

高等院校计算机应用技术系列教材

信息论与编码技术

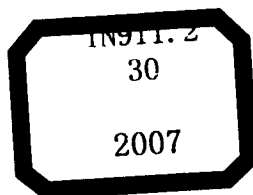
- ◆ 信息论基本概念
- ◆ 信源及其熵
- ◆ 信道及其容量
- ◆ 信息率失真函数
- ◆ 信源编码
- ◆ 信道编码



冯桂 林其伟 陈东华 编著

清华大学出版社

高等院校计算机应用技术系列教材



信息论与编码技术

冯 桂 林其伟 陈东华 编著

清华大学出版社

北 京

内 容 简 介

本书系统介绍了两部分内容：其一，香农信息论的三个基本概念(信源熵、信道容量和信息率失真函数)，以及与这三个概念相对应的三个编码定理；其二，信源编码和信道编码的基本原理与方法。为了便于教学和加深对概念的理解，以及读者自检，每章后面都附有思考题与习题。

本书不追求高深的数学推导，尽量用通俗、生动的语言对概念进行描述，用例题和图表形象地说明基本概念和原理，特别适合于教学和自学。已掌握工科高等数学和工程数学的读者都能读懂本书。

本书的电子课件可通过 <http://www.tupwk.com.cn/download/> 下载。

本书可作为高等院校通信工程、信息工程和电子工程，以及相关专业的本科生、研究生的教材或教学参考书，也可供从事相关专业的科研人员和工程技术人员参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

信息论与编码技术/冯桂,林其伟,陈东华 编著. —北京:清华大学出版社,2007.3

(高等院校计算机应用技术系列教材)

ISBN 978-7-302-14655-1

I. 信… II. ①冯…②林…③陈… III. ①信息论—高等学校—教材②信源编码—编码理论—高等学校—教材③信道编码—编码理论—高等学校—教材 IV.TN911.2

中国版本图书馆 CIP 数据核字(2007)第 018894 号

责任编辑:胡伟卷 刘金喜

装帧设计:康 博

责任校对:胡雁翎

责任印制:何 芊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编:100084

c-service@tup.tsinghua.edu.cn

社 总 机:010-62770175

邮购热线:010-62786544

投稿咨询:010-62772015

客户服务:010-62776969

印 刷 者:北京鑫丰华彩印有限公司

装 订 者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:185×260 印 张:14.25 字 数:329 千字

版 次:2007 年 3 月第 1 版 印 次:2007 年 3 月第 1 次印刷

印 数:1~5000

定 价:25.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103 产品编号:021447-01

前 言

1948年, C.E.Shannon 的开创性文章《通信的数学理论》, 为信息论和编码技术奠定了理论基础。信息论是信息科学中最成熟、最完整、最系统的一部分, 以活跃、新颖的思路和高效解决问题的方法而显示出独特的魅力。在此基础上发展起来的数据通信和计算机技术, 又反过来为信息编码技术的发展和應用创造了有利的环境。随着社会信息化的不断深入, 信息论和编码技术已经渗透到许多应用领域, 展示出勃勃生机和巨大的发展前景。

信息论不仅在方法论的层面上解决通信的有效性和可靠性问题, 而且在认识论的层面上帮助认识事物的本质。学完信息论之后, 再重新审视一下周围的事物, 会有许多新的看法和认识。用信息论可以宏观地认识某些政治问题, 也可以定量地解决某些经济问题, 还可以分析、解释外语学习中存在的问题等。总之, 信息论是高层次信息技术人才必不可少的基础知识。因此, 目前各高等院校的电子信息类专业的本科生、研究生都把信息论和编码技术作为一门重要的专业基础理论课。

由于信息论牵涉到众多学科, 需要广泛的数学基础, 许多读者虽然认识到信息论和编码技术的重要性, 但在繁杂的公式面前只好望而却步。针对这种情况, 作者根据多年的教学经验, 在编写过程中强调基本原理的理解, 选材时充分考虑其实用性, 把信息论涉及的数学知识限制在工科高等数学和工程数学的范畴内, 并尽量以通俗形象的语言描述定义、性质和结论的物理概念, 叙述中重概念描述、少理论推导, 每章还附有相应的思考题与习题。因此, 本书适于作为通信工程及信息类专业的本科生、研究生的教材, 也可作为其他专业学生及有关科技人员的参考书。

本书主要内容包括经典信息论的基本内容和主要结论、信息压缩编码的基本原理、提高通信可靠性的纠错编码理论和方法。全书共分6章, 遵照由浅入深、循序渐进的教学规律, 系统地组织教学内容。第1章绪论, 介绍信息论与编码的基本概念、数字通信系统模型, 以及信息论与编码理论研究的主要内容和意义; 第2章信源及其熵, 介绍信源的数学模型和分类、离散信源的信息熵及其性质、连续信源的信息熵、信源的冗余度和离散无失真信源编码定理; 第3章信道及其容量, 介绍信道的数学模型与分类、信道疑义度与平均互信息、离散信道的信道容量、连续信道的信道容量、信源与信道的匹配, 以及信道编码定理; 第4章信息率失真函数, 介绍失真测度、信息率失真函数及其性质、离散无记忆信源的信息率失真函数、连续无记忆信源的信息率失真函数和保真度准则下的信源编码定理; 第5章信源编码, 介绍编码器和相关概念、变长编码、限失真信源编码和一些实用的信源编码方法; 第6章信道编码, 介绍信道编码的概念、线性分组码、循环码和卷积码。其中第1~3章由冯桂编写, 第4章由陈东华编写, 第5和第6章由林其伟编写, 全书由冯桂统稿。

本书获得了华侨大学教材建设基金的资助，在此表示感谢。

限于作者的水平加上时间比较仓促，书中难免有欠妥和错误之处，殷切希望读者指正，作者不胜感激。本书责编的 E-mail：hnliujinxi@163.com。服务邮箱：wkservice@tsinghua.edu.cn。

编著者

2006年12月

目 录

第 1 章 绪论	1
1.1 基本概念	1
1.1.1 信息的一般概念	1
1.1.2 香农信息定义	4
1.1.3 信息论与编码技术发展简史	7
1.2 数字通信系统模型	8
1.3 信息论与编码理论研究的主要内容和意义	10
1.3.1 信息论研究的主要内容	10
1.3.2 香农信息论对信道编码的指导意义	10
1.3.3 香农信息论对信源编码的指导意义	11
思考题与习题	11
第 2 章 信源及其熵	13
2.1 信源的数学模型和分类	13
2.1.1 信源的数学模型	13
2.1.2 信源的分类	14
2.2 离散信源的信息熵及其性质	18
2.2.1 自信息	19
2.2.2 信息熵	20
2.2.3 熵的基本性质	22
2.3 离散无记忆信源的扩展信源	26
2.4 离散平稳信源	28
2.4.1 平稳信源的概念	28
2.4.2 二维平稳信源	29
2.4.3 一般离散平稳信源	33
2.5 连续信源的信息熵	34
2.5.1 单符号连续信源的熵	35
2.5.2 波形信源的熵	37
2.5.3 最大熵定理	38
2.6 信源的冗余度	39
2.6.1 信源效率	40
2.6.2 信源冗余度	40
2.7 离散无失真信源编码定理	42

2.7.1	信源编码器	42
2.7.2	香农第一定理	45
	思考题与习题	46
第3章	信道及其容量	49
3.1	信道的数学模型与分类	49
3.1.1	信道的分类	50
3.1.2	信道的数学模型	51
3.1.3	单符号离散信道	53
3.2	信道疑义度与平均互信息	55
3.2.1	信道疑义度	55
3.2.2	平均互信息	56
3.2.3	平均互信息的性质	60
3.3	离散无记忆的扩展信道	63
3.4	离散信道的信道容量	65
3.4.1	信道容量的定义	65
3.4.2	简单离散信道的信道容量	66
3.4.3	对称离散信道的信道容量	69
3.4.4	离散无记忆 N 次扩展信道的信道容量	71
3.5	连续信道的信道容量	72
3.5.1	连续单符号加性高斯噪声信道的信道容量	72
3.5.2	多维无记忆加性连续信道的信道容量	74
3.5.3	限频限时限功率的加性高斯白噪声信道的信道容量	78
3.6	信源与信道的匹配	81
3.7	信道编码定理	82
	思考题与习题	83
第4章	信息率失真函数	87
4.1	失真测度	88
4.1.1	系统模型	88
4.1.2	失真度和平均失真度	88
4.2	信息率失真函数及其性质	90
4.2.1	信息率失真函数的定义	90
4.2.2	信息率失真函数的性质	91
4.3	离散无记忆信源的信息率失真函数	94
4.3.1	等概率、对称失真信源的 $R(D)$ 计算	94
4.3.2	离散无记忆信源的信息率失真函数的参量表述	97
4.4	连续无记忆信源的信息率失真函数	102

4.4.1	连续无记忆信源的信息率失真函数的定义	102
4.4.2	高斯信源的信息率失真函数	103
4.4.3	连续无记忆信源的信息率失真函数的参量表述	105
4.4.4	差值失真度量下连续无记忆信源的信息率失真函数	106
4.5	保真度准则下的信源编码定理	110
	思考题与习题	111
第 5 章	信源编码	114
5.1	编码器和相关概念	114
5.1.1	码的分类	115
5.1.2	码树	118
5.1.3	Kraft 不等式	120
5.2	变长编码	121
5.2.1	香农码	123
5.2.2	费诺码	124
5.2.3	霍夫曼码	126
5.3	限失真信源编码	132
5.4	实用信源编码方法	132
5.4.1	游程编码	132
5.4.2	算术编码	136
5.4.3	预测编码	142
5.4.4	变换编码	144
	思考题与习题	152
第 6 章	信道编码	157
6.1	信道编码的概念	157
6.1.1	信道编码的分类	157
6.1.2	与纠错编码有关的基本概念	159
6.1.3	检错与纠错原理	164
6.1.4	检错与纠错方式和能力	166
6.2	线性分组码	168
6.2.1	线性分组码的基本概念	168
6.2.2	生成矩阵和一致校验矩阵	171
6.2.3	线性分组码的译码	177
6.2.4	线性分组码的纠错能力	180
6.2.5	汉明码	183
6.3	循环码	185
6.3.1	循环码的多项式描述	185

6.3.2	循环码的生成矩阵	187
6.3.3	系统循环码	190
6.3.4	多项式运算电路	191
6.3.5	循环码的编码电路	192
6.3.6	循环码的译码电路	195
6.3.7	常用的循环码	198
6.4	卷积码	203
6.4.1	卷积码的编码	203
6.4.2	卷积码的译码	211
	思考题与习题	216
	参考文献	220

第1章 绪 论

现代科学技术的飞速发展使我们对周围世界的认识和理解不断加深，特别是 20 世纪 60 年代以来，计算机技术的迅猛发展、计算机及相关设备的迅速更新换代和个人微型计算机的普及，极大地提高了人们处理信息、存储信息、控制和管理信息的能力，人类社会进入了信息时代。作为现代科学技术基础理论之一的信息论在各个领域的应用和推广使许多经典的概念有了全新的解释，使过去曾经不确切的描述有了精确的定量分析方法。

信息论是人们在长期通信工程的实践中，由通信技术、概率论、随机过程和数理统计等相结合逐步发展起来的一门学科。通常人们公认信息论的奠基人是美国科学家香农 (C.E.Shannon)，他于 1948 年发表的著名论文《通信的数学理论》，为信息论的诞生和发展奠定了理论基础。

信息理论在学术界引起了巨大的反响，在香农信息论的指导下，为提高通信系统信息传输的有效性和可靠性，人们在信源编码和信道编码两个领域进行了卓有成效的研究，取得了丰硕的成果。随着信息理论的迅猛发展和信息概念的不断深化，信息论所涉及的内容早已超越了通信工程的范畴，进入了信息科学这一更广、更新的领域，并渗透到许多学科，得到多个领域的科学工作者的重视。

本章重点内容：

- 信息的定义及其基本特性，信息论与编码的发展历史；
- 数字通信系统模型；
- 信息论与编码技术所研究的主要内容；
- 香农信息论对信源编码和信息编码研究的指导意义。

1.1 基本概念

1.1.1 信息的一般概念

当今社会，人们在各种生产、科学研究和社会活动中，无处不涉及信息的交换和利用。可以说，在我们周围充满了信息，我们正处在“信息社会”中。通过电话、电报、传真和电子邮件，人们可以自由地交流信息；通过报纸、书刊、电子出版物和因特网等媒介，人们可以有选择地获取大量信息；通过电台、电视台等视听媒体，人们可以“身临其境”地感受最新信息。但以上所述还远不能概括信息的全部含义：四季交替透露的是自然界的信

息，而牛顿定律揭示的是物体运动内在规律的信息。信息含义之广几乎可以涵盖整个宇宙，且内容庞杂，层次混叠，不易理清。因此，迅速获取信息，正确处理信息，充分利用信息，既能促进科学技术和国民经济的飞跃发展，又能在各种形式的竞争中占得先机。

如今有关信息的新名词、新术语层出不穷，信息产业在社会经济中所占份额也越来越大，信息基础设施建设速度之快成了当今社会的重要特征之一，物质、能源、信息构成了现代社会生存发展的三大基本支柱。

信息的价值在于它为人们能动地改造外部世界提供了可能，信息所揭示的事物运动规律为人们应用这些规律提供了可能，而信息所描述的事物状态也为人们推动事物向有利的方向发展提供了可能。掌握的资源 and 能量越多，面对同样的信息时人们能以改造世界的可能性也越大。今天我们所掌握的物质力量比过去增大了不知多少倍，因此，信息对于当今社会的发展和人们生活的重要性较之几百年前、几十年前甚至十几年前都有很大的提高。这是信息社会的一个重要特征。

信息的重要性不言而喻，那么，如此神通广大、无处不在而又无所不能的信息究竟是什么呢？

信息是信息论中最基本、最重要的概念，既抽象又复杂。关于信息的科学定义，到目前为止，国内外已提出近百种，它们从不同的侧面和不同的层次来揭示信息的本质。从本质的意义上说，信息是人类社会活动所产生的各种状态和消息的总称，信息是人们对客观事物运动规律及其存在状态的认识。

在信息论和通信理论中经常会遇到信息、消息和信号这三个既有联系又有区别的名词。在学习信息论与编码技术之前，先介绍这几个基本概念。

对信息、消息和信号的定义比较如下：

信息：信息是任何随机事件发生后所包含的内容。人们在对外部世界的观察中获得信息，信息是抽象的意识或知识，它是看不见、摸不着的。而且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。从这一点上我们可以得知，信息量的大小与随机事件发生的概率有直接的关系，概率越小的随机事件一旦发生，它所包含的信息量就越大，而出现概率大的随机事件一旦发生，它所包含的信息量就比较小。

消息：消息是信息的载体。在世界各地的人要想知道其他地方发生的事情的内容，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上的文字与图像等。可见，消息是具体的，它载荷信息，但它不是物理性的。信息只与随机事件的发生有关。每时每刻在世界上的每个地方，都会有各种事件发生，这些事件的发生绝大多数是随机的，即这些随机事件的消息中含有信息；如果事件的发生不是随机而是确定的，那么该消息中就不含信息，该消息的传输也就失去了意义。

信号：信号是消息的物理体现。为了在信道上传输消息，就必须把消息加载(调制)到具有某种物理特征的信号上去。信号是信息的载体，是具有物理性的，如电信号、声信号、光信号等。以人类的语言为例，当人们说话时，发出声信号，这种声信号经过麦克风的转换变成了电信号。这里的声信号和电信号都是我们所指的信号，在本书中涉及的信号主要是指电信号。

按照信息论的观点,信息不等于消息。在日常生活中,人们往往对消息和信息不加区别,认为得到了消息,就是获得了信息。例如,当人们收到一封电报,接到一个电话,收听了广播或看了电视等以后,就认为获得了“信息”。的确,人们从接收到的电报、电话、广播和电视的消息中能获得各种信息,信息与消息有着密切的联系。但是,信息与消息并不等同。人们收到消息后,如果消息告诉了我们很多原来不知道的新内容,我们会感到获得了很多的信息,而如果消息是我们基本已经知道的内容,我们得到的信息就不多。所以信息应该是可以测度的。

我们知道,在电报、电话、广播、电视(也包括雷达、导航、遥测)等通信系统中传输的是各种各样的消息。这些被传送的消息有着各种不同的形式,如文字、数据、语言、图像等。所有这些不同形式的消息都是能被人们的感觉器官所感知的,人们通过通信,接收到消息后,得到的是关于描述某事物状态的具体内容。例如,电视中转播球赛,人们从电视图像中看到了球赛进展情况,而电视的活动图像则是对球赛运动状态的描述。当然,消息也可用来表达人们头脑里的思维活动。例如,朋友给您打电话说:“我想去北京”,您从这条消息得知了您的朋友的想法,该语言消息反映了人的主观世界——大脑物质的思维运动所表现出来的思维状态。

因此,用文字、符号、数据、语言、音符、图像等能够被人们的感觉器官所感知的形式,把客观物质运动和主观思维活动的状态表达出来就成为消息。可见,消息中包含信息,是信息的载体,得到消息,进而获得信息。

同一则信息可用不同的消息形式来载荷,如前所述的球赛进展情况可用电视图像、广播语言、报纸文字等不同消息来表述。而一则消息也可载荷不同的信息,它可能包含非常丰富的信息,也可能只包含很少的信息。因此,信息与消息是既有区别又有联系的。

在各种实际通信系统中,往往为了克服时间或空间的限制而进行通信,必须对消息进行加工处理。把消息转换成适合于信道传输的物理量(如声、光、电等),这种物理量即为信号。信号携带着消息,它是消息的运载工具。如前例中,携带球赛进展情况的电视图像转换成电信号,电信号经过调制变成高频调制电信号,才能在信道中传输;在通信系统的接收端,通过解调还原出原始电信号,在电视屏幕中呈现给观众,从而使观众获得信息。

同样,同一消息可用不同的信号来表示,同一信号也可表示不同的消息。例如,红、绿灯信号:若在十字路口,红、绿灯信号表示能否通行的信息;而在电子仪器面板上,红、绿灯信号却表示仪器是否正常工作或者表示高低电压等信息。所以,信息、消息和信号是既有区别又有联系的三个不同的概念。

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可能是一种“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。

信息的基本概念在于它的不确定性,任何已确定的事物都不含有信息。信息具有以下特征:

- 信息是可以识别的。我们知道信息离不开物理载体，人们可以通过对这些物理载体的识别来获得信息。有些可以用人的感官直接识别信息，例如承载于语言、文字中的信息可以直接用耳、目接收进而识别；而有些则需借助于各种传感器间接识别信息，例如在遥感测量中要利用对电磁波敏感的传感器来间接进行。
- 信息是可以存贮的。信息可以用多种方式存储起来，在需要的时候把存储的信息调取出来。相同的信息可以用文字的形式记录在书刊笔记中，也可以用录音、录像的方式存储在磁性介质中，或者利用计算机存储设备存储起来。
- 信息是可以传递的。信息可以通过多种途径进行传递，人与人之间的信息传递，既可以通过语言、文字，也可以通过体态、动作或表情；社会规模的信息传递，常通过报纸、杂志、电话、广播、电视和网络等。从原则上来说，各种物质的运动形式都可以用于信息的传递。
- 信息是可以量度的。信息量有大小的差别，出现概率越大的随机事件一旦发生，它所包含的信息量就越小；出现概率越小的随机事件一旦发生，它所包含的信息量就越大。
- 信息是可以加工的。人们在收到各种原始信息之后，经过各种方式的加工可以产生新的信息，如研究人员通过收集资料或实验获得的原始信息，经过加工处理可能提出新的见解；计算机对输入的信息通过加工处理，可为人们提供更有意义的结果。
- 信息是可以共享的。信息可以像实物一样作为商品出售，但信息的知识特性使其交易又不同于一般的实物交易，信息交易后，信息出售者与信息购买者共同享有信息。
- 信息的载体是可以转换的。同样内容的信息，可以有不同的形态，可以被包含在不同的物体变化之中，可以从一种形态转换到另一种形态。如我们用感官识别出来的声音、味道、颜色等信息可以转换成语言、文字等形式。在这种转换中，信息的物理载体发生了变化，但信息的内容可以保持完好无损。信息的这个特性，为人们借助于仪器间接地识别信息提供了基础，也为信息的传递、存储和处理带来了方便。

1.1.2 香农信息定义

信息仅仅与随机事件的发生相关，用数学的语言来说，不确定性就是随机性，具有不确定性的事件就是随机事件。因此，可运用研究随机事件的数学工具——概率论和随机过程来测度不确定性的大小。若从直观概念来说，不确定性的大小可以直观地看作事先猜测某随机事件是否发生的难易程度。

某一事物状态的不确定性的大小，与该事物可能出现的不同状态数目和各状态出现的概率大小有关。既然不确定性的大小能够测度，那么信息也是可以测度的。

信息如何测度？当人们收到一封电报，或者听了广播、看了电视，到底能得到多少信息量？由于信息量与不确定性消除的程度有关，我们用消除不确定性的多少，来测度信息量。

例 1.1.1 假设有 A、B 两个布袋，在袋内各装有大小均匀、手感完全一样的 100 个球。A 袋内有红、白球各 50 个，B 袋内有红、白、蓝、黑四种球各 25 个。如果随意从 A 袋或 B 袋中取出一个球，猜测取出的球是什么颜色。该事件当然具有不确定性。按给定条件，显然，“从 A 袋中取出是红球”要比“从 B 袋中取出是红球”要容易得多。因为，在 A 袋中只是在“红”、“白”两种颜色中选择一种，而且“红”与“白”机会均等，即概率为 1/2。而在 B 袋中，红球只占 1/4，取出红球的可能性就小。因此，事件“从 A 袋中取出的是红球”比事件“从 B 袋中取出的是红球”发生的不确定性要小。从这个例子可知，不确定性的与可能发生的不同状态数目和各状态发生的概率有关。

例 1.1.2 足球的魅力在于其比赛结果的不确定性。如果实力接近的两个队进行比赛，在比赛之前，我们很难预测谁能获得胜利，所以这个事件的不确定性很大，当得知比赛结果时，我们会获得较大的信息量。如果实力相差悬殊的两个队进行比赛，一般结果是强队取得胜利，所以当得知比赛结果是强队获胜时，我们并不觉得奇怪，因为结果与我们的猜测是一致的，所以消除的不确定性较小，获得的信息量也较小；当得知比赛结果是弱队取胜时，我们会感到非常惊讶，认为出现了“黑马”，这时将获得很大的信息量。

上述两个例子告诉我们：某一事物状态的不确定性的与该事物可能出现的不同的状态数目和各状态出现的概率的大小有关。某一事物状态出现的概率越小，其不确定性越大，一旦出现，带来的信息量就越大；反之，某一事物状态出现的概率接近于 1，即预料中肯定会发生的事件，那它的不确定性就接近于 0，如果出现，带来的信息量就很小。

香农信息反映的就是事物的不确定性。在其著名的论文《通信的数学理论》中，香农根据概率测度和数理统计学系统地研究了通信中的基本问题，并给出了信息的定量表示，得出了带有普遍意义的重要结论，由此奠定了现代信息论的基础。

我们把某事物各种可能出现的不同状态，即所有可能选择的消息的集合，称为**样本空间**，用 X 表示。在样本空间中，每个可能选择的消息是这个样本空间的一个元素。对于离散消息的集合，概率测度就是对每一个可能选择的消息指定一个概率(非负，且总和为 1)。一个样本空间和它的概率测度一起构成一个**概率空间**。

一般概率空间用 $[X, P]$ 来表示。在离散情况下， X 可写成 $\{a_1, a_2, \dots, a_q\}$ 。在样本空间中选择任一元素 a_i 的概率用 $p_x(a_i)$ 表示，这里下标 X 表示所考虑的概率空间是 X 。如果不会引起混淆，下标可以略去，写成 $p(a_i)$ 。所以在离散情况下，概率空间表示成

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1, & a_2, & \dots, & a_q \\ p(a_1), & p(a_2), & \dots, & p(a_q) \end{bmatrix}$$

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**。在接收端，对是否选择这个消息(或符号) 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 $p(a_i)$ 的某一函数。

在概率空间中， a_i 本身携带的信息量(香农信息)定义为

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

$I(a_i)$ 称为消息(符号) a_i 的自信息。

由上式可知： a_i 出现的先验概率 $p(a_i)$ 越大，则其自信息 $I(a_i)$ 越小；反之， a_i 出现的概率越小，则自信息越大。因此，自信息可描述消息 a_i 出现的先验不确定性，其关系如图1.1.1所示。

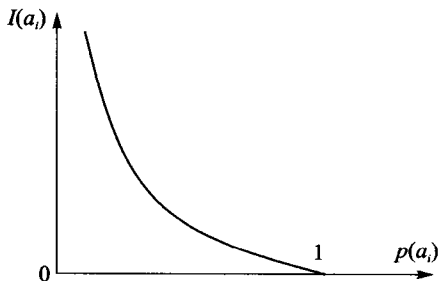


图 1.1.1 自信息与其对应的先验概率的关系

由于在信道中存在干扰，假设接收端收到的消息(符号)为 b_j ，这个 b_j 可能与 a_i 相同，也可能与 a_i 不同，则条件概率 $p(a_i/b_j)$ 反映接收端收到消息 b_j 而发送端发出的是 a_i 的概率，此概率称为后验概率。这样，接收端收到 b_j 后，发送端发送的符号是否为 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 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：

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

定义 $I(a_i; b_j)$ 为发送 a_i 接收 b_j 的互信息。

如果信道没有干扰，信道的统计特性使 a_i 以概率1传送到接收端。这时，收信者接到消息尚存在的不确定性就等于0，即

$$p(a_i/b_j) = 1,$$

$$\log \frac{1}{p(a_i/b_j)} = 0$$

不确定性全部消除。此时互信息 $I(a_i; b_j) = I(a_i)$ 。

以上就是香农关于信息的定义和测度方法，通常也称为概率信息。

香农关于信息的定义是合理的，它以事物的不确定性作为信息定义，非常便于利用数学工具进行定量研究。这也是香农信息论取得成功的重要原因。

1.1.3 信息论与编码技术发展简史

信息论从诞生至今已经经历半个多世纪，目前已成为一门独立的学科。而编码理论与技术研究也有半个世纪的历史，并从刚开始时作为信息论的一个组成部分逐步发展成为比较完善的独立体系。回顾它们的发展历史，我们可以清楚地看到该理论是如何在实践中经过抽象、概括、提高而逐步形成和发展的。

信息理论与编码理论是在长期的通信工程实践和理论研究的基础上发展起来的。一百多年来，物理学中的电磁理论以及后来的电子学理论一旦取得某些突破，很快就会促进电信系统的创造发明或改进。当法拉第于1820—1830年间发现电磁感应规律后不久，莫尔斯就建立起人类第一套电报系统(1832—1835年)，1876年贝尔发明了电话系统，人类由此进入了非常方便的语音通信时代。1864年麦克斯韦预言了电磁波的存在，1888年赫兹用实验证明了这一预言，接着英国的马可尼和俄国的波波夫就发明了无线电通信。1907年福雷斯特发明了能把电信号进行放大的电子管，之后很快就出现了远距离无线电通信系统。20世纪20年代大功率超高频电子管发明以后，人们很快就建立起了电视系统(1925—1927年)。电子在电磁场运动过程中能量相互交换的规律被人们认识后，就出现了微波电子管，接着，在20世纪30年代末和40年代初，微波通信、雷达等系统就迅速发展起来。20世纪60年代发明的激光技术和70年代初光纤传输技术的突破，使人类进入到光纤通信的新时代，由于其带宽、微损、成本低等优点，光纤通信已成为信息高速公路的主干道。

最早对信息进行科学定义的是哈特莱(R.V.L Hartley)，他在1928年发表的《信息传输》一文中，首先提出“信息”这一概念。他认为，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，并主张用所选择的自由度来测度信息。哈特莱的这种理解在一定程度上能够解释通信工程中的一些信息问题，但它存在没有考虑各种可能选择方法的统计特性的局限。正是这些缺陷严重地限制了它的适用范围。

1948年，控制论的创始人之一，美国科学家维纳(N.Wiener)出版了《控制论——动物和机器中通信与控制问题》一书。维纳在该书中将“信息”上升到“最基本概念”的位置。

关于信息的定义，意大利学者朗格(G.Longe)提出用差异量来测度信息，认为“信息就是差异”。在宇宙中到处存在着差异，差异的存在使人们存在着“疑问”和“不确定性”。从这个角度看，差异的确是信息。但是，并不能说没有差异就没有信息。所以，这种关于信息的定义也是不全面、不确切的。

香农在 1948 年发表了著名的论文《通信的数学理论》。他从研究通信系统传输的实质出发,对信息作了科学的定义,并进行了定性和定量的描述。他用概率测度和数理统计的方法系统地讨论了通信的基本问题,得出了无失真信源编码定理和有噪环境下的信道编码定理,由此奠定了现代信息论的基础。1959 年香农又发表了论文“保真度准则下的离散信源编码定理”,以后发展成为信息率失真理论,这一理论是信源编码的核心,至今仍是信息论的研究课题。1961 年香农的论文《双路通信信道》开拓了多用户信息论的研究,而随着卫星通信和网络技术的发展,多用户信息理论成为当前信息论研究的重要课题之一。

在香农编码定理的指导下,信道编码理论和技术逐步发展成熟。20 世纪 50 年代初期,汉明(R.W. Hamming)提出了一种重要的线性分组码——汉明码,此后人们把代数方法引入到纠错码的研究,形成了代数编码理论。1957 年普兰奇(Prange)提出了循环码,在随后的十多年里,纠错码理论的研究主要是围绕着循环码进行的,取得了许多重要成果。1959 年霍昆格姆(Hocquenghrm)、1960 年博斯(Bose)和查德胡里(Chaudhari)各自分别提出了 BCH 码,这是一种可纠正多个随机错误的码,是迄今为止所发现的最好的线性分组码之一。1955 年埃利斯(Elias)提出了不同于分组码的卷积码,接着伍成克拉夫(J.M. Wozencraft)提出了卷积码的序列译码。1967 年维特比(Viterbi)提出了卷积码的最大似然译码法,该译码方法效率高、速度快、译码较简单,得到了极为广泛的应用。1966 年福尼(Forney)提出级联码概念,用两次或更多次编码的方法组合成很长的分组码,以获得性能优良的码,尽可能接近香农极限。随着科学的进步和工程实践的需要,纠错码理论还将进一步发展,它的应用范围也必将进一步扩大。

信源编码的研究由维纳于 1942 年就进行了开创性的工作,以均方量化误差最小为准则,建立最优预测原理,为后来的线性预测压缩编码铺平了道路。1952 年霍夫曼(Huffman)提出了一种重要的无失真信源编码方法——Huffman 码。这是一种非等长码,它可以很好地达到香农无失真信源编码定理所指出的压缩极限,已被证明是平均码长最短的最佳码。为进一步提高有记忆信源的压缩效率,20 世纪 60 年代—70 年代,人们开始将各种正交变换用于信源压缩编码,先后提出 DFT、DCT、WHT、ST、KLT 等多种变换,其中 KLT 为最佳变换,但其实用性不强,综合性能最好的是离散余弦变换(DCT),目前 DCT 已被多种图像压缩国际标准用作主要压缩手段,得到了极为广泛的应用。除了上述几类经典的信源压缩编码方法的研究外,从 20 世纪 90 年代初开始,主要针对图像类信源的特点,人们提出了多种新的压缩原理和方法,包括小波变换编码、分形编码、模型编码、子带编码等。这些方法可有效地消除图像信源的各种冗余,在目前还有很大的发展空间,有关其实际应用问题,还在继续探讨之中。

1.2 数字通信系统模型

从 1.1 节关于信息概念的讨论中,我们已经看到:各种通信系统如电报、电话、电视、