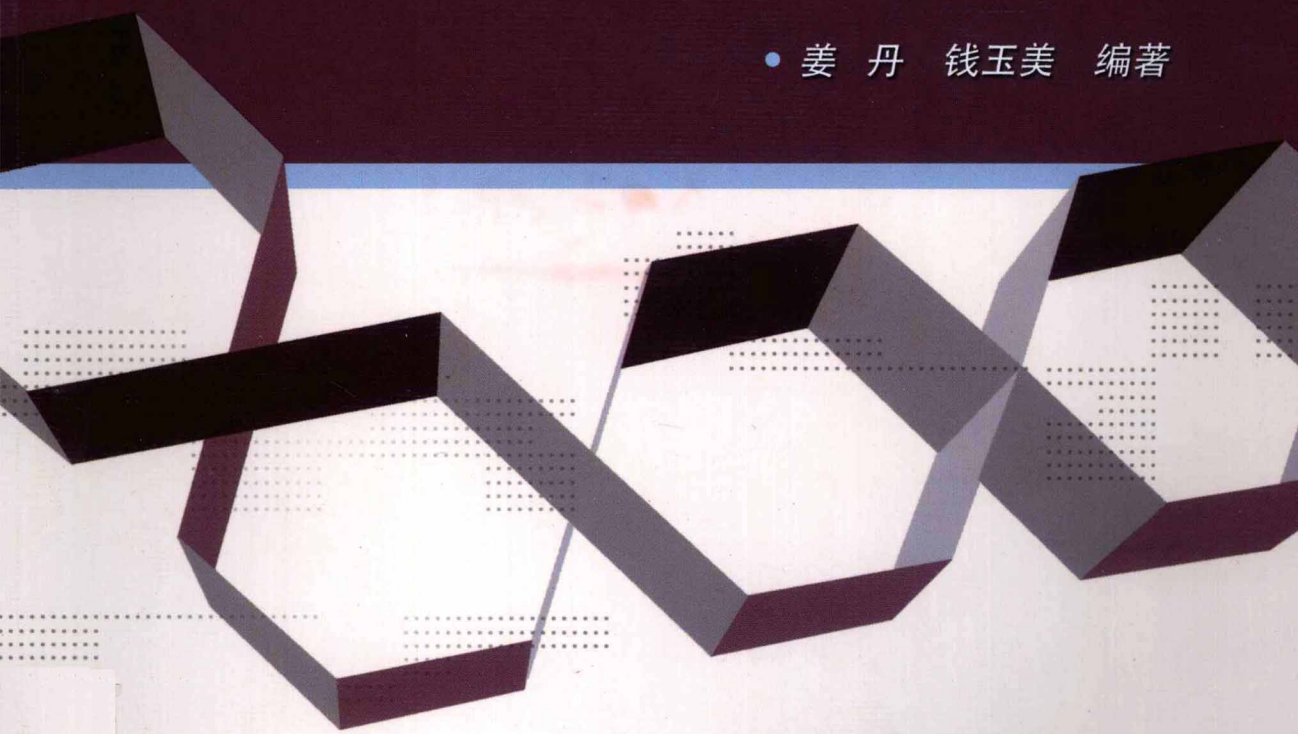
普通高等教育“十二五”规划教材

电子信息类精品教材

信息论与编码基础

Fundamentals of Information Theory and Coding

姜 丹 钱玉美 编著



電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY http://www.phei.com.cn

普通高等教育“十二五”规划教材
电子信息类精品教材

信息论与编码基础

姜 丹 钱玉美 编著

電子工業出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书系统论述香农(shannon)信息论的基础理论和编码的基本理论及方法。内容包括:信息的定义、自信息、互信息、信息熵、平均互信息、信道容量与匹配信源、串接信道与数据处理;扩展信源的信息熵、平均符号熵、马尔柯夫(Markov)信源的极限熵、剩余度、扩展信道的平均互信息、独立并列信道的信道容量;连续信源的相对熵、熵功率、高斯白噪声加性信道的最高信息传输速率;单义可译码的结构定理、信源符号速率极限定理、霍夫曼(Huffman)码编码方法及其性能评估、费诺(Fano)码和香农(shannon)码的编码方法;最小平均误码率译码规则、几种纠错码的编码方法及其最小误码率、误码率极限定理、线性分组码的代数结构和编码译码方法、系统完备码的最小平均误码率、汉明(Hamming)码的最优化;信息率 - 失真函数 $R(D)$ 的定义和性质、离散信源 $R(D)$ 的表达式、扩展信源的 $R(D)$ 与数据压缩的关系等。

本书可作为高等院校高年级本科生的教材,也可供相关专业的研究生和从事信息理论、信息技术的科研、教学和工程技术人员参考。

图书在版编目(CIP)数据

信息论与编码基础/姜丹,钱玉美编著. —北京 :电子工业出版社, 2013.2
电子信息类精品教材
ISBN 978-7-121-17490-2

Ⅰ. ① 信… Ⅱ. ① 姜… ② 钱… Ⅲ. ① 信息论 - 高等学校 - 教材 ② 信源编码 - 高等学校 - 教材
Ⅳ. ① TN911.2

中国版本图书馆 CIP 数据核字(2012)第 143255 号

责任编辑:韩同平 特约编辑:林宏峥
印 刷:涿州市京南印刷厂
装 订:涿州市京南印刷厂
出版发行:电子工业出版社
北京市海淀区万寿路 173 信箱 邮编 100036
开 本:787×1 092 1/16 印张:24 字数:680 千字
印 次:2013 年 2 月第 1 次印刷
印 数:2 000 册 定价:59.90 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zlts@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

前　言

信息论产生于通信领域，是解决通信问题的有力工具。经半个多世纪的发展，它已成为整个信息科学中最完整、最系统、最成熟的一门学科，是通信和信息处理有关专业的基础学科。由于解决问题的思路和方法的独特、新颖和有效，在当今信息时代，它与其他自然科学，甚至社会、人文科学的有关学科，相互渗透、密切结合，显示出它的勃勃生机和不可估量的发展前景。随着时代的发展和文明程度的不断提高，高等院校和科研院所，普遍开设本科生、研究生的信息论有关课程。

作者于2001年编著，由中国科学技术大学出版社出版发行的《信息论与编码》一书，在2003年入选中华人民共和国教育部研究生工作办公室推荐的“研究生教育用书”，并分别于2004年、2009年由中国科学技术大学出版社出版发行第二版、第三版。

本书是作者总结从事30余年高等院校和科研院所本科生、研究生信息论课程的教学经验，面向高等院校高年级本科生编写的信息论及其相关课程的教材。

“着眼基础，强化基础”是本书的宗旨。

本书重点介绍和论述香农(Claude E. Shannon)信息论和编码的基础理论和方法。

在“引言”中，首先提出信息论的“三大理论支柱”，作为一条主线，贯穿本书始终。它也是交给读者进入信息论学术殿堂大门，必备的一把“钥匙”。

本书内容共设7章，分为两大部分。

第一部分(第1、2、3、4章)论述信息论的基础理论。为了强化“系统”总体概念，由“单符号离散通信系统”(第1、2章)、“多符号离散通信系统”(第3章)、“连续通信系统”(第4章)三个“横向”教学板块，构成“横向”教学结构体系。在教学进程中，以“系统”为一个整体，由简到繁、由浅入深、循序渐进、逐步提升。

第二部分(第5、6、7章)论述编码的基础理论和方法。重点论述“无失真信源编码定理”、“抗干扰信道编码定理”，以及信息率—失真函数的基础理论。侧重介绍“霍夫曼(Huffman)码”、“线性分组码”的编码理论和方法，阐明信息率—失真函数与“限失真信源编码”(数据压缩)之间的关系。列举大量例题，说明采用编码手段，实施通信系统“最优化”的具体途径和方法，展示通信系统“最优化”的光明前景。

本书各章、节以及下属标题设置，体现信息论基础理论体系的结构框架。在论述中，用定理的形式，表述信息论的基本概念、重要结论和编码的基本理论。用深入细致、逻辑严密、步步推进的数学分析过程，推导、论证每一个定理。体现信息论基础理论的完整演绎体系。

为了帮助读者排除学习信息论过程中经常遇到的数学分析方面的困难，结合有关内容，适当介绍必要的数学基础知识，提供不同的证明途径和方法。同时，用通俗易懂、富有哲理的比喻，诠释定理的物理意义和内涵，阐明香农信息论向人们揭示的信息产生、信息传输、信息处理的规律，彰显香农信息论的信息观。

本书各章、节基本上覆盖了香农信息论基础理论的主要内容。在教学过程中，可根据限定学时数和其他相关实际情况，从中挑选适当章、节作为课堂教学内容。有些内容，可供教师备课参考，或供学生自修阅读。各章所附“习题”，难易程度不一，亦可适当选择使用。

本书除供高等院校高年级本科生作为教材外，也可供有关专业的研究生和从事信息理论、信息技术、通信工程的教学、科研、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

于北京

目　录

引　言

信息论是人们在长期通信实践活动中，将通信技术与概率论、随机过程、数理统计等学科相结合，逐步发展起来的一门新兴交叉学科。

美国科学家香农(Claude E. Shannon)于1948年发表的著名论文《通信的数学理论》(Claude E. Shannon, A Mathematical Theory of Communication)，奠定了信息论的理论基础。信息论以通信活动为背景，以通信的有效性、可靠性为主要研究对象，用富有创意的思路和方法，分析探究实现既有效、又可靠通信的途径和方法，论证通过编码实现最优化通信的可能性。它是解决通信问题的理论基础和有力工具。经过半个多世纪的发展、完善和提高，它已成为信息科学中最完善、最系统、最成熟的基础理论学科。

若某同学收到两封来信：一封是谈关于同学们最近的工作、学习情况的；另一封是谈关于家人的健康情况的。现若要问：他从哪一封信中获取了更多的信息？也许，按某种想当然的感觉，他会给出某种模糊的回答，如“从家信中获取了更多的信息”。这个结论可靠吗？就算这个结论不错，如进一步问：“从家信中获取的信息，比从同学来信中获取的信息多了多少？”一般来说，人们很难回答这个问题。问题的症结就在于，人们通常是经验性地、习惯性地把“信息”和“消息”不加区分地混为一谈。

那么，为什么把“信息”和“消息”不加区分地混为一谈，信息的度量就会显得十分困难呢？众所周知，“消息”是用文字、符号、数据、语言、图片、音符、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。“消息”由“形式”、“语义”和“语用”三个因素组成。不同的消息，不仅有不同的“形式”，而且含有不同的“语义”和不同的“语用”效果。例如，“北京获得二〇〇八年第二十九届奥运会主办权”这条消息，它的“形式”，可看作是汉字表中挑选20个字的一种选择，是20个汉字的一个时间序列。在“语义”上，这条消息含有多个“语义”层次结构：是中国的“北京”，而不是法国的“巴黎”、加拿大的“多伦多”、日本的“大坂”、土耳其的“伊斯坦布尔”等其他国家的城市；是“获得”，而不是“丢失”；是“二〇〇八年”，而不是“二〇〇〇”、“二〇〇四”和“二〇一二年”等其他时间；是“第二十九届”，而不是“第二十八届”、“第三十届”等其他届数；是“奥运会”，而不是“世界杯”、“世界锦标赛”等其他赛事；是“主办权”，而不是“参赛权”、“电视转播权”等其他权利。从“语用”效果上来看，它不仅与消息的“语义”内容有关，而且与消息接收者的主观因素有关。从电视实况转播中看到，当萨马兰奇主席宣布：“第二十九届奥运会的主办城市是北京”这一消息时，在场的中国代表团成员情不自禁地跳跃欢呼、互相拥抱、热泪盈眶。而在场的其他国家代表团，绝大多数仍然坐在自己的座位上，鼓掌表示祝贺。也有少数代表团感到失落，甚至沮丧。你看，同一条消息，对不同接收者来说，做出的反应的反差如此巨大！由此可见，一条“消息”既有“形式”，又有“语义”，还有“语用”效果。而且“消息”的“形式”、“语义”和“语用”这三个因素是捆绑在一起的。“消息”是其“形式”、“语义”和“语用”三因素互相交织在一起的一个混合体。

要解决信息的度量问题，必然要用数学工具，进行量的运算。数学是刻画物质运动形式的工具，用数学对“消息”的“形式”进行描述，不存在法则上的困难。但用数学刻画“消息”的“语义”，乃至“语用”效果，至今仍然是一个巨大的难题。所以，如经验性、习惯性地把“消息”和“信息”不加区分地混为一谈，把“消息”的“形式”、“语义”和“语用”三因素捆绑在一起，综合地解决信息的度量问题，必然面临头绪纷繁、无从下手的僵局。

信息论的奠基人——香农，针对通信活动的特点，精辟地提出了“形式化假说”、“非决定论”和“不确定性”三个观点，明确了通信领域中“信息”的特定含义，打破了僵局，解决了信息的度量问题，开创了信息理论的新局面。

1. 形式化假说

通过对通信活动功能的观察分析，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选择的消息”。这就是说，通信工程的职责，只是在接收端把发送端发出的消息，从形式上精确地或近似地复制出来。通信工程并不需要对复制出来的消息的“语义”做任何处理和判断。对消息的“语义”内容的解读、处理和判断，是接收者自己的事，不是通信工程本身的任务，与通信工程无关。正如邮递员一样，邮递员的职责，只是把信尽量完好无损地送到收信者手中。至于对信的内容的解读，乃至看了信以后收信者是高兴还是悲痛，那完全是收信者自己的事，与邮递员毫无关系。又如电视实况转播一场精彩的 NBA 篮球比赛，电视转播系统的职责，只是把画面和声音尽量精准地传播到每家的电视接收机的屏幕上。至于对这场球赛中运动员的表现、裁判的裁决、教练员的指挥等因素的解读和评价，都是看电视节目的观众自己的事，与电视转播系统无关。当然，看完球赛后观众的情绪是高兴还是沮丧，是热烈欢呼还是气得把电视机从楼上扔下去，更是观众自己的事，与电视转播系统毫无关系。这就是香农对通信活动的“形式化假说”。

对于通信工程的“形式化假说”，大胆地去掉了消息的“语义”和“语用”因素，巧妙地保留了能用数学描述的“形式”这一因素。这使应用数学工具，定量描述信息成为了可能，打开了信息理论进入科学殿堂的大门。

2. 非决定论

通过对通信活动对象的分析研究，香农指出，“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。”“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。”“各种消息的选择都是随机的，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。例如，面对公众的“公用电话”，什么人、什么时候使用公用电话，以及通话人声音的最高频率、频带宽度、峰值功率、平均功率、持续时间等技术参数都是随机的，工程设计人员是无法事先预料的。显然，为了使公用电话尽可能地被广泛使用，面对广大通话者，工程设计人员不可能把某一特定的通话者的技术参数作为设计的依据，而是要用概率论、随机过程和数理统计等数学工具，从大量的不可预料的通话者发出的随机消息中，寻求其统计规律，作为工程设计的依据。用随机的观点和概率统计的方法来观察、处理信息，这就是香农的“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。它从原则上回答了应采用什么类型的数学工具来解决信息的度量问题。

3. 不确定性

通过对通信活动机制的剖析研究，香农指出，人们只在两种情况下有通信的需求：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”（或“不完全知道”）这个消息；其二，是自己有某种“疑问”要询问对方，而估计对方能做出“解答”（或“部分”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，不能做出明确的判断，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“部分知道”），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（或部分消除）了通信前存在的“不确定性”。通信的作用，就是通过消息的传递，接收者从收到的消息中，获取了一样“东西”，用这个

“东西”，消除（或部分消除）了通信前存在的“不确定性”。这种“东西”，就是“信息”。这样，在通信领域中，可以给“信息”下一个明确的定义：“信息”就是用来消除不确定性的东西，在数量上就等于通信前、后“不确定性”的消除量。这就是香农从“不确定性”观点出发，给通信领域中的“信息”下的明确的定义。

我们知道，“可能性”的大小，在数学上可以用概率的大小来表示：概率大，表示出现的“可能性”大；概率小，表示出现的“可能性”小。我们同样知道，“不确定性”与“可能性”是有联系的：“可能性”大，意味着“不确定性”小；“可能性”小，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2012年奥运会冠军”这条消息，根据中国女子乒乓球队历来的表现，夺取奥运会冠军的概率很大，即“可能性”很大，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，夺取世界杯冠军的概率很小，即“可能性”很小，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。球迷们会惊喜万分，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。显然，通信前、后“不确定性”的消除量——信息量，同样应是概率的某一函数。对于随机消息来说，我们虽然不能确定它能否发生，但表示随机消息发生的“可能性”大小的“概率”，是一个精准的数量。那么，只要找到这个函数，通信前、后获得的信息就可予以度量，而且一定是一个精准的数量。所以，香农对通信领域中的“信息”的定义，从理论原则上解决了信息的度量问题。

在通信领域中，“信息”与“消息”两者之间既有联系，又有区别，两者不能混为一谈。“消息”是表达“信息”的形式，是载荷“信息”的客体；“信息”是“消息”的统计特性的函数，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”。它是一门具有完整数学演绎体系的高度抽象和概括的基础理论学科。

“形式化假说”、“非决定论”和“不确定性”是构建香农信息论这个理论大厦的三个理论支柱，也是我们打开信息论这门学科的大门，并达到其“顶峰”的一把必备的“钥匙”。

信息论是通信领域中的一门学科，是信息科学的重要组成部分。以它独到、新颖的观念和方法，使它与其他学科相互渗透、相互结合，取得了令人惊喜的成果。随着科学技术的迅猛发展和人类文明的不断提高，必将显示出它的勃勃生机和光明前景！

只要坚持不懈地在信息论科学殿堂里潜心学习、勤奋耕耘、努力探索、勇于创新，必将达到科学与艺术融会贯通的美妙境界，享受到人类文明的无穷乐趣！

第 1 章　单符号离散信源

信源就是信息的源泉。信息不是消息本身，但又包含在消息之中。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，则这种信源称为单符号离散信源。单符号离散信源是最简单、最基本的信源。

1.1　信源的信息熵

建立单符号离散信源的数学模型、构建信源符号含有信息量的度量函数、确立信源的总体信息测度，是单符号离散信源的首要课题。

1.1.1　信源的数学模型

单符号离散信源中某符号要含有一定信息，信源发这一符号必须有随机性，以一定的概率发这一符号。单符号离散信源，是具有一定概率分布的离散符号的集合，可用一个离散随机变量 X 来表示：用 X 的状态空间，表示信源可能发出的各种不同符号；用 X 的概率空间，表示信源发出各种不同符号的概率分布。

一般地说，若信源可能发出 r 种不同符号 $a_1, a_2, \cdots, a_r$，相应的先验概率分别是 $p(a_1)$，$p(a_2), \cdots, p(a_r)$。用离散随机变量 X 表示这个信源，构成信源 X 的信源空间

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

$$0 \leqslant p(a_i) \leqslant 1 \qquad (i = 1, 2, \cdots, r)$$

信源 X 发出的符号，只可能是信源符号集合 $X: \{a_1, a_2, \cdots, a_r\}$ 中的某一符号，不可能是集合以外其他任何符号。$[X \cdot P]$ 中的概率空间 $P: \{p(a_1), p(a_2), \cdots, p(a_r)\}$ 应是一个完备集，即有

$$\sum_{i=1}^{r} p(a_i) = 1 \tag{1.2}$$

信源空间 $[X \cdot P]$ 完整地描述了信源 X 的信息特征，是信源 X 的数学模型。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。构建信源空间的关键，是给定（或测定）信源 X 各种不同符号的先验概率 $p(a_i)(i = 1, 2, \cdots, r)$。信源 $X: \{a_1, a_2, \cdots, a_r\}$ 各符号的先验概率 $p(a_i)$ 可知，是香农信息论的假设前提。不可多次重复试验，不存在先验概率的事件，不属于香农信息论的范畴。

1.1.2　信源符号的自信量

信源空间为式（1.1）所示的信源 X 中，每一种不同符号 a_i 含有的信息量 $I(a_i)(i = 1, 2, \cdots, r)$ 称为符号 a_i 的“自信量”。

信源 X 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（如图 1.1 所示）。根据信息的定义，收信者收到 b_j 后，从 b_j 中获取关于 a_i 的信息量

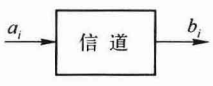


图　1.1

$$I(a_i;b_j)=\{\text{收到 } b_j \text{ 前,收信者对信源发 } a_i \text{ 的不确定性}\}-\{\text{收到 } b_j \text{ 后,收信者对信源发 } a_i \text{ 仍然存在的不确定性}\}=\{\text{收信者收到 } b_j \text{ 前、后,对信源发 } a_i \text{ 的不确定性的消除}\} \tag{1.3}$$

为了导出 a_i 的自信量 $I(a_i)$,可先把问题推向极致。设信道中没有噪声干扰(无噪信道),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

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

这时,式(1.3)可改写为

$$I(a_i;a_i)=\{\text{收到 } a_i \text{ 前,收信者对 } a_i \text{ 存在的不确定性}\} \tag{1.5}$$

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,即信源符号 a_i 含有的全部信息量

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

式(1.6)表明,信源符号 a_i 的自信量 $I(a_i)$,等同于信源发 a_i 的不确定性。而信源发符号 a_i 的不确定性 $I(a_i)$,一定是信源符号 a_i 的先验概率 $p(a_i)$ 的某一函数

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

至此,找出函数 $f[p(a_i)]$ 的具体表达式,已成为解决信源符号 a_i 所含自信息的度量问题的关键。

函数 $I(a_i)=f[p(a_i)]$ 必须满足以下四个公理条件:

(1) 若有两条消息:"中国男子乒乓球队获得 2012 年奥运会冠军"和"中国男子篮球队获得 2012 年奥运会冠军"。根据历来的表现,"中国男子乒乓球队获得 2012 年奥运会冠军"的概率,比"中国男子篮球队获得 2012 年奥运会冠军"的概率大,"中国男子乒乓球队获得 2012 年奥运会冠军"的不确定性,比"中国男子篮球队获得 2012 年奥运会冠军"的不确定性小。如这两条消息在 2012 年奥运会上真的都实现了,收信者收到"中国男子乒乓球队获得 2012 年奥运会冠军"这条消息后,消除的不确定性,比收到"中国男子篮球队获得 2012 年奥运会冠军"后消除的不确定性小。从"中国男子乒乓球队获得 2012 年奥运会冠军"这条消息中获得的信息量,比从"中国男子篮球队获得 2012 年奥运会冠军"这条消息中获取的信息量小。这种感觉是人们的实践经验和习惯概念所公认的。

一般来说,若信源 $X:\{a_1,a_2,\cdots,a_r\}$ 中符号 a_q 和 a_l 的概率分别为 $p(a_q)$ 和 $p(a_l)$,且有 $p(a_q)>p(a_l)$,则有

$$\{I(a_q)=f[p(a_q)]\}<\{I(a_l)=f[p(a_l)]\} \qquad (q,l=1,2,\cdots,r) \tag{1.8}$$

这表明,函数 $I(a_i)=f[p(a_i)]$ 是 $p(a_i)$ 的单调递减函数。

(2) 人们公认,"太阳从西边升起"是概率为零的不可能事件,存在无限大的不确定性。如这个事件一旦发生,就可消除无限大的不确定性,获取无限大的信息量。到时,将会天翻地覆、乾坤扭转。

一般地说,若信源 $X:\{a_1,a_2,\cdots,a_r\}$ 中符号 a_i 的先验概率 $p(a_i)=0$,则有

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

(3) 人们同样公认,"太阳从东边升起"是概率等于 1 的确定事件,不存在任何不确定性。如有一天,某人告诉你"明天太阳从东边升起"。你听了以后,没有消除任何不确定性,也没有从这条消息中获得任何信息量。

一般地说,若信源 $X:\{a_1,a_2,\cdots,a_r\}$ 中符号 a_i 的先验概率 $p(a_i)=1$,则

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

(4) 若甲与乙两人互不相干、统计独立地同时掷一个六面质地均匀的骰子。甲掷的骰子朝上一面呈现“6”,乙掷的骰子朝上一面呈现“4”。当同时知道这两个结果后,人们获得的信息量,应该等于分别单独从甲、乙二人掷骰子的结果中获得的信息量之和。这也是人们一致公认的。

一般地说,若两个统计独立的信源 X 和 Y 的信源空间分别为

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

$$[Y \cdot P]\begin{cases} Y: & b_1 & b_2 & \cdots & b_s \\ P(Y): & p(b_1) & p(b_2) & \cdots & p(b_s) \end{cases}$$

且有 $a_i(i=1,2,\cdots,r)$ 和 $b_j(j=1,2,\cdots,s)$ 的联合概率

$$p(a_ib_j)=p(a_i)p(b_j) \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s)$$

则 (a_ib_j) 的自信量

$$I(a_ib_j)=I(a_i)+I(b_j) \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{1.11}$$

综上所述,式(1.8)、式(1.9)、式(1.10)、式(1.11)就是函数 $I(a_i)=f[p(a_i)]$ 要满足的四个公理条件。在数学上可以证明,满足这四个公理条件的函数是

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

它是信源空间 $[X \cdot P]$ 中,概率分量为 $p(a_i)$ 的信源符号 a_i 的自信量的度量函数,称之为符号 a_i 的“自信函数”(图 1.2)。自信函数式(1.12)表明,信源 X 的任一符号 a_i 的自信量 $I(a_i)$,由 a_i 的先验概率 $p(a_i)$ 唯一确定。只要知道 $p(a_i)$,就可唯一确定地、精准地得到相应符号 a_i 的自信量 $I(a_i)$。

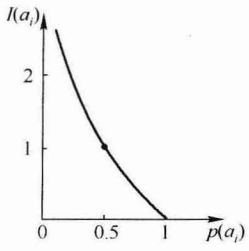


图 1.2

因为信源 X 的任一符号 a_i 的先验概率 $p(a_i)\in[0,1]$,为了确保由自信函数式(1.12)计算所得符号 a_i 的自信量 $I(a_i)\geqslant 0$,在信息论中,人为地规定式(1.12)所示自信函数中的对数的“底”,取大于 1 的数。若以“2”为底,则自信量 $I(a_i)$ 的单位为“比特”(binary unit,缩写为 bit—比特)

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

若以“e”为底,则自信量 $I(a_i)$ 的单位为“奈特”(nature unit,缩写为 nat—奈特)

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

若以“10”为底,则自信量 $I(a_i)$ 的单位为“哈特”(Hartley,缩写为 Hat—哈特)

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

若以大于 1 的正整数 r 为底,则自信量 $I(a_i)$ 的单位为“r 进制信息单位”

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

本书在后续章节中,如不加说明,一般采用以“2”为底的对数,并以“ log ”代表“ $\log_2$ ”。

1.1.3 信源的信息熵

对于式(1.1)所示信源空间 $[X \cdot P]$ 的信源 X 来说,自信量 $I(a_i)$ 表示信源 X 的某一特定符号 a_i 所含有的信息量。信源 X 发符号 a_i,是概率为 $p(a_i)$ 的随机事件,相应的自信量 $I(a_i)$ 也是一个概率为 $p(a_i)$ 的随机量。信源 X 的 r 个随机的自信量 $I(a_i)(i=1,2,\cdots,r)$ 中任何一个,都不

能作为信源 X 的总体信息测度。显然，信源 X 的总体信息测度，应是信源 X 每发一个信源符号（不论是哪一个符号）提供的平均信息量。因为信源空间 $[X \cdot P]$ 中的概率空间 $P:\{p(a_1), p(a_2), \cdots p(a_r)\}$ 是一个完备集，所以这个平均信息量，应是 r 个自信量 $I(a_i)(i=1,2,\cdots,r)$ 在信源 X 的概率空间 $P:\{p(a_1), p(a_2), \cdots p(a_r)\}$ 中的统计平均值

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

式(1.13)所示信源 X 的总体信息测度 $H(X)$，称为信源 X 的“信息熵”。它已不像自信量 $I(a_i)$ 那样，是概率为 $p(a_i)$ 的随机量，而是 $I(a_i)$ 在信源概率空间 $P:\{p(a_1), p(a_2), \cdots, p(a_r)\}$ 中的统计平均值，是一个确定的量；它不再像自信量 $I(a_i)$ 那样，表示信源 X 的某一特定符号 a_i 含有的信息量，而是从总体上、从平均的意义上表示信源 X 每一个符号（不论哪一个符号）所含有的平均信息量（或信源发符号前，每一符号存在的平均不确定性；信源每发一个符号提供的平均信息量；收信者确切无误地收到信源每一符号所获取的平均信息量）。若式(1.13)中的对数的底取“2”，则信息熵的单位是（比特/信源符号）。

【例 1.1】 掷一个六面均匀的骰子，若把骰子朝上一面的点数作为这个随机试验的结果，并把试验的结果作为信源的输出，那么，这个信源就是一个单符号离散信源，可用离散随机变量 X 表示。这个信源 X 可能出现的符号（数字）组成的信源符号集，用 X 的状态空间 $X:\{1,2,3,4,5,6\}$ 表示。信源 X 各种不同符号（数字）出现的先验概率，用 X 的概率空间 $P:\{1/6,1/6,1/6,1/6,1/6,1/6\}$ 表示。信源 X 的信源空间为

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

信源 X 的符号集 $X:\{1,2,3,4,5,6\}$ 中每一符号的先验概率相等，即

$$P\{X=1\} = P\{X=2\} = P\{X=3\} = P\{X=4\} = P\{X=5\} = P\{X=6\} = 1/6$$

则信源 $X:\{1,2,3,4,5,6\}$ 中每一符号的自信量都相等，即

$$I(1) = I(2) = I(3) = I(4) = I(5) = I(6) = \log\frac{1}{1/6} = \log 6 = 2.59 \quad (\text{比特})$$

信源 X 的信息熵

$$\begin{aligned} H(X) &= -\sum_{i=1}^{6} p(a_i)\log p(a_i) = -\sum_{i=1}^{6} P(X=i)\log P\{(X=i)\} \\ &= -\frac{1}{6}\log\frac{1}{6} - \frac{1}{6}\log\frac{1}{6} - \frac{1}{6}\log\frac{1}{6} - \frac{1}{6}\log\frac{1}{6} - \frac{1}{6}\log\frac{1}{6} - \frac{1}{6}\log\frac{1}{6} \\ &= \log 6 = 2.59 \qquad (\text{比特/信源符号}) \end{aligned}$$

若骰子六面不是均匀的，任一面出现的概率与该面的点数成正比，是该面点数的归一化值

$$P\{X=1\} = \frac{1}{1+2+3+4+5+6} = \frac{1}{21} \qquad P\{X=2\} = \frac{2}{1+2+3+4+5+6} = \frac{2}{21}$$

$$P\{X=3\} = \frac{3}{1+2+3+4+5+6} = \frac{3}{21} \qquad P\{X=4\} = \frac{4}{1+2+3+4+5+6} = \frac{4}{21}$$

$$P\{X=5\} = \frac{5}{1+2+3+4+5+6} = \frac{5}{21} \qquad P\{X=6\} = \frac{6}{1+2+3+4+5+6} = \frac{6}{21}$$

则该信源的信源空间为

$$[X \cdot P]\begin{cases} X: & 1 & 2 & 3 & 4 & 5 & 6 \\ P(X): & \frac{1}{21} & \frac{2}{21} & \frac{3}{21} & \frac{4}{21} & \frac{5}{21} & \frac{6}{21} \end{cases}$$

朝上一面出现各点的自信量为：

$$I(1) = -\log\frac{1}{21} = \log 21 = 4.39 \text{（比特）} \qquad I(2) = -\log\frac{2}{21} = \log 10.5 = 3.39 \text{（比特）}$$

$$I(3) = -\log\frac{3}{21} = \log 7 = 2.81 \text{（比特）} \qquad I(4) = -\log\frac{4}{21} = \log 5.25 = 2.39 \text{（比特）}$$

$$I(5) = -\log\frac{5}{21} = \log 4.2 = 2.07 \text{（比特）} \qquad I(6) = -\log\frac{6}{21} = \log 3.5 = 1.81 \text{（比特）}$$

每掷一次骰子，朝上一面提供的平均信息量，即信源 X 的信息熵

$$H(X) = -\sum_{i=1}^{6} P\{X=i\}\log P\{X=i\} = \sum_{i=1}^{6} p(i)I(i)$$

$$= \frac{1}{21}\times 4.39 + \frac{2}{21}\times 3.39 + \frac{3}{21}\times 2.81 + \frac{4}{21}\times 2.39 + \frac{5}{21}\times 2.07 + \frac{6}{21}\times 1.81$$

$$= 2.40 \text{（比特/信源符号）}$$

由此例看出，在信源发符号前，等概信源的平均不确定性大于非等概信源的平均不确定性，等概信源是最令人捉摸不定的。在信源发符号后，等概信源每符号提供的平均信息量，大于非等概信源每符号提供的平均信息量。

【例 1.2】 在一个箱子中，装有 m 个黑球和 $(n-m)$ 个白球。设试验 X 随机地从箱子中取出一个球而不再放回箱子；试验 Y 从箱子中取出第二个球。

（1）计算试验 X 所获取的平均信息量；

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

解 令 W 代表白球；B 代表黑球。

（1）设试验 X 中，取出白球的概率为 $P_X(W)$；取出黑球的概率为 $P_X(B)$，则有

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

信源 X 的信源空间为

$$[X \cdot P]\begin{cases} X: & W & B \\ P(X): & \dfrac{n-m}{n} & \dfrac{m}{n} \end{cases}$$

试验 X 获取的平均信息量，就是信源 X 的信息熵

$$H(X) = -P_X(W)\log P_X(W) - P_X(B)\log P_X(B)$$

$$= -\frac{n-m}{n}\log\frac{n-m}{n} - \frac{m}{n}\log\frac{m}{n}$$

（2）若试验 X 中，取出的第一个球是白（W）球，则试验 Y 取第二个球是白球和黑球的概率分别是

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

若试验 X 中，取出的第一个球是黑（B）球，则试验 Y 取第二个球是白球和黑球的概率分别是

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

这样，试验 Y 中出现白球的概率为

$$P_Y(W) = P_X(W)P_Y(W/W) + P_X(B)P_Y(W/B)$$

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

出现黑球的概率

$$P_Y(B) = P_X(W)P_Y(B/W) + P_X(B)P_Y(B/B)$$

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