普通高等教育“十五”国家级规划教材

高等院校信息与通信工程系列教材

信息论与编码

曹雪虹　张宗橙　编著

清华大学出版社

普通高等教育“十五”国家级规划教材

高等院校信息与通信工程系列教材

信息论与编码

曹雪虹　张宗橙　编著

清华大学出版社
北京

内 容 简 介

本书重点介绍由香农理论发展而来的信息论的基本理论以及编码理论和实现原理。全书共分7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、信道编码和密码学中的理论知识及其实现原理。

全书注重概念，采用通俗的文字，联系目前实际的通信系统，用较多的例题和图示阐述了基本概念、基本理论及实现原理，尽量减少繁杂的公式定理证明。在各章的最后还附有大量习题，便于加深理解。

本书可作为理工科高等院校信息工程、通信工程及相关专业的本科学生教材，亦可作为信息、通信、电子工程等相关专业科技人员的参考书。

图书在版编目(CIP)数据

信息论与编码/曹雪虹，张宗橙编著. —北京：清华大学出版社，2004
(普通高等教育"十五"国家级规划教材　高等院校信息与通信工程系列教材)
ISBN 7-302-08026-7

Ⅰ. 信…　Ⅱ. ①曹…　②张…　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(2004)第 006368 号

出 版 者：清华大学出版社
http://www.tup.com.cn
社 总 机：010-62770175
地　　址：北京清华大学学研大厦
邮　　编：100084
客户服务：010-62776969
组稿编辑：陈国新
文稿编辑：马幸兆
封面设计：傅瑞学
版式设计：刘祎淼
印 装 者：北京鑫海金澳胶印有限公司
发 行 者：新华书店总店北京发行所
开　　本：185×260　印张：15.25　字数：350 千字
版　　次：2004 年 3 月第 1 版　2004 年 3 月第 1 次印刷
书　　号：ISBN 7-302-08026-7/TN·171
印　　数：1～5000
定　　价：23.00 元

前言

信息理论与编码是信息、通信、电子工程类专业的基础，对理论研究和工程应用均有重要的指导作用，广大信息类专业的本科生及科技人员迫切需要掌握信息论与编码的基本知识。

由于信息理论与编码介绍的是信息论基础和编码理论，内容本身理论性很强，现有的一些教材除了介绍理论和公式外，都用了大量篇幅来证明这些理论和公式，这些用作研究生教材是比较适合的。而作为电子、信息、通信工程的本科生及相关专业的工程技术人员，由于其理论基础的不足以及实际应用的需要，不可能花很多精力去研读那些在他们看来是非常难懂而枯燥乏味的证明，而迫切需要一本介绍有关信息理论的基本知识，且与实际应用紧密联系的书籍，本书就是出于这样的目的而编写的。

本书共分 7 章，第 1 章是绪论。第 2 章介绍信息论的一些基本概念，包括自信息量、条件自信息量、互信息量、条件互信息量、平均互信息量、单符号熵、随机序列的熵、熵的性质以及连续信源熵、最大熵定理等，对信源的信息给出定量描述，并解释了冗余度的由来及作用。这一章是后续章节的基础。

第 3 章介绍信道的分类及其表示参数，讨论各种信道能够达到的最大传输速率，即信道的容量及其计算方法。

第 4 章介绍失真函数和信息率失真函数的定义及性质，给出了在一定失真限度内信源必须输出的最小传输速率。

第 5 章介绍信源编码。首先给出无失真信源编码定理和限失真信源编码定理，其中无失真信源编码定理包括定长编码定理和变长编码定理，并详细阐述最佳无失真编码中的香农码、费诺(Fano)码和哈夫曼(Huffman)码的编码方法及其性能比较。最后简单提及常用的几种信源编码方法。

第 6 章介绍信道编码，在阐述信道编码定理、差错控制与信道编译码的基本原理之后，详细介绍最基本，也是最常用的几种信道编码方法，包括线性分组码、卷积码、级联码等。

第 7 章在介绍密码体制的基础知识及其与熵的关系后，简述具有代表性的秘密密钥加密算法 DES，IDEA 和公开密钥加密算法 RSA，MD5 等。还引入信息安全性概念以及数字签名、防火墙等技术。

本书注重基本概念，用较通俗的文字解释其物理意义，辅以一定的例题和图示说明，不再用繁杂的公式来证明这些早已是人们非常成熟的公理。本书联系当前实际通信技术，使读者研读本书后概念清晰，可有目标地将概念应用于实际工作中。

本书由曹雪虹主编。第6章由张宗橙编写，其余各章由曹雪虹编写。在编写过程中，本书得到了徐澄圻教授、胡建彰教授的大力帮助，在此表示衷心的感谢。

限于编者的水平，书中不妥或谬误之处难免，殷切希望读者指正。

编　者

2003年11月

高等院校信息与通信工程系列教材

目 录

CHAPTER 1

第1章 绪论

科学技术的发展使人类跨入了高度发展的信息化时代。在政治、军事、经济等各个领域，信息的重要性不言而喻，有关信息理论的研究正越来越受到重视。

人们在自然和社会活动中，获取信息并对其进行传输、交换、处理、检测、识别、存储、显示等操作，对这方面科学的研究就是信息科学。信息论(information theory)是信息科学的主要理论基础之一。它主要研究可能性和存在性问题，为具体实现提供理论依据。与之对应的是信息技术(information technology)，信息技术主要研究如何实现、怎样实现的问题。

通过本章的学习，可以了解下列问题：信息论的形成和发展；信息论研究的内容及信息的基本概念。本章还结合通信系统模型介绍了模型中各部分的作用、编码的种类和研究内容。

1.1 信息论的形成和发展

信息论理论基础的建立，一般来说开始于香农(Shannon)在研究通信系统时所发表的论文。随着研究的深入与发展，信息论有了更为宽广的内容。

信息在早些时期的定义是由奈奎斯特(Nyquist，H.)和哈特利(Hartley，L. V. R.)在20世纪20年代提出来的。1924年奈奎斯特解释了信号带宽和信息速率之间的关系；1928年哈特利最早研究了通信系统传输信息的能力，给出了信息度量方法；1936年阿姆斯壮(Armstrong)提出增大带宽可以使抗干扰能力加强。这些研究工作都给香农很大的影响，他在1941年至1944年对通信和密码进行深入研究，并用概率论的方法研究通信系统，揭示了通信系统传递的对象就是信息，并对信息给以科学的定量描述，提出了信息熵的概念。还指出通信系统的中心问题是在噪声下如何有效而可靠地传送信息，而实现这一目标的主要方法是编码等。这一成果于1948年以“通信的数学理论”(a mathematical theory of communication)为题公开发表。这是一篇关于现代信息论的开创性的权威论文，为信息论的创立作出了独特

的贡献。香农因此成为信息论的奠基人。

20 世纪 50 年代信息论在学术界引起了巨大的反响。1951 年美国 IRE 成立了信息论组，并于 1955 年正式出版了信息论汇刊。20 世纪 60 年代信道编码技术有了较大进展，成为信息论的又一重要分支。信道编码技术把代数方法引入到纠错码的研究，使分组码技术的发展到了高峰，找到了大量可纠正多个错误的码，而且提出了可实现的译码方法。20 世纪 70 年代卷积码和概率译码有了重大突破，提出了序列译码和 Viterbi 译码方法，并被美国卫星通信系统采用，这使香农理论成为真正具有实用意义的科学理论。1982 年 Ungerboeck G. 提出了将信道编码和调制结合在一起的网格编码调制方法，这种方法无需增大带宽和功率，以增加设备的复杂度换取编码增益，受到了广泛关注，在目前的通信系统中占据统治地位。

信源编码的研究落后于信道编码。香农在 1948 年的论文中提出了无失真信源编码定理，也给出了简单的编码方法——香农码。1952 年费诺(Fano)和哈夫曼(Huffman)分别提出了各自的编码方法，并证明其方法都是最佳编码法。1959 年香农的文章《Coding theorems for a discrete source with a fidelity criterion》系统地提出了信息率失真理论和限失真信源编码定理。这两个理论是数据压缩的数学基础，为各种信源编码的研究奠定了基础。随着传输内容和传输信道的发展，人们针对各种信源的特性，提出了大量实用高效的信源编码方法。

到 20 世纪 70 年代，有关信息论的研究，从点与点间的单用户通信推广发展到多用户系统的研究。1972 年 Cover 发表了有关广播信道的研究，以后陆续进行了有关多接入信道和广播信道模型和信道容量的研究。近 30 多年来，这一领域的研究活跃，大量的论文被发表，使多用户信息论的理论日趋完整。

信息论是在信息可以量度的基础上，对如何有效、可靠地传递信息进行研究的科学，它涉及信息量度、信息特性、信息传输速率、信道容量、干扰对信息传输的影响等方面的知识。通常把上述范围的信息论称为狭义信息论，又因为它的创始人是香农(Shannon, C. E.)，故又称为香农信息论。广义信息论则包含通信的全部统计问题的研究，除了香农信息论之外，还包括信号设计、噪声理论、信号的检测与估值等。当信息在传输、存储和处理的过程中，不可避免地要受到噪声或其他无用信号的干扰时，信息理论会为可靠、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质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，如度量信息量的熵值就是概率性质的。信息论、概率论、随机过程和数理统计学是信息论应用的基础和工具。

本书讲述的信息理论的基本内容是与通信科学密切相关的狭义信息论，涉及到信息理论中的很多基本问题。例如：

① 什么是信息？如何度量信息？

② 在信息传输中，基本的极限条件是什么？

③ 对于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的条件是什么？

⑤ 设计什么样的设备才能达到这些极限？

⑥ 实际上接近极限的设备是否存在?

信息论主要应用于通信领域,在含噪信道中传输信息的最优方法到今天还不十分清楚,特别是在数据的信息量大过信道容量的情况下更是毫无所知,这是经常遇到的情况。因为从信源提取的信息常常是连续的,即信号的信息含量为无限大。在一般信道中传输这样的信号不可能不产生误差。引入信道容量和信息量的概念以后,这类问题便可以得到满意的解释,这样就为设计具最佳效果的通信系统提供了理论依据。

在通信理论中经常会遇到信息、消息和信号这 3 个既有联系又有区别的名词,下面将对它们定义并作一比较。

信息是指各个事物运动的状态及状态变化的方式。人们从对周围世界的观察得到的数据中获得信息。信息是抽象的意识或知识,它是看不见、摸不到的。当由人脑的思维活动产生的一种想法它仍被存储在脑子中时,它就是一种信息。

消息是指包含信息的语言、文字和图像等。例如我们每天从广播节目、报纸和电视节目中获得的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。单个符号消息的情况,例如用 x_1 表示晴天,x_2 表示阴天,x_3 表示雨天;符号序列消息的情况,例如"今天是晴天"这一消息由 5 个汉字构成。可见消息是具体的,它载荷信息,但它不是物理性的。

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。信号是信息的载荷子或载体,是物理性的,如电信号、光信号等。

在通信系统中传送的本质内容是信息,发送端需将信息表示成具体的消息,再将消息载至信号上,才能在实际的通信系统中传输。信号到了接收端(信息论里称为信宿)经过处理变成文字、话音或图像等形式的消息,人们再从中得到有用的信息。在接收端将含有噪声的信号经过各种处理和变换,从而取得有用信息的过程就是信息提取,提取有用信息的过程或方法主要有检测和估计两类。载有信息的可观测、可传输、可存储及可处理的信号,均称为数据。

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。信息的特征如下:

① 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;

② 信息是能使认识主体对某一事物的未知性或不确定性减少的有用知识;

③ 信息可以产生,也可以消失,同时信息可以被携带、被存储及处理;

④ 信息是可以量度的,信息量有多少的差别。

1.2　通信系统的模型

图 1-1 是目前较常用的、也是较完整的通信系统物理模型。下面介绍模型中各个部分的作用及需要研究的核心问题。

(1) 信源

信源是向通信系统提供消息 u 的人和机器。信源本身十分复杂,在信息论中我们仅对信源的输出进行研究。信源输出的是以符号形式出现的具体消息,它载荷信息。信源

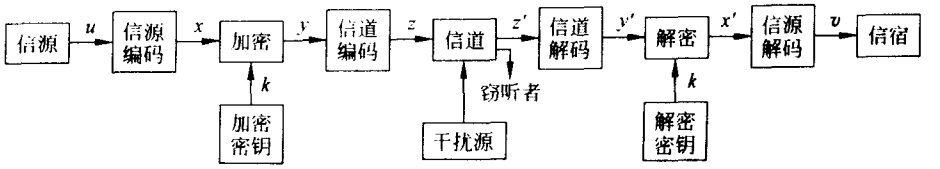

图 1-1　通信系统的物理模型

输出的消息可以有多种形式，但可归纳成两类：离散消息，例如由字母、文字、数字等符号组成的符号序列，或者单个符号；连续消息，例如话音、图像和在时间上连续变化的电参数等。因为通信系统的接收者（信宿）在收到消息之前并不知道信源所发出消息的内容，所以一般地说信源发出的是随机性的消息。但因信源发出的消息都携带着信息，消息的变化具有一定规律性，因此严格地说信源发出的消息并不是完全随机性的。信源的核心问题是它包含的信息到底有多少，怎样将信息定量地表示出来，即如何确定信息量。

（2）信宿

信宿是消息传递的对象，即接收消息的人或机器。根据实际需要，信宿接收的消息 v 的形式可以与信源发出的消息 u 相同，也可以不相同。当两者形式不相同时，v 是 u 的一个映射。信宿需要研究的问题是能收到或提取多少信息。

（3）信道

信道是传递消息的通道，又是传送物理信号的设施。信道可以是一对导线、一条同轴电缆、传输电磁波的空间、一条光导纤维等传输信号的介质。信道的问题主要是它能够传送多少信息，即信道容量的大小。

（4）干扰源

干扰源是整个通信系统中各个干扰的集中反映，用以表示消息在信道中传输时遭受干扰的情况。对于任何通信系统，干扰的性质和大小是影响系统性能的重要因素。

（5）密钥源

密钥源是产生密钥 k 的源。信道编码器输出的信号 x 经过 k 的加密运算后，就把明文 x 变换为密文 y。若窃听者未掌握发送端采用的密钥 k，则很难从窃听到的信号 z' 解出明文 x。而接收端的信宿因知道事先已约定好的密钥 k，因此能从收到的信号 z' 解出明文 x。对于二进制的代码而言，加密相当于 $y=z\oplus p$ 运算（其中序列 p 通常是受密钥控制的伪随机序列），而解密相当于 $x'=y'\oplus p$ 运算。这里 x'，y'，z' 之所以不同于发送端的 x，y，z，是因为考虑到信号 z 在信道中传输时所受到的干扰影响。但在正常通信条件下，总会有 $x'\approx x$，$y'\approx y$ 和 $z'\approx z$ 的结果。

一般地说，通信系统的性能指标主要是有效性、可靠性、安全性和经济性。通信系统优化就是使这些指标达到最佳。除了经济性外，这些指标正是信息论的研究对象，可以通过各种编码处理来使通信系统的性能最优化。根据信息论的各种编码定理和上述通信系统的指标，编码问题可分解为 3 类：信源编码、信道编码和加密编码。

（1）信源编码

信源编码器的作用有两个：一是把信源发出的消息变换成由二进制码元（或多进制

码元)组成的代码组,这种代码组就是基带信号;另一个作用是通过信源编码可以压缩信源的冗余度(即多余度),以提高通信系统传输消息的效率。信源编码可分为无失真信源编码和限失真信源编码。前者适用于离散信源或数字信号;后者主要用于连续信源或模拟信号,如话音、图像等信号的数字处理。从提高通信系统的有效性意义上说,信源编码器的主要指标是它的编码效率,即理论上所需的码率与实际达到的码率之比。一般来说,效率越高,编译码器的代价也将越大。信源译码器的作用是把信道译码器输出的代码组变换成信宿所需要的消息形式,它的作用相当于信源编码器的逆过程。

(2) 信道编码

信道编码器的作用是在信源编码器输出的代码组上有目的地增加一些监督码元,使之具有检错或纠错的能力。信道译码器具有检错或纠错的功能,它能将落在其检错或纠错范围内的错传码元检测出来并加以纠正,以提高传输消息的可靠性。信道编码包括调制解调和纠错检错编译码。信道中的干扰常使通信质量下降,对于模拟信号,表现在收到的信号的信噪比下降;对于数字信号,就是误码率增大。信道编码的主要方法是增大码率或频带,即增大所需的信道容量。这恰与信源编码相反。

(3) 加密编码

加密编码是研究如何隐蔽消息中的信息内容,以便在传输过程中不被窃听,提高通信系统的安全性。将明文变换成密文,通常不需要增大信道容量,例如在二进码信息流上叠加一密钥流。但也有些密码要求占用较大的信道容量。

在实际问题中,上述3类编码应统一考虑,以提高通信系统的性能。这些编码的目标往往是相互矛盾的。提高有效性必须去掉信源符号中的冗余部分,此时信道误码会使接收端不能恢复原来的信息,这就需要相应提高传送的可靠性,不然会使通信质量下降;反之,为了提高可靠性而采用信道编码,往往需增加码值,也就降低了有效性。安全性也有类似情况。编成密码,有时需扩展码位,这样就降低了有效性;有时还会因收、发两端不同步而使授权用户无法获得信息,必须重发而降低有效性,或丢失信息而降低可靠性。从理论上说,若能把3种编码合并成一种编码来编译,即同时考虑有效性、可靠性和安全性,可使编译码器更理想化,在经济上也能更优越。这种三码合一的设想是当前众所关心的课题;但从理论上和技术上的复杂性看,要取得有用的结果,还是相当困难的。值得注意的是,信息论分析的问题是存在性问题,即符合条件的编码是存在的,但并没有给出寻找编码的方法。

本书用了大量篇幅讨论编码问题,着重介绍信源和信道的编码定理。限于课时,主要从概念上解释了这些定理的结论,并没有从严格意义上加以证明。而对于加密编码仅介绍了保密通信中的一些基本知识。这里首先举几个例子来说明编码的应用,例如电报常用的莫尔斯(Morse)码就是按信息论的基本编码原则设计出来的,又如在一些商品上面有一张由粗细条纹组成的标签,从这张标签可以得知该商品的生产厂家、生产日期和价格等信息。这些标签是利用条形码设计出来的,非常方便,非常有用,应用越来越普遍。再如,计算机的运算速度很高,要保证它几乎不出差错,相当于要求它在100年的时间内不得有一秒钟的误差,这就需要利用纠错码来自动、及时地纠正所发生的错误。每出版一本书,都给定一个国际标准书号(ISBN),这大大方便了图书的销售、编目和收藏工作。可以

说，人们在日常生活和生产实践中，正在越来越多地使用编码技术。

顺便指出：不是所有的通信系统都采用如图 1-1 所示的那样全面的技术。例如点对点的有线电话，只要有一对电话机和一条电话线路（铜线）就够了，话音基带信号通过电话机变成相应的电信号（模拟信号），就能在电话线上传送。接收端的电话机再把电信号恢复成人耳能听得清的话音。如果是点对点的无线电话，则在发送端需要一台发信机，把模拟信号调制到射频上，再用大功率发射机经天线发射出去，然后在无线信道中传输。在接收端应使用收信机把收到的调制射频信号解调恢复为发送端的原始话音。若在这样的系统中增加加密和解密装置，就构成无线保密通信系统。在干扰大、信道容量有限的通信系统中，需要采用信源编码和信道编码技术，以提高传输消息的有效性和可靠性。

思 考 题

1-1 信息、消息、信号的定义是什么？三者的关系是什么？

1-2 简述一个通信系统包括的各主要功能模块及其作用。

CHAPTER 2

第 2 章　信源与信息熵

在信息论中，信源是发出消息的源，信源输出以符号形式出现的具体消息。如果符号是确定的且已知，那么该消息就无信息可言。只有当符号的出现是随机的，预先无法确定时，该符号的出现才给观察者提供了信息。而这些符号的出现在统计上具有某些规律性，因此可用随机变量或随机矢量来表示信源，运用概率论和随机过程的理论来研究信息，这是香农信息论的基本点。本章首先介绍各种信源，再研究不同信源所含信息量的计算方法。

2.1　信源的描述与分类

实际应用中分析信源所采用的方法往往要由信源的特性而定。按照信源发出的消息在时间上和幅度上的分布情况可将信源分成离散信源和连续信源两大类。离散信源是指发出时间和幅度上都是离散分布的离散消息的信源，如文字、数字、数据等符号都是离散消息；连续信源是指发出在时间或幅度上是连续分布的连续消息（模拟消息）的信源，如话音、图像、图形等都是连续消息。

另外按照信源发出的符号之间的关系还可细分为下列几种类型：

信源
- 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
- 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

2.1.1　无记忆信源

例如在一个布袋内放 100 个球，其中 80 个球是红色的，20 个球是白色的。若随机摸取一个球，看它的颜色，则摸到的球要么是红色，要么是白色。若将这样的实验看成一种信源，则该信源输出的消息数量是有限的，这种消息数量有限的信源就是离散信源。它每次只出现一种消息，出现哪一种消息是随机的，这样的信源又叫做发出单个符号的信源。若每次看过的球又放回布袋中再做下次实验，那么大量统计证明，出现红色球的概率是 0.8，出现白色球的概率是 0.2。因此可用一个

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 X 的样本空间就是符号集 $A=\{a_1=$"红色"$,a_2=$"白色"$\}$。而 X 的概率分布为 $P(X=a_1)=p(a_1)=0.8$,$P(X=a_2)=p(a_2)=0.2$,这个概率分布就是各消息出现的先验概率。它不随实验次数变化,也不与先前的实验结果相关,因而该信源是无记忆的,可将每次实验结果独立处理。上述这种每次只发出一个符号代表一个消息的信源叫做发出单个符号的无记忆信源。

在实际应用中,存在着很多这样的信源,例如扔骰子、十进制数字码和字母等。这些信源输出都是单个符号的消息,出现的消息数是有限的,且只可能是符号集中的一种,即符合完备性。若各符号出现的概率已知,则该信源就定了;反之,若信源已知,则各符号出现的概率就确定了。所以信源出现的符号及其概率分布就决定了信源,因此可用下列概率空间来表示这种信源:

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ p(a_1) & p(a_2) & \cdots & p(a_n) \end{bmatrix} \tag{2-1-1}$$

其中符号集 $A=\{a_1,a_2,\cdots,a_n\}$,$X\in A$。显然有 $p(a_i)\geqslant 0$,$\sum_{i=1}^{n} p(a_i)=1$。

有的信源输出的消息也是单个符号,但消息的数量是无限的,如符号集 A 的取值是介于 a 和 b 之间的连续值,或者取值为实数集 $\mathbf{R}$ 等。例如在一个袋中有很多干电池,随机摸出一节干电池,用电压表测量其电压值作为输出符号,该信源每次输出一个符号,但符号的取值是在[0,1.5]之间的所有实数,每次测量值是随机的,可用连续型随机变量 X 来描述,这样的信源就是发出单个符号的连续无记忆信源。一般用符号分布的概率密度函数 $p_X(x)$ 来表示,连续信源的概率空间如下:

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} (a,b) \\ p_X(x) \end{bmatrix} \quad 或 \quad \begin{bmatrix} \mathbf{R} \\ p_X(x) \end{bmatrix} \tag{2-1-2}$$

显然应满足 $p_X(x)\geqslant 0$,$\int_a^b p_X(x)\mathrm{d}x=1$ 或 $\int_R p_X(x)\mathrm{d}x=1$。

在有些情况下,也可将符号的连续幅度进行量化,使其取值转换成有限的或可数的离散值,也就是把连续信源转换成离散信源来处理。

然而,很多实际信源输出的消息往往是由一系列符号组成,这种用每次发出 1 组含 2 个以上符号的符号序列来代表一个消息的信源叫做发出符号序列的信源。需要用随机序列(或随机矢量)$\boldsymbol{X}=(X_1,X_2,\cdots,X_l,\cdots,X_L)$ 来描述信源输出的消息,用联合概率分布来表示信源特性。最简单的符号序列信源是 L 为 2 的情况,此时信源 $\boldsymbol{X}=(X_1,X_2)$,其信源的概率空间为

$$\begin{bmatrix} \boldsymbol{X} \\ P \end{bmatrix} = \begin{bmatrix} (a_1,a_1) & (a_1,a_2) & \cdots & (a_n,a_n) \\ p(a_1,a_1) & p(a_1,a_2) & \cdots & p(a_n,a_n) \end{bmatrix} \tag{2-1-3}$$

显然有 $p(a_i,a_j)\geqslant 0$,$\sum_{i,j=1}^{n} p(a_i,a_j)=1$。

上述布袋摸球的实验,若每次取出两个球,由两个球的颜色组成的消息就是符号序列。例如先取出一个球,记下颜色后放回布袋,再取另一个球。由于两次取球时布袋中的红球、白球个数没有变化,第二个球取什么色与第一个球的颜色无关,是独立的,因而该信