是信源 X 的“信息熵”. 它表示信源 X 每发一个符号(不论发什么符号)所提供的平均信息量. 信息熵 $H(X)$的单位取决于(1.13)式中对数的底. 若以正整数 r 为底，则有

$$H_r(X)=-\sum_{i=1}^{r}p(a_i)\log_r p(a_i) \quad r\text{ 进制信息单位 / 信源符号} \tag{1.14}$$

由对数换底公式，可得 r 进制信息熵 $H_r(X)$与 2 进制信息熵 $H(X)$之间的换算关系

$$\begin{aligned}H_r(X)&=-\sum_{i=1}^{r}p(a_i)\log_r p(a_i)\\&=\frac{-\sum_{i=1}^{r}p(a_i)\log p(a_i)}{\log r}=\frac{H(X)}{\log r}\end{aligned} \tag{1.15}$$

我们不妨通过一个简单例子，从另一角度进一步领会信息熵的含义. 若有一布袋内放 100 个球. 其中：70 个是红色；30 个是白色. 现随意摸出一球，猜是什么颜色. 这个随机试验相当于一个单符号离散信源，其信源空间为

$$[X \cdot P]: \begin{cases} X: & \text{红}(a_1) & \text{白}(a_2) \\ P(X): & 0.7 & 0.3 \end{cases}$$

摸出红球(a_1)的信息量

$$I(a_1) = -\log 0.7 \quad \text{比特}$$

摸出白球(a_2)的信息量

$$I(a_2) = -\log 0.3 \quad \text{比特}$$

若每次摸出一个球后又放回袋中，再进行第二次摸取. 在摸取 N(N 足够大)次中，红球(a_1)出现的次数约为 $n_1 = Np(a_1)$次，白球出现的次数约为 $n_2 = Np(a_2)$次. 摸取 N 次后总共所获取的信息量为

$$\begin{aligned} I_N &= n_1 I(a_1) + n_2 I(a_2) \\ &= Np(a_1)I(a_1) + Np(a_2)I(a_2) \end{aligned}$$

平均每摸取一次所获得的平均信息量为

$$\begin{aligned} I &= \frac{I_N}{N} = \frac{Np(a_1)I(a_1) + Np(a_2)I(a_2)}{N} \\ &= p(a_1)I(a_1) + p(a_2)I(a_2) = -p(a_1)\log p(a_1) - p(a_2)\log p(a_2) \\ &= -\sum_{i=1}^{2} p(a_i)\log p(a_i) \qquad \text{比特 / 信源符号} \end{aligned} \tag{1.16}$$

(1.16)式正好就是由(1.13)式构架定义的信息熵. 这再次说明，由(1.13)式定义的信息熵 $H(X)$，确实表示信源 X 每发一个符号所能提供的平均信息量，是在平均的意义上的信源总体信息测度.

信息函数 $I(a_i)$既表示收信者确切无误收到信源符号 a_i 后，从 a_i 中获取的信息量，同时也表示收到符号 a_i 前，收信者对于信源发符号 a_i 存在的不确定性. 作为信息函数 $I(a_i)$在信源概率空间中的统计平均值的信息熵 $H(X)$，除了表示信源 X 每发一个符号能提供的平均信息量之外，同样也表示收信者在接到符号前，对信源 X 存在的平均不确定程度. 例如，有三个信源 X_1、X_2 和 X_3，它们的信源空间分别是：

$$[X_1 \cdot P]: \begin{cases} X_1: & a_1 & a_2 \\ P(X_1): & 0.5 & 0.5 \end{cases}$$

$$[X_2 \cdot P]: \begin{cases} X_2: & a_1 & a_2 \\ P(X_2): & 0.7 & 0.3 \end{cases}$$

$$[X_3 \cdot P]: \begin{cases} X_3: & a_1 & a_2 \\ P(X_3): & 0.99 & 0.01 \end{cases}$$

我们把这三个信源作一番比较. 对于信源 X_1 来说，发符号 a_1 和发符号 a_2 的概率相等，即发 a_1 和发 a_2 的可能性相同. 对收信者来说，要判断信源 X_1 是发 a_1 还是发 a_2，是最捉摸不定的. 对于信源 X_2 来说，由于发 a_1 的概率 0.7 大于发 a_2 的概率 0.3，即发 a_1 的可能性大于发 a_2 的可能性. 对于收信者来说，要判断信源 X_2 发 a_1 还是发 a_2，要比信源 X_1 容易一些. 对于信源 X_3 来说，由于发 a_1 的概率 0.99 远大于发 a_2 的概率 0.01，即发 a_1 的可能性远大于发 a_2 的可能性. 对于收信者来说，要判断信源 X_3 发 a_1 还是发 a_2，是这三个信源中最容易的. 这就意味着，在发符号前，要判断信源 X_1 是发 a_1 还是发 a_2 的不确定性最大，要判断信源 X_3 是发 a_1 还是 a_2 的不确定性最小. 而信源 X_1、X_2 和 X_3