高等学校信息与通信工程“十一五”规划教材

信息论与编码技术基础

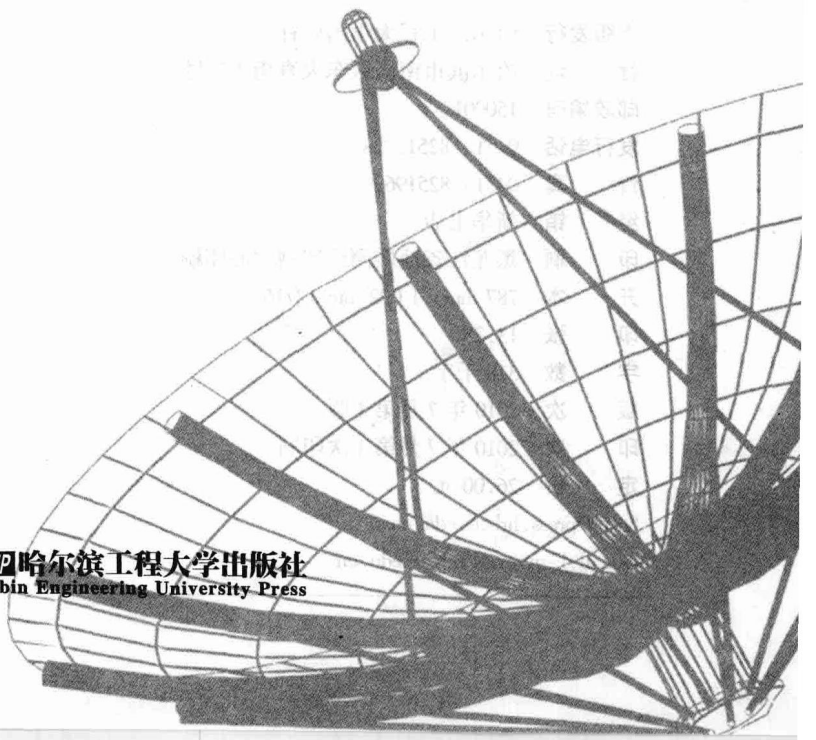
主编/林　云　高洪元　郜丽鹏　蒋伊琳

HEUP 哈尔滨工程大学出版社
Harbin Engineering University Press

高等学校信息与通信工程“十一五”规划教材

信息论与编码技术基础

主编/林　云　高洪元　郜丽鹏　蒋伊琳



HEUP 哈尔滨工程大学出版社
Harbin Engineering University Press

内容简介

本书主要介绍了信息与编码理论的基本原理和技术问题。全书共分为六章,其主要内容包括绪论、信源与信源熵、信道与信道容量、信息率失真函数、信源编码和信道编码等。

本书可以作为高等院校电子工程、通信工程、信息对抗等专业的本科教材和教学参考书,也可以作为研究生入学考试参考用书或者工程技术人员的参考用书。

图书在版编目(CIP)数据

信息论与编码技术基础/林云主编.—哈尔滨:
哈尔滨工程大学出版社,2010.7
ISBN 978-7-81133-838-6

Ⅰ.①信… Ⅱ.①林… Ⅲ.①信息论②信源编码-编码理论③信道编码-编码理论 Ⅳ.①TN911.2

中国版本图书馆 CIP 数据核字(2010)第 133525 号

出版发行	哈尔滨工程大学出版社
社　　址	哈尔滨市南岗区东大直街 124 号
邮政编码	150001
发行电话	0451-82519328
传　　真	0451-82519699
经　　销	新华书店
印　　刷	黑龙江省地质测绘印制中心印刷厂
开　　本	787 mm×1 092 mm　1/16
印　　张	13.25
字　　数	318 千字
版　　次	2010 年 7 月第 1 版
印　　次	2010 年 7 月第 1 次印刷
定　　价	26.00 元

http://press.hrbeu.edu.cn
E-mail:heupress@hrbeu.edu.cn

前言

从1948年C.E.Shannon发表开拓性的论文《通信的数学理论》，为信息论与编码技术奠定了理论基础开始，信息理论就成为了信息科学中最活跃的研究内容，并且以其新颖的思路和高效解决问题的方法而显示出其独特的魅力。在此基础上发展起来的现代通信理论和计算机技术，反过来又为信息论与编码技术的发展和应用创造了有利的环境。随着社会信息化程度的不断深入，信息论和编码技术已经渗透到许多应用领域，展现出勃勃的生机和巨大的发展前景。

目前，信息论和编码技术已经成为信息、通信、电子工程等专业的基础，对理论研究和工程应用均具有重要的指导作用，是高层次信息技术人才必不可少的基础知识。因此，各高等院校均将信息论和编码技术作为信息类专业本科生和研究生的一门重要的专业基础理论课。

由于信息论和编码技术介绍的是信息理论基础和编码理论基础，涉及到众多的学科，内容本身理论性很强，需要深厚的数学基础，许多读者虽然认识到信息论和编码技术的重要性，但是面对繁杂的数学公式只好望而却步，因此迫切需要一本介绍信息理论和编码理论的基本知识，并且与实际应用联系紧密的书籍。

针对这些问题，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和应用，把信息论和编码技术所涉及的数学知识限制在工科高等数学的范畴内，尽量使用通俗形象的语言描述定义、性质和结论的物理概念，减少理论推导，并且在每章后面都附有相应的思考题和习题。因此，本书适合作为通信工程及信息类专业本科生、研究生的教材，也可以作为其他专业学生和相关研究人员的参考书籍。

本书一共分为六章，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章是概论，介绍信息理论与编码理论的基本概念、通信系统的基本模型，以及信息论与编码技术的主要研究内容和应用；第2章介绍信息论的一些基本概念，主要包括信源的分类和数学模型、信源信息量的度量、离散信源的熵和性质、连续信源的熵和性质、信源的冗余度和离散无失真信源编码定理；第3章介绍信道及信通容量，主要包括信道的分类和数学模型、信道容量的基本概念、离散信道容量的计算方法、连续信道容量的计算方法、有噪声信道编码定理、理想通信系统以及网络信息理论；第4章介绍信息率失真函数，主要包括信息率失真的基本概念、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以及限失真信源编码定理；第5章介绍信源编码，主要包括编码器的相关概率、定长和变长无失真信源编码定理，以及几种常用的编码方法，包括香农码、费诺码、

霍夫曼码、游程编码、预测编码和变换编码方法及其性能比较；第6章介绍信道编码，主要包括信道编码的概念、伪噪声编码、检错码和简单纠错码、线性分组码及编译码方法、循环码及编译码方法，以及卷积码的矩阵、多项式和图表。其中第1章、第2章、第3.1节、第3.3节由林云编写；第4章和第5章（除5.4节外）由郜丽鹏编写；第6章、第3.4节、第3.5节和第3.6节由高洪元编写；第3.2节、第3.7节和第5.4节由蒋伊琳编写；全书由林云统稿，高洪元审阅。在本书的编写过程中，得到了刁鸣教授、李一兵教授、孙志国副教授和窦峥副教授的大力帮助，在此表示衷心的感谢。

本书的出版得到了三个基金项目的资助，分别为国家重点基础研究发展计划（“973”计划）基金“61393010101－1”，国家自然科学基金“60802059”，中央高校基本科研业务专项资金“HEUCF100800111”。

由于编者的水平有限，书中难免出现不妥和错误之处，希望读者批评指正。

编　者

2010年3月30日

目录

第1章 概 论

现代科学技术的快速发展使得我们对周围世界的认识和理解不断地加深，特别是20世纪60年代以来，计算机技术的飞速发展、计算机及其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、存储、控制和管理信息的能力，人类社会进入了信息时代。在政治、经济、文化、军事等各个领域，信息的重要性不言而喻，有关信息理论的研究越来越受到重视。

人们在自然和社会活动中，获得信息并对其进行传输、交换、处理、检测、识别、存储、显示等操作，针对这方面内容的研究就是信息科学，信息论是信息科学中的主要基础理论之一。通常人们认为信息论的奠基人是美国科学家香农(C.E.Shannon)，他于1948年发表了著名的论文《通信的数学理论》，为信息论的诞生和发展奠定了理论基础，并且在学术界引起了巨大的反响。在香农信息论的指导下，为了提高通信系统的有效性和可靠性，人们在信源编码和信道编码这两个领域进行了卓有成效的研究，取得了丰硕的成果。随着信息理论的飞速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，进入了信息科学这个更广阔、更新的领域，已渗透到了许多其他学科，并且得到了各个领域科学家和工程师的重视。

本章的重点内容：

- 信息的基本概念；
- 通信系统的基本模型；
- 信息与编码理论的主要研究内容。

1.1 信息的基本概念

信息自古就有，但是古代社会文明程度很低，信息传递手段落后，获取信息困难，人们没有意识到信息的存在。随着人类社会的不断进步，人们才意识到信息的存在，并且对信息的认识随着社会文明程度的提高不断地提高和深入。然而，信息学科毕竟还是一门新兴的学科，人们对信息还没有一个全面的、系统的、准确的、一致的认识，即还未从不同的学科、不同的角度、不同的方面、不同的层次、不同的深度，对信息有不同的认识。

信息的概念十分广泛，不同的定义在百种以上。例如，“信息是事物之间的差异”、“信息是事物联系的普遍形式”、“信息是物质和能量在时间和空间中分布的不均匀性”、“信息是物质的普遍属性”、“信息是收信者事先所不知道的报道”、“信息是用以消除随机不确定性的东西”、“信息是负熵”、“信息是作用于人类感觉器官的东西”、“信息是通信传输的内容”、“信息是加工知识的原材料”、“信息是控制的指令”、“信息就是数据”、“信息就是情报”、“信息就是知识”，等等。

1928年,美国数学家哈特莱(Hartley)在《贝尔系统电话杂志》上发表了一篇题为《信息传输》的论文,把信息理解为选择通信符号的方式,并用选择的自由度来计量这种信息的大小。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。例如,如果从符号表中选择了这样一些符号"I am well",他就发出了"我平安"的信息;如果选择了"I am sick",他就发出了"我病了"的信息。发信者选择的自由度越大,所能发出的信息量也就越大。此外,哈特莱还注意到,选择的具体物理内容是无关紧要的,重要的是选择的方式。也就是说,不管符号代表的意义是什么,只要符号表的符号数目一定,字的长度一定,那么,发信者所能发出的信息的数量就被限定了。所以,他认为"信息是选择的自由度"。

时隔20年,另一位美国数学家香农(C.E.Shannon)在《贝尔系统电话杂志》上发表了题为《通信的数学理论》的长篇论文。这篇论文以概率论为工具,深刻阐述了通信工程的一系列基本理论问题,给出了计算信源信息量和信道容量的方法和一般公式,得到了一组表征信息传递关系的重要编码定理,从而创立了信息论。但是香农并没有给出信息的确切定义,他认为"信息就是一种消息"。

后来,随着认识的进一步深化,人们把信息理解为广义通信的内容。美国数学家、控制论的主要奠基人维纳(Winner)在1950年出版的《控制论与社会》一书中写到:"人通过感觉器官感知周围世界","我们支配环境的命令就是给环境的一种信息",因此,"信息就是我们在适应外部世界,并把这种适应反作用于外部世界的过程中,同外部世界进行交换的内容的名称","接收信息和使用信息的过程,就是我们适应外界环境的偶然性的过程,也是我们在这个环境中有效地生活的过程"。在这里,维纳把人与外部环境交换信息的过程看作是一种广义的通信过程,认为"信息是人与外界相互作用的过程中所交换的内容的名称"。

上面这些关于信息的定义都或多或少地从某种程度上描述了信息的一些特征,但是都不够全面、系统和准确。从本质意义上讲,信息是一个既复杂又抽象的概念,是人类社会活动所产生的各种状态和消息的总称,是人们对客观事物运动规律及其存在状态的认识。信息的基本概念在于它的不确定性,任何已经确定的事物都不含有信息。

在信息理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概念,因此,介绍这三个基本概念是十分必要的。

信息是事物运动状态或存在方式的不确定性的描述。人们从对周围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有关,非随机事件不包含任何信息。从这点上我们可以得知,信息量的大小与随机事件发生的概率有关,概率越小,它所含有的信息量越大,概率越大,它所含有的信息量越小。

消息是信息的载体,我们每天从广播、报纸和电视中获得各种新闻及其他消息。在通信中,消息是指担负着传送信息任务的单个符号或符号序列,这些符号包括字母、文字、数字和语言等。消息是具体的,它包含信息,但它不是物理性的。一个随机事件的消息中含有信息,一个确定事件的消息中不含有信息,此时传输该消息也就失去了意义。

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,是物理性的,如无线电波、光信号等。在通信系统中,实际传输的是信号,但本质的内容是信息。

可见,消息中包含信息,是信息的载体;信号中携带消息,是消息的运载工具。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。

从上面的讨论中我们可以看出，信息、消息、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。作为系统的设计人员，我们接触的只是信号，而这种信号最终要被变成消息才能被大众所接受。

由上面的讨论可知，信息的基本概念在于它的不确定性，已确定的事物都不含信息，信息具有如下的特征。

(1)信息是普遍存在的，信息的本质是事物的运动和变化，只要有事物存在，就会有事物的运动和变化，就会产生信息。绝对静止的事物是没有的。

(2)接收者在收到信息之前，对其内容是未知的，所以信息是新知识、新内容。

(3)信息是能使认识主体对某一时间的未知性或不确定性减少的有用知识。

(4)信息可以产生，也可以消失，同时信息可以被携带、被存储以及被处理。

(5)信息是可以度量的，信息量的多少是有区别的。

(6)信息具有相对性，对于同一事物，不同观察者获得的信息不同。

1.2 通信系统的基本模型

从前面的讨论可知，各种通信系统如电报、电话、电视、广播、遥感、遥测、雷达和导航等，虽然它们的形式和用途各不相同，但本质都是相同的，都是信息的传输系统。为了更好地研究信息传输和处理的共同规律，我们将各种通信系统中具有共同特性的部分提取出来，构成一个统一的通信系统模型，如图 1.1 所示。

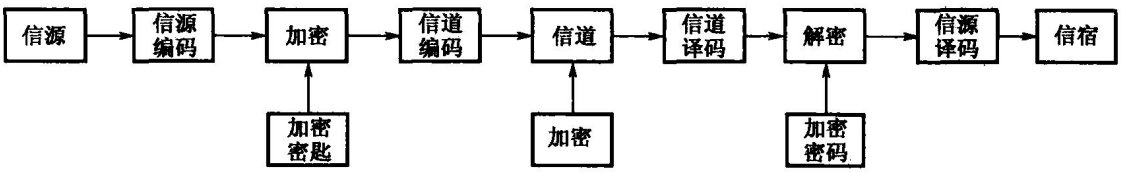


图 1.1 通信系统模型

事实上，这个通信系统模型同样也适用于其他的信息流通系统，它主要由以下六个部分组成。

1.信源

信源是产生消息(或消息序列)的源，消息通常是符号序列或时间函数。例如在电报系统中，消息是由文字、符号、数字组成的报文(符号序列)，称为离散消息；在电话系统中，消息是语声波形(时间函数)，称为连续消息。消息取值服从一定的统计规律，故信源的数学模型是一个在信源符号集合中取值的随机变量、随机序列或随机过程。

2.信宿

信宿是消息传递的对象，即接收消息的人或机器。信宿接收消息的形式与信宿发出的消息可以相同也可以不同，不同时，信宿的消息是信源消息的一个映射。

3.信源编码

信源编码主要有两个作用:①将信源产生的消息变换为适合在信道中传输的一个数字序列(通常为二进制数字序列)或代码组;②压缩信源的冗余度,以提高通信系统传输消息的效率。

4.信道

在实际的通信系统中,信道是指传输信号的媒质或通道,如架空明线、电线、射频波束、人造卫星等。在信息论的模型里,有时为了研究方便,可以将发送端和接收端的一部分如调制器和解调器归入信道,而且将系统中各部分的噪声和干扰也都归入信道中考虑。根据噪声和干扰的统计特性,信道有多种模型。最简单的是离散无记忆(恒参)信道。

5.信道编码

信道编码(或称纠错编码)就是通过编码引进冗余度以提高信息传输的可靠性。

6.加密编码

在密码系统中,信源产生的消息或经信源编码后的数字序列称为明文;加密编码将明文变换为密文(通常是信源符号序列和数字序列间的一一变换)。密码学的研究包括两个方面,即密码的设计和密码的分析与破译,这两方面是紧密联系的。

1.3 信息与编码理论的主要研究内容

信息论是在信息可以度量的基础上,对如何进行有效地、可靠地传递信息进行全面研究的科学,它涉及信息的度量、信息的特性、信息的传输、信道容量、干扰对信息传输的影响等方面内容。通常把上述研究范围内的信息论称为狭义信息论,因为它的创始人是香农,故又被称为香农信息论。一般信息论主要研究信息传输和处理问题。除了香农信息论以外,一般信息论还包括噪声理论、信号滤波和预测、统计检测与估计理论、调制理论、信息处理理论以及保密理论等,这一部分内容以美国科学家维纳为代表。广义信息论不仅包括上述两方面的内容,还包括所有与信息有关的自然和社会科学领域,如模式识别、计算机翻译、心理学、遗传学、神经生理学、语言学、语义学甚至包括社会学中有关信息的问题。

信息在传输、存储和处理过程中不可避免地要受到各种噪声和干扰的影响,信息论为可靠、有效地从数据中提取信息提供必要的根据和方法,因此,必须研究噪声和干扰的性能以及它们与信息本质上的差别。噪声与干扰往往具有某种统计规律的随机特性,信息则具有一定的概率特性,如度量信息量的熵就是概率性质的。因此,集合理论、概率理论、随机过程理论和数理统计理论是信息论应用的基础和工具。

作为一门经典理论,狭义信息论在研究如何度量信息的同时,也研究如何提高图 1.1 所示的通信系统中信息传输的可靠性和有效性,因而狭义信息论也是信源编码和信道编码的理论基础。

本书主要研究与信息和通信科学密切相关的狭义信息论,涉及到信息论的很多基本问题。例如:

(1)什么是信息,如何度量信息?

(2)如何更加有效地传输信息?

(3)对信息的压缩和恢复的极限条件是什么?
(4)从环境中抽取信息极限条件是什么?
(5)如何对信息进行有效性编码?
(6)如何实现信息可靠性编码?
(7)怎样设计高效、可靠的通信系统?

第 2 章　信源与信源熵

在信息论中，信源是发出消息的源，信源输出的消息是以符号形式出现的。如果信源发出的符号是确定的或者是预先知道的，那么该消息对于信宿来说就无信息而言；只有当信源输出的某个(些)符号是随机出现的，无法预先确定时，该符号的出现才会给观察者带来信息。因此，应该用随机变量、随机矢量或随机过程来表示信源，运用集合理论、概率理论、随机过程理论和数据统计理论来研究信息，这就是香农信息论的基本观点。本章首先讨论信源，重点研究信源的统计特性和数学模型，以及不同信源所含信息量的计算方法，从而引入信息论的一些基本概念和重要结论。本章内容是香农信息论的基础。

本章重点内容：

- 信源的数学模型与分类；
- 离散信源的自信息量及其性质；
- 离散信源的熵及其性质；
- 连续信源的熵及其性质；
- 信源的冗余度；
- 离散无失真信源编码定理。

2.1　信源的数学模型与分类

信源是信息的来源，是产生消息或消息序列的源泉。信息是抽象的，而消息是具体的，消息不是信息本身，但它包含和携带信息。所以，要通过信息的表达者——消息来研究信源。本节主要研究信源的数学模型和分类。

2.1.1　信源的数学模型

在通信系统中，信源发出的和信宿收到的消息都是随机的，所以可以用随机变量、随机矢量(随机序列)或随机过程来描述信源输出的消息。或者说，可以用一个样本空间及其概率分布函数(或概率密度函数)——概率空间来描述信源。当信源给定时，它对应的概率空间就已给定；反之，如果概率空间给定，就表示相应的信源已给定。所以，概率空间能表征信源的统计特性，有时也把概率空间称为信源空间，即

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_i & \cdots & a_n \\ p(a_1) & \cdots & p(a_i) & \cdots & p(a_n) \end{bmatrix} \tag{2.1}$$

或

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

式中　X——信源；

$P(X)$——信源符号的概率分布(或概率密度函数)。

2.1.2 信源的分类

实际应用中，分析信源所采用的方法往往依据信源的特性而定。因此，我们有必要先研究一下信源的分类。通常情况下，按照信源发出的消息在时间和幅度上的分布可以将信源分为离散信源和连续信源两大类。

1.离散信源

离散信源是指发出的时间和幅度都是离散分布的离散消息的信源，如文字、数字、数据、计算机代码、电报符号等，这些信源输出的消息都是由单个符号或符号序列组成的，它们符号集合中符号的个数是有限的或是可数的，可以用一维离散型随机变量 X 来描述这些信源的输出。离散信源的数学模型就是离散型的信源空间，即

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_i & \cdots & a_n \\ p(a_1) & \cdots & p(a_i) & \cdots & p(a_n) \end{bmatrix} \tag{2.2}$$

其中，$p(a_i)$代表信源输出符号 a_i 的先验概率，并且应该满足

$$\sum_{i=1}^{n} p(a_i) = 1,\ 0 \leqslant p(a_i) \leqslant 1$$

上式表明信源的信源空间必定是一个完备集，信源输出的消息数是有限的或者是无限可数的，而且每次必定只能从符号集合 $A:\{a_1,\cdots,a_i,\cdots,a_n\}$ 中选取一个消息输出，这就是最基本的离散信源。

另外，如果按照信源发出消息之间的关系还可以将离散信源细分成下列四种类型：

离散信源
- 离散无记忆信源
  - 发出单个符号的无记忆信源
  - 发出符号序列的无记忆信源
- 离散有记忆信源
  - 发出符号序列的有记忆信源
  - 发出符号序列的马尔可夫信源

离散无记忆信源所发出的每个符号之间是相互独立的，发出的符号序列中每个符号之间没有统计关联性，每个符号出现的概率就是它的先验概率；而离散有记忆信源发出的每个符号之间存在统计关联性，每个符号出现的概率不仅和它的先验概率有关，还和整个符号序列或者前面有限个符号有关。

单个符号信源是指信源每次只发出一个符号来代表一条消息；而符号序列信源是指信源每次发出一组包含两个以上符号的符号序列来代表一条消息。

(1)发出单个符号的离散无记忆信源

如果在一个布袋内放有 100 个球，其中 80 个红色球，20 个白色球，随机取出一个球，则取出的不是红色球就是白色球，记下颜色后放回布袋。若将这个实验看成一种信源，则该信源输出的消息是有限的，这种信源就是离散信源。另外，这种信源每次只出现一条消息，出现哪条消息是随机的，因此又被称为发出单个符号的信源。如果经过大量的统计实验，出现红色球的概率是 0.8，出现白色球的概率是 0.2，可以用一个一维离散型随机变量 X 来描述这个信源输出的消息，这个随机变量的样本空间就是符号集合 $A=\{a_1=\text{“红色球”},a_2=\text{“白色球”}\}$。$X$ 的概率分布为 $P(X=a_1)=0.8$，$P(X=a_2)=0.2$，这个概率分布就是各条消息出现的先验概率。它不随试验次数变化，也不和先前的实验结果相关，因而这个信源是无记忆的。上面这种每次只发出一个符号代表一条消息的，并且发出的每个符号之间是相互独立

的信源被称为发出单个符号的离散无记忆信源。

在实际应用中，存在很多这样的信源，例如投硬币、书信文字、计算机的代码、电报符号、阿拉伯数字码等。这些信源输出的消息数都是有限的或可数的，而且每次只能输出其中一条消息，因此可用式(2.2)的信源空间来表示，即

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 0.8 & 0.2 \end{bmatrix}$$

(2)发出符号序列的离散无记忆信源

在多数情况下，信源的输出往往是由一系列符号序列组成的。在上述布袋摸球的实验中，若先取出一个球，记下颜色后放回布袋，再取另一个球，记下颜色。因为第一次取球记下颜色后又放回了布袋，所以第二个球是什么颜色与第一个球是什么颜色无关，是独立的，这个信源是无记忆的。如果按照上述方法取 L 个球，记下相应的颜色，则可以得到一条由 L 个符号序列组成的消息。实际信源输出的消息往往是由这样的符号序列组成，这种用两个以上符号组成的符号序列来代表一个消息，并且符号序列中每个符号都是独立的信源被称为发出符号序列的离散无记忆信源。对于上面这个例子，该信源的信源空间为

$$\begin{bmatrix} \boldsymbol{X} \\ P(\boldsymbol{X}) \end{bmatrix} = \begin{bmatrix} a_1a_1 & a_1a_2 & a_2a_1 & a_2a_2 \\ 0.64 & 0.16 & 0.16 & 0.04 \end{bmatrix}$$

这类信源输出的消息是按一定概率选取的符号序列，所以可以把这类信源输出的消息看作是在时间或空间上离散的一系列随机变量，即随机矢量，也被称为随机序列。这样可以用 L 维随机矢量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_i, \cdots, X_L)$ 来描述这类信源的输出，用联合概率分布来表示信源特性，L 为有限正整数或者可数的无限值。最简单的符号序列信源是 $L=2$ 的情况，此时信源 $\boldsymbol{X}=(X_1, X_2)$，其信源的信源空间为

$$\begin{bmatrix} \boldsymbol{X} \\ P(\boldsymbol{X}) \end{bmatrix} = \begin{bmatrix} (a_1,a_1) & (a_1,a_2) & \cdots & (a_i,a_j) & \cdots & (a_n,a_n) \\ p(a_1,a_1) & p(a_1,a_2) & \cdots & P(a_i,a_j) & \cdots & p(a_n,a_n) \end{bmatrix} \tag{2.3}$$

显然要求 $p(a_ia_j) \geqslant 0, \sum_{i=1}^{n}\sum_{j=1}^{n} p(a_ia_j) = 1$。

如果随机矢量 $\boldsymbol{X}$ 的各维概率分布都与时间起点无关，也就是说，在任意两个不同时刻随机矢量的各维概率分布都相同，则这样的信源被称为离散平稳信源。一般来说，信源输出随机矢量的统计特性比较复杂，分析起来也比较困难。为了便于分析，我们假设信源输出的是平稳随机序列，这样序列的统计特性就与时间的推移无关，很多实际信源都满足这样的假设。

在一些简单的离散平稳信源情况下，如果信源先后发出的一个个符号彼此是无记忆的，即在信源输出的随机矢量 $\boldsymbol{X}$ 中，每个随机变量 X_i 之间都是统计独立的，则 L 维随机矢量的联合概率分布满足

$$P(\boldsymbol{X}) = p(X_1X_2\cdots X_L) = p(X_1)p(X_2)\cdots p(X_L) \tag{2.4}$$

对于离散平稳信源来说，不同时刻的随机变量 X_i 的一维概率分布都相同，即 X_i 的一维概率分布与下标无关，则有

$$p(\boldsymbol{X}) = p(X_1X_2\cdots X_L) = \prod_{i=1}^{L} p(X_i) \tag{2.5}$$

如果不同时刻的随机变量均取值于同一符号集 $A:\{a_1, a_2, \cdots, a_n\}$，则有

$$p(\boldsymbol{X}=\boldsymbol{a}_i)=p(a_{i_1}a_{i_2}\cdots a_{i_k}\cdots a_{i_L})=\prod_{k=1}^{L}p(a_{i_k}) \tag{2.6}$$

其中，$\boldsymbol{a}_i$ 是 L 维随机矢量 $\boldsymbol{X}$ 的一个取值，即 $\boldsymbol{a}_i=(a_{i_1}a_{i_2}\cdots a_{i_k}\cdots a_{i_L})$，$p(a_{i_k})$是符号集合 A 的一维概率分布。

对于上面这样一个离散平稳信源，如果输出随机序列 $\boldsymbol{X}=(X_1,X_2,\cdots,X_L)$中每一个随机变量 X_i 都取值于符号集合 $A:\{a_1,a_2,\cdots,a_n\}$，则把这个信源称为离散无记忆信源 X 的 L 次扩展信源，一般用 $\boldsymbol{X}^L$ 表示。离散无记忆信源 X 的 L 次扩展信源是由离散无记忆信源 X 输出 L 长的随机序列构成的信源。若 $X_i\in A$ 共有 n 种取值可能性，那么随机序列 $\boldsymbol{X}^L$ 有 n^L 种可能性。它的数学模型就是 X 信源空间的 L 重空间，即

$$\begin{bmatrix}\boldsymbol{X}^L\\P(\boldsymbol{X}^L)\end{bmatrix}=\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_{n^L}\\p(\boldsymbol{a}_1) & p(\boldsymbol{a}_2) & \cdots & p(\boldsymbol{a}_{n^L})\end{bmatrix} \tag{2.7}$$

其中，$\boldsymbol{a}_i=(a_{i_1}a_{i_2}\cdots a_{i_k}\cdots a_{i_L})$，$i=1,2,\cdots,n^L$，并满足

$$p(\boldsymbol{a}_i)=p(a_{i_1}a_{i_2}\cdots a_{i_k}\cdots a_{i_L})=p(a_{i_1})p(a_{i_2})\cdots p(a_{i_k})\cdots p(a_{i_L})=\prod_{k=1}^{L}p(a_{i_k})$$

$$\sum_{i=1}^{n^L}p(\boldsymbol{a}_i)=\sum_{i=1}^{n^L}\prod_{k=1}^{L}p(a_{i_k})=1$$

(3)发出符号序列的离散有记忆信源

在一般情况下，信源在不同时刻发出的符号之间是相互依赖的，也就是说，在信源输出的平稳随机序列 $\boldsymbol{X}$ 中各随机变量 X_i 之间是有依赖关系的，这种信源被称为离散有记忆信源。此时，L 维随机矢量的联合概率分布比较复杂，需要引入条件概率分布来说明符号序列内各个符号之间的记忆特征，即

$$p(a_{i_1}a_{i_2}\cdots a_{i_L})=p(a_{i_1})p(a_{i_2}\mid a_{i_1})p(a_{i_3}\mid a_{i_2}a_{i_1})\cdots p(a_{i_L}\mid a_{i_{L-1}}\cdots a_{i_1}) \tag{2.8}$$

在实际中存在很多这样的信源，例如，在汉字组成的中文序列中，只有根据中文语法、习惯用语、修辞制约和表达的实际意义所构成的中文序列才是有意义的中文句子或文章。所以，在汉字序列中前后文字的出现不仅是有依赖的，也是有记忆的。其他如英文、德文、法文等自然语言也都是如此。

(4)发出符号序列的马尔可夫信源

根据前面的例子，表述有记忆信源要比表述无记忆信源困难很多。而在实际中，信源发出的符号往往只对前面若干个符号的依赖性较强，而对更前面的符号依赖性弱。因此，在分析时可以限制随机序列的记忆长度，这样可以获得更加简单的信源表述形式。

如果记忆长度为 $m+1$，则称这种有记忆信源为 m 阶马尔可夫信源，也就是说，信源每次发出的符号只与前面 m 个符号有关系，与更前面的符号无关。此时描述信源符号之间依赖关系的条件概率为

$$p(a_{i_L}\mid a_{i_{L-1}}a_{i_{L-2}}\cdots a_{i_1})=p(a_{i_L}\mid a_{i_{L-1}}a_{i_{L-2}}\cdots a_{i_{L-m}}) \tag{2.9}$$

如果上述条件概率与时间起点 i 也无关，那么信源输出的符号序列可被看作时齐马尔可夫链，此时信源被称为时齐马尔可夫信源。

2.连续信源

连续信源是指输出在时间和幅度上都是连续分布消息(又称连续消息或模拟消息)的信

源。在实际中有些信源输出单个符号的消息 X，但其可能出现的消息数是不可数的无限值，即输出消息 X 的取值是连续区间(a,b)，或取值是实数集 $\mathbf{R}$。例如，语音信号、图像信号、噪声信号在某个时间的取值是连续的；遥测系统中对电压、电流、压力、重力等测得的数据也是连续的。这类输出单个随机符号的连续信源被称为单符号连续信源，可以用一维连续性随机变量 X 来描述，其数学模型是连续型的信源空间，即

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} (a,b) \\ p(x) \end{bmatrix} \quad 或 \quad \begin{bmatrix} \mathbf{R} \\ p(x) \end{bmatrix} \tag{2.10}$$

并且满足

$$\int_a^b p(x)\mathrm{d}x = 1 \quad 或 \quad \int_{\mathbf{R}} p(x)\mathrm{d}x = 1$$

其中，$p(x)$是随机变量 X 的概率密度函数，式(2.10)表明连续型信源空间满足完备条件。上述信源是连续信源中最简单的情况，信源只输出一个符号代表一条消息，所以可以用一维连续型随机变量来描述。

如果信源输出的消息需要用 L 维随机矢量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_i,\cdots,X_L)$来描述，其中每个随机变量 X_i 都是连续型随机变量，并且随机矢量 $\boldsymbol{X}$ 的各维概率密度函数与时间起点无关，这样的信源被称为连续平稳信源。例如，语音信号 $X(t)$、热噪声信号 $n(t)$，它们在时间上取样离散化后的信源为 $\boldsymbol{X}=(\cdots,X_1,X_2,\cdots,X_i,\cdots,X_L,\cdots)$和 $\boldsymbol{n}=(\cdots,n_1,n_2,\cdots,n_i,\cdots,n_L,\cdots)$，它们在时间上是离散的，但每个随机变量 X_i 或 n_i 的取值都是连续的，因此它们是连续平稳信源。

通常情况下，实际信源输出的消息常常在时间和取值上都是连续的。例如，语音信号、热噪声信号、电视图像信号等本身是时间连续函数，而在某一固定时间 t_0，它们可能的取值又是连续的和随机的。对于这种信源输出的消息，可用随机过程来描述，这类信源被称为随机波形信源。

要分析一般的随机波形信源是比较复杂和困难的，而一般情况下常见的随机波形信源输出的消息是时间或频率受限的随机过程。因此，可以根据采样定理的要求对随机过程进行取样，把随机过程用一系列时间或频率上离散的取样点来表示，每个取样都是连续型随机变量，通常每个这样的取值点被称为一个自由度。这样，就可以把随机过程转换成时间或频率上离散的随机序列来处理。如果随机过程是平稳随机过程，则取样后可转换成平稳的随机序列，这样，随机波形信源可以转换成连续平稳信源来处理。若对每个取样值进行量化，就可以将连续的取值转换成有限的或可数的离散取值，也就把连续信源转换成离散信源来处理。

综上所述，对有着不同统计特性的信源可用随机变量、随机矢量(或随机序列)和随机过程来描述其输出的消息，这样的描述方法能够很好地反映信源的随机性质。图 2.1 简要地列出了信源的分类、相应的数学模型和信源之间的转换关系。