

信息与电子学科百本精品教材工程

新编电气与电子信息类本科规划教材

信息论与纠错编码

孙丽华 主编 谢仲华 陈荣伶 副主编

赵长奎 主审

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



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

内 容 简 介

本书重点介绍了信息论与纠错编码的基础内容及应用,共分10章,内容包括:信息及信息的度量,离散信源及信源熵,离散信道及信道容量,信源编码定理和信道编码定理,平均失真测度和信息率失真函数,率失真编码定理,纠错编码代数基础,线性分组码,循环码和卷积码。

本书内容深入浅出,概念清晰,对一些较难理解的概念,配有较多的例题。本书适合作为高等院校电子技术、信息工程、通信工程、雷达、计算机、自动化、仪器仪表等相关专业的本科教材,亦可供从事信息科学、系统工程等的科研人员参考。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

信息论与纠错编码 / 孙丽华主编. —北京: 电子工业出版社, 2005.3

新编电气与电子信息类本科规划教材

ISBN 7-121-00918-8

I. 信… II. 孙… 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 (2005) 第 008281 号

责任编辑: 王 颖

印 刷: 北京牛山世兴印刷厂

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

经 销: 各地新华书店

开 本: 787×1092 1/16 印张: 13.5 字数: 346 千字

印 次: 2005 年 3 月第 1 次印刷

印 数: 5 000 册 定价: 18.00 元

凡购买电子工业出版社的图书,如有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系。
联系电话:(010) 68279077。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。

信息与电子学科百本精品教材工程

《新编电气与电子信息类本科规划教材》

电子信息类专业教材编委会

主任委员：鲍泓（北京联合大学）

副主任委员：徐科军（合肥工业大学）

江国强（桂林电子工业学院）

秦会斌（杭州电子工业学院）

胡先福（电子工业出版社）

委员：崔桂梅 陈新华 陈启祥 段吉海 黄智伟 胡学龙
(按拼音排序)

李霞 李金平 孙丽华 谭博学 王辉 袁家政

姚远程 邹彦 周德新 周宇 张恩平 王颖

编辑出版组

主任：胡先福

成员：王颖 凌毅 韩同平 张孟玮

冉哲 李岩 李维荣 张昱

《新编电气与电子信息类本科规划教材》参编院校

(按拼音排序)

- ▶ 安徽大学
- ▶ 北京联合大学
- ▶ 北华大学
- ▶ 常州工学院
- ▶ 成都理工大学
- ▶ 哈尔滨工程大学
- ▶ 杭州电子科技大学
- ▶ 合肥工业大学
- ▶ 合肥电子工程学院
- ▶ 湖北工业大学
- ▶ 湖南科技大学
- ▶ 河海大学
- ▶ 河北工业大学
- ▶ 扬州大学
- ▶ 华北电力大学
- ▶ 淮海工学院
- ▶ 桂林电子工业学院
- ▶ 桂林工学院
- ▶ 广西工学院
- ▶ 济南大学
- ▶ 南京邮电学院
- ▶ 南京工业大学
- ▶ 南昌大学
- ▶ 南华大学
- ▶ 南通大学
- ▶ 内蒙古科技大学
- ▶ 山东大学
- ▶ 山东理工大学
- ▶ 山东科技大学
- ▶ 青岛大学
- ▶ 上海第二工业学院
- ▶ 上海海运学院
- ▶ 太原理工大学
- ▶ 太原重型机械学院
- ▶ 天津理工大学
- ▶ 厦门大学
- ▶ 西南科技大学
- ▶ 西安建筑科技大学
- ▶ 武汉工业学院
- ▶ 云南大学

前 言

信息论是应用近代概率统计方法研究信息传输、交换、存储和处理的一门学科，也是源于通信实践发展起来的一门新兴应用学科。当前人类已步入信息社会，随着信息概念的不断发展，信息在科学技术上的重要性也早已超越了狭义的通信工程的范畴，受到越来越多的关注。

在高等院校中，信息工程类专业是最为热门的专业之一，信息技术已经改变了很多传统电子类专业的知识结构。在这种形式下，许多高校都在相关专业开设了信息论方面的相关课程，作为本科生、研究生的必修或选修课。一方面这门课是电子类专业的核心课程，很多学校都把它作为必修课或必选课，但另一方面教材可选余地太小，一直有教师难教、学生难学的说法，希望能尽快出版面向本科生的适用教材。本着这种精神，本书力求在内容筛选及编排上以读者最易接受的方式介绍信息理论的知识。

本书包括“信息论与纠错编码”的基本内容及应用。

第1章为“信息论基础”，介绍了信息论的基本概念，以及本书的研究对象，即各种信源和信道。

第2~6章属信息论部分。这部分介绍了信息的度量，内容主要围绕香农三大定理展开，研究在不允许失真情况下信息传输率的极限值，以及给定信源并且允许一定失真条件下信息传输率的极限值，并研究在误码率小于给定值的条件下如何最有效地利用信道的传输能力。

第8~10章属纠错编码部分，第7章讲述了纠错编码所必须的数学知识。

纠错编码是后人沿着香农指明的可行方向为寻求有效而可靠的编译码方法而发展起来的，主要研究在有噪信道条件下各种可行的编码方案及实施技术。

与现有的各种“信息论与编码”教材相比，本教材有如下特色。

1. 本书力图在编排上由浅入深，循序渐进，使读者能较容易地接受信息论与纠错编码方面的基本理论知识；

2. 对于部分具有结论性、指导性的定理，本教材省去了冗繁的定理证明，注重物理概念的阐述及对后人工作的指导意义。

3. 着重介绍了纠错编码部分的内容。教材第8、9和10章分别论述纠错编码中最基本的线性分组码、循环码和卷积码的编译码理论，并举了几种常用的码，如汉明码、BCH码和卷积码，介绍了它们主要的编译码方法。

4. 对一些难以理解的概念，本书配有较多的例题，以帮助学生对抽象定理的理解。各章后面配有一些难易程度不等的思考题和习题，以供选用。

本书全部内容约需60学时，不同专业可根据需要进行选择。

本书第1、7章由陈荣伶编写，第2~6章由孙丽华编写，第8、9、10章由谢仲华编写，由孙丽华负责全书的策划、修改和统编。

本书在编写过程中得到电子工业出版社的大力支持，王颖编辑做了大量的工作，在此一并表示衷心感谢！

由于本书涉及的知识领域广泛，而且变化日新月异，加上时间和水平的限制，难免有差错和不足之处，敬请读者指正！

孙丽华

目 录

第 1 章 信息论基础	(1)
1.1 信息的概念	(2)
1.2 信息传输系统	(3)
1.3 离散信源及其数学模型	(4)
1.3.1 离散无记忆信源	(5)
1.3.2 离散无记忆的扩展信源	(5)
1.3.3 离散平稳有记忆信源	(6)
1.3.4 马尔可夫信源	(7)
1.4 离散信道及其数学模型	(8)
1.4.1 离散无记忆信道	(9)
1.4.2 离散无记忆的扩展信道	(10)
本章小结	(11)
思考题与习题	(12)
第 2 章 信息的度量	(13)
2.1 自信息量和互信息量	(14)
2.1.1 自信息量和条件自信息量	(15)
2.1.2 互信息量和条件互信息量	(18)
2.2 离散集的平均自信息量	(22)
2.2.1 信息熵	(22)
2.2.2 熵函数的性质	(26)
2.3 离散集的平均互信息量	(31)
2.3.1 平均互信息量	(31)
2.3.2 平均互信息量的性质	(33)
2.3.3 有关平均互信息量的两条定理	(36)
2.4 N 维扩展信源的熵和平均互信息量	(39)
2.4.1 N 维扩展信源的熵	(39)
2.4.2 N 维扩展信源的平均互信息量	(41)
2.4.3 有关 N 维平均互信息量的两条定理	(42)
本章小结	(44)
思考题与习题	(44)
第 3 章 离散信源无失真编码	(47)
3.1 概述	(48)
3.1.1 码的分类	(49)
3.1.2 平均码长的计算	(52)
3.1.3 信息传输率	(54)

3.2	等长码及等长编码定理	(55)
3.3	变长码及变长编码定理	(58)
3.3.1	变长码	(58)
3.3.2	克拉夫特不等式	(59)
3.3.3	变长编码定理	(61)
3.4	变长码的编码方法	(66)
3.4.1	香农编码法	(66)
3.4.2	费诺编码法	(68)
3.4.3	霍夫曼编码法	(69)
	本章小结	(73)
	思考题与习题	(74)
第4章	率失真编码	(78)
4.1	失真测度与平均失真	(79)
4.2	信息率失真函数 $R(D)$	(82)
4.2.1	率失真函数的定义	(82)
4.2.2	率失真函数的值域、定义域	(83)
4.2.3	率失真函数的性质	(84)
4.3	率失真函数的计算	(87)
4.3.1	两种特殊情况下的求解	(87)
4.3.2	$R(D)$ 的参数表示法	(91)
4.4	率失真信源编码定理	(95)
	本章小结	(95)
	思考题与习题	(96)
第5章	离散信道的信道容量	(99)
5.1	信道容量的定义	(100)
5.2	离散无记忆信道容量的计算	(100)
5.2.1	达到信道容量的充要条件	(101)
5.2.2	几类特殊信道	(105)
5.3	组合信道的容量	(112)
5.3.1	独立并行信道	(112)
5.3.2	和信道	(113)
5.3.3	串行信道	(114)
	本章小结	(116)
	思考题与习题	(117)
第6章	有噪信道编码	(119)
6.1	信道编码的基本概念	(120)
6.2	译码规则及错误概率	(122)
6.3	信道编码定理	(126)
6.4	费诺不等式及信道编码逆定理	(129)
6.4.1	费诺不等式	(129)

6.4.2	信道编码逆定理	(131)
	本章小结	(132)
	思考题与习题	(133)
第7章	纠错编码代数基础	(135)
7.1	群	(136)
7.1.1	群的定义	(136)
7.1.2	子群	(138)
7.1.3	群的陪集分解	(139)
7.2	环	(140)
7.2.1	环的定义	(140)
7.2.2	整数剩余类环	(142)
7.2.3	多项式剩余类环	(142)
7.3	域	(143)
7.3.1	域的定义	(143)
7.3.2	有限域	(144)
7.3.3	有限域的本原元	(145)
7.3.4	有限域的结构	(146)
7.3.5	有限域的共轭根组	(147)
7.3.6	有限域的综合举例	(148)
	本章小结	(150)
	思考题与习题	(150)
第8章	线性分组码	(152)
8.1	纠错编码的基本概念	(153)
8.1.1	信道纠错编码	(153)
8.1.2	差错控制系统模型及分类	(153)
8.1.3	纠错编码的分类	(154)
8.1.4	差错类型	(155)
8.2	线性分组码的编码	(156)
8.2.1	生成矩阵	(156)
8.2.2	校验矩阵	(158)
8.2.3	编码的实现	(160)
8.3	伴随式与译码	(161)
8.3.1	码的距离和重量	(161)
8.3.2	线性码的纠检错能力	(162)
8.3.3	陪集分解与伴随式	(163)
8.3.4	标准阵列与译码表	(165)
8.4	汉明码	(168)
8.4.1	汉明码的构造	(168)
8.4.2	汉明限与完备码	(169)
	本章小结	(170)

思考题与习题	(170)
第 9 章 循环码	(172)
9.1 循环码的一般概念	(173)
9.1.1 循环码的定义	(173)
9.1.2 循环码的多项式描述	(173)
9.1.3 循环码的生成多项式	(175)
9.2 循环码的编码	(177)
9.2.1 利用生成多项式 $g(x)$ 实现编码	(177)
9.2.2 除法电路	(178)
9.2.3 编码电路	(179)
9.3 循环码的译码	(180)
9.3.1 伴随式计算	(181)
9.3.2 循环码的译码	(182)
9.3.3 Meggit 通用译码器	(184)
9.4 一些重要的循环码	(184)
9.4.1 循环汉明码	(184)
9.4.2 BCH 码	(185)
本章小结	(188)
思考题与习题	(188)
第 10 章 卷积码	(190)
10.1 卷积码的基本概念	(191)
10.1.1 卷积码概述	(191)
10.1.2 卷积码的矩阵描述	(192)
10.2 卷积码的概率译码	(195)
10.2.1 状态图和网格图	(195)
10.2.2 最大似然译码	(196)
10.2.3 维特比译码算法	(198)
本章小结	(201)
思考题与习题	(202)
参考文献	(203)

第 1 章

信息论基础

内容提要

本章首先引出信息的概念, 简述信息传输系统模型的各个组成部分, 进而讨论离散信源和离散信道的数学模型, 最后简单介绍几种常见的离散信源和离散信道。

知识要点

信息、消息的概念, 信息论的研究范畴, 信息传输系统的组成, 离散信源和离散信道的数学模型, 离散无记忆信源和离散无记忆信道的特点。

教学建议

本章主要介绍信息论的基本概念, 后续信息的度量、信源及信道的编码都是围绕本章的概念而展开的, 建议学时数为 3 学时。

1.1 信息的概念

人类利用各种信息使得社会不断发展，而在人类社会的发展过程中信息产生、存储、传输的方式也在不断改进，因此，可以说人类是与信息共同进步的。那么，到底什么是信息呢？

人们认为，物质、能量和信息是构成客观世界的三大要素。信息是物质和能量在空间和时间上分布的不均匀程度，或者说信息是关于事物运动的状态和规律。物质、能量和信息三者相辅相成，缺一不可。没有物质和能量就不存在事物的运动，也就没有运动状态和规律，当然也就没有信息；反过来，事物在运动，这种运动的状态和规律就成为信息。事物运动的状态和规律一旦体现出来，就可以脱离原来的事物而相对独立地载荷于别的事物上，因此信息是可以脱离具体的事物而被摄取、传输、存储和处理的。

我们知道，通信系统中传输的是消息，消息与信息有何区别呢？

信息是一个抽象的概念，而消息是具体的，如一场足球赛事的状况，我们可以分别通过电视、广播和报纸来了解，这其中就涉及到图像、语言、文字这些不同形式的消息。可以说消息是能被人们感觉器官感知的客观物质和主观思维的运动状态或存在状态。在通信之前，收信者无法判断发信者将发送何种状态的消息。通过消息的传递，收信者知道了消息的具体内容，我们说收信者获得了信息。由此看来，通信系统中形式上传输的是消息，实质上传输的是信息，消息中包含信息，消息是信息的载体。

信息论研究的是信息的基本性质及度量方法，研究的是信息的获取、传输、存储和处理的一般规律。从通信的角度讲，信息论是人们在长期的通信工程的实践中，应用近代概率统计方法研究信息传输、交换、存储和处理的一门学科。现代信息论的奠基人——美国科学家香农（Shannon）1948年在《贝尔系统技术》杂志上发表了重要著作《通信的数学理论》。他以概率论为工具，定量地描述了信息的含义，给出了信源产生信息的数学模型；导出了度量信息概率的数学公式，描述了信道传输信息的过程，给出了信道传输能力的容量公式；指出在通信系统中采用适当的编码后能够在随机噪声干扰下有效而可靠地传输信息，并得出了信源编码定理和信道编码定理，从理论上论证了信息传输的一些基本界限。

信息论是在长期的通信工程实践中发展起来的，但是，由于信息问题本身具有极为广泛的意义，信息论很快就渗透到其他领域，如生物学、心理学、语言学、经济学、组织理论等，并在各个领域都取得了新的进展。

对于信息论的研究，一般划分为以下三个不同的范畴。

- 狭义信息论，即通信的数学理论，主要研究狭义信息的度量方法，研究各种信源、信道的描述和信源、信道的编码。
- 实用信息论，研究信息传输和处理问题，也就是狭义信息论方法在调制解调、编码译码及检测等领域的应用。
- 广义信息论，包括信息论在自然和社会中的新的应用，如模式识别、机器翻译、自主学习自组织系统，以及在心理学、生物学、经济学、社会学等一切与信息问题有关的领域中的应用。

在信息时代，人们对于信息的理解远远超出了狭义信息论的讨论范围，要求进一步认识和发展信息概念和信息理论。由于信息科学的很多问题还在探索之中，本书只限于讨论在通信学

科中已建立的完整理论和已取得重大成果的狭义信息论。

1.2 信息传输系统

通信的基本问题是在彼时彼地精确地或近似地再现此时此地发出的消息。各种通信系统，如电报、电视、遥控和雷达系统，虽然它们的形式和用途各不相同，但从信息传输的角度来看，在本质上有许多共同之处。对有收发两端的单向传输系统，一般可概括为图 1-1 所示的系统模型。

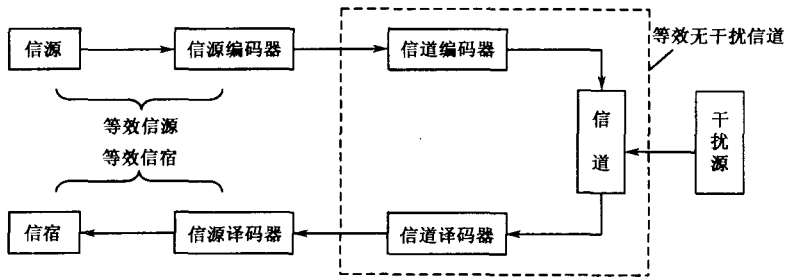


图 1-1 信息传输系统模型

这个模型包括以下五个部分。

1. 信源

信源是产生消息的源。消息可以有多种形式，如语言、文字、图像等。消息可以是离散的，也可以是连续的。消息中包含着信息，消息是信息的载体，通信的结果是为了获得信息。信源输出的消息是随机的，在没有收到这些消息之前，收信者所获得的信息的多少是不确定的。在信息论中用随机变量或随机过程来描述消息。

2. 编码器

编码器是将消息变换成适合于信道传送的信号的设备。编码器分为两种：信源编码器和信道编码器。信源编码器对信源输出的消息进行适当的变换和处理，以达到减少或消除信源的冗余度来提高信息传输率的目的。例如，发中文电报时，需将每个汉字编码为 5 位等重码，为提高传输率，可在不改变语意情况下将词语或句子压缩。比如“奥林匹克运动会”可压缩为“奥运会”，这样，原意不变，而冗余度大大减少。

信道编码是为了抵抗信道的干扰，提高通信的可靠性而对信源编码器的输出进行的变换和处理。信道编码是以增加冗余度为代价来提高可靠性的。如奇偶检验码，通过增加一位奇/偶检验位可检验出奇数位错，该奇/偶检验位与信息的内容无关，是个冗余码元。

3. 信道

信道是信息传输和存储的媒介，如光纤、电缆、无线电波等。信道上不可避免地存在各种干扰源，为了分析方便，我们将系统其他部分产生的各种干扰都等效成信道干扰。信道的输出是信道输入信号和干扰的叠加，由于干扰往往具有随机性，所以信道的特性也用

随机过程来描述。

4. 译码器

译码器是编码器的逆变换，分为信道译码器和信源译码器。信道译码器是从受干扰的信号中尽可能地再现信源的输出。信源译码器是将信道中传输的各种信号还原成收信者能感知的消息。

5. 信宿

信宿是消息的接收者，可以是人，也可以是机器。

通信的最终目的是要有效地、可靠地传递消息。有效性和可靠性往往相互矛盾，要提高有效性，就要减少信源的冗余度，但这样势必会降低抗干扰能力；而要提高可靠性，就要增加纠错检错码元，这样增加了信道的冗余度，从而降低了信息传输率。如上例中，若发电报“奥运会”，当收到电报“×运会”时，无法判断所发电报是“奥运会”、“亚运会”还是“农运会”……，可见所发电文虽然冗余度很小，但容错能力较差；而如果发电报“奥林匹克运动会”，当收到电报“×林匹克运动会”时，我们很容易纠正电文的错误，译为“奥林匹克运动会”，说明信源的冗余度对于抵抗信道的干扰是有益的。香农的信息理论的提出兼顾了这两者的矛盾，在一定的准则下，可以实现有效而可靠的通信。

信息传输系统的模型不是一成不变的，它是根据实际情况而定的。例如，如果研究信息传输的有效性时，可将信道编码器、信道译码器和信道组合起来，等效为一个无干扰信道，这样信源编码器的研究只和信源、信宿有关；如果研究信息传输的可靠性时，可将信源和信源编码器等效为一个对于信道编码器而言的信源，而将信源译码器和信宿等效为信宿，这样信道编码器的研究只和信道有关，与信源、信宿无关。理论表明这种简化方法对大多数理论结果没有太大限制。

1.3 离散信源及其数学模型

信源是产生消息的源，消息是随机的，信息论中用随机变量或随机过程来描述消息，或者说用一个样本空间及其概率测度 $\{X, q(X)\}$ 来描述信源。

根据 X 的不同情况，信源可分为以下类型。

- 离散信源：消息集 X 为离散集合，即时间和空间都离散的信源。
- 连续信源：时间离散而空间连续的信源，如温度、压力等。
- 波形信源：时间连续的信源，如语言、图像等。

连续信源和波形信源输出的消息都可以经过抽样和量化处理成为时间和空间都离散的消息。因此本书中只讨论离散信源的情况。

根据信源的统计特性，离散信源又分为以下两种。

- 无记忆信源： X 的各时刻取值相互独立。
- 有记忆信源： X 的各时刻取值互相有关联，如中文句子中前后文字的出现是有依赖的。有记忆信源的数学模型通常采用联合概率空间来描述。

1.3.1 离散无记忆信源

离散无记忆信源 (DMS, Discrete Memoryless Source) 输出的是单个符号的消息, 不同时刻发出的符号之间彼此统计独立, 而且符号集中的符号数目是有限的或可数的。离散无记忆信源的数学模型为离散型的概率空间, 即

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_i \\ q(x_1) & q(x_2) & \cdots & q(x_i) \end{bmatrix}$$

式中, $q(x_i)$ 是信源输出符号消息 x_i 的先验概率, 满足 $0 \leq q(x_i) \leq 1$, $1 \leq i \leq l$, 且 $\sum_{i=1}^l q(x_i) = 1$ 。信

源每次输出的符号消息 $x_i \in \{a_1, a_2, \dots, a_k\}$, 即 x_i 的取值必定是 k 个符号 $\{a_1, a_2, \dots, a_k\}$ 中的某一个。

【例 1.1】 二进制对称信源只能输出符号 0 或 1, 输出 0 的概率为 p , 输出 1 的概率为 $1-p$, 其概率空间可描述为

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$$

【例 1.2】 随机掷一个无偏骰子, 可能出现的点数与其概率分布为

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$$

1.3.2 离散无记忆的扩展信源

实际情况中, 信源输出的消息往往不是单个符号, 而是由许多不同时刻发出的符号所组成的符号序列。设序列由 N 个符号组成, 若这 N 个符号取自同一符号集 $\{a_1, a_2, \dots, a_k\}$, 并且先、后发出的符号彼此间统计独立, 将这样的信源称做离散无记忆的 N 维扩展信源。其数学模型为 N 维概率空间, 即

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ q(x_1) & q(x_2) & \cdots & q(x_m) \end{bmatrix}$$

式中, \mathbf{x} 为各种长为 N 的符号序列, $\mathbf{x} = x_1 x_2 \cdots x_N$, $x_i \in \{a_1, a_2, \dots, a_k\}$, $1 \leq i \leq N$, 序列集 $\mathbf{X} = \{a_1 a_1 \cdots a_1, a_2 a_2 \cdots a_2, \dots, a_k a_k \cdots a_k\}$, 共有 k^N 种序列, $\mathbf{x} \in \mathbf{X}$ 。序列的概率 $q(\mathbf{x}) = q(x_1 x_2 \cdots x_N) = \prod_{i=1}^N q(x_i)$, 说明符号序列的概率是序列中各个符号概率的乘积, 满足 $\sum_{\mathbf{x}} q(\mathbf{x}) = 1$ 。

【例 1.3】 将二进制对称信源进行三维扩展, 则信源序列共有 2^3 种: 000, 001, 010, 011, 100, 101, 110, 111。由 $q(0)=p$, $q(1)=1-p$ 可得序列的概率依次为

$$\begin{aligned} q(000) &= q(0) \cdot q(0) \cdot q(0) = p^3 & q(001) &= q(0) \cdot q(0) \cdot q(1) = p^2(1-p) \\ q(010) &= q(0) \cdot q(1) \cdot q(0) = p^2(1-p) & q(011) &= q(0) \cdot q(1) \cdot q(1) = p(1-p)^2 \\ q(100) &= q(1) \cdot q(0) \cdot q(0) = p^2(1-p) & q(101) &= q(1) \cdot q(0) \cdot q(1) = p(1-p)^2 \\ q(110) &= q(1) \cdot q(1) \cdot q(0) = p(1-p)^2 & q(111) &= q(1) \cdot q(1) \cdot q(1) = (1-p)^3 \end{aligned}$$

将这 8 种序列看成 8 个符号, 得到一个新的信源, 即

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ p^3 & p^2(1-p) & p^2(1-p) & p(1-p)^2 & p^2(1-p) & p(1-p)^2 & p(1-p)^2 & (1-p)^3 \end{bmatrix}$$

1.3.3 离散平稳有记忆信源

汉字或英文字母组合成中、英文句子时，往往要考虑到语法、习惯用语、修辞等，因此中、英文句子中前后出现的汉字、字母往往是相互关联的。如英文字母 T 后面最常出现 H 和 R，而根本不会出现 Q、F、X。这种关联性称做有记忆。

离散有记忆信源的输出需要用联合概率空间 $\{X, q(X)\}$ 来描述，信源输出的消息可表示为符号序列 $x = x_1 x_2 \cdots x_i \cdots$ ，其中任一变量 $x_i \in \{a_1, a_2, \cdots, a_k\}$ ， $i = 1, 2, \cdots$ ， x_i 表示在 i 时刻信源所发出的符号。很明显信源在 i 时刻发出什么符号与 i 时刻以前信源所发出的符号有关，即由条件概率 $p(x_i | x_{i-1} x_{i-2} \cdots)$ 确定。

如果该条件概率分布与时间起点无关，只与关联长度有关，则该信源为平稳信源。

对于离散平稳有记忆信源，有

$$\begin{aligned} p(x_1 = a_1) &= p(x_2 = a_1) = \cdots \\ p(x_2 = a_2 | x_1 = a_1) &= p(x_3 = a_2 | x_2 = a_1) = \cdots \\ p(x_3 | x_2 x_1) &= p(x_4 | x_3 x_2) = \cdots \\ &\vdots \\ p(x_{i+L} | x_{i+L-1} x_{i+L-2} \cdots x_i) &= p(x_{j+L} | x_{j+L-1} x_{j+L-2} \cdots x_j) = \cdots \\ &\vdots \end{aligned}$$

【例 1.4】 某离散平稳信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ \frac{4}{9} & \frac{3}{9} & \frac{2}{9} \end{bmatrix}$ ，设信源发出的符号只与前一个符号有

关，其关联程度用表 1-1 所示的联合概率 $p(x_i x_j)$ 表示 (x_i 为前一个符号， x_j 为后一个符号)。

表 1-1 $p(x_i x_j)$ 的取值 (1)

$x_i \backslash x_j$	0	1	2
0	1/3	1/9	0
1	1/9	1/18	1/6
2	0	1/6	1/18

解：满足 $\sum_j p(x_i x_j) = q(x_i)$ ，由 $p(x_j | x_i) = \frac{p(x_i x_j)}{q(x_i)}$ 可计算出当已知前一个符号 x_i 时，后一个符号 x_j 为 0、1、2 时的概率，如表 1-2 所示。

表 1-2 $p(x_j | x_i)$ 的取值 (2)

$x_i \backslash x_j$	0	1	2
0	3/4	1/4	0
1	1/3	1/6	1/2
2	0	3/4	1/4

1.3.4 马尔可夫信源

多数有记忆信源的记忆长度是有限的,即某一时刻信源发出的符号只与前面已发出的若干个符号有关。为了描述这种有限的记忆关系,常引入“状态”的概念。这样,信源发出的符号消息与信源所处状态有关。

设信源 r 时刻发出的符号 x_r 与前 m 个符号 $x_{r-1}, x_{r-2}, \dots, x_{r-m}$ 有关(称做 m 阶),这 m 个时间上依次相邻的符号组成一个状态 s ,若 $x_i \in \{a_1, a_2, \dots, a_k\}$,则可能的状态有 k^m 种: s_1, s_2, \dots, s_{k^m} 。用 e_r 表示 r 时刻的状态, $e_r = x_{r-1} x_{r-2} \dots x_{r-m} = s_i$,当符号 x_r 发出后,状态将发生改变,记为 $e_{r+1} = x_r x_{r-1} \dots x_{r-m+1} = s_j$ 。

当状态转移概率和已知状态下发符号的概率与时刻无关,即 $p(e_{r+1} = s_j | e_r = s_i) = p(s_j | s_i)$ 和 $p(x_r = a_l | e_r = s_i) = p(a_l | s_i)$ 时,称为时齐的。

若信源输出的消息序列与信源的状态满足下列条件时,就称做马尔可夫信源。

(1) 某一时刻信源的输出只与当时的信源状态有关,而与以前的状态无关。

有
$$p(x_r = a_l | e_r = s_i, e_{r-1} = s_l, e_{r-2} = s_n, \dots) = p(x_r = a_l | e_r = s_i)$$

且满足
$$\sum_{l=1}^k p(x_r = a_l | e_r = s_i) = 1。$$

(2) 某一时刻信源所处的状态只由当前的输出符号和前一时刻的状态惟一决定。

$$p(e_{r+1} = s_j | x_r = a_l, e_r = s_i) = \begin{cases} 0 \\ 1 \end{cases}$$

当时齐马尔可夫信源达到平稳分布时,满足

$$\begin{cases} p(s_i) = \sum_r p(s_r) p(s_i | s_r) \geq 0 \\ \sum_i p(s_i) = 1 \end{cases}$$

【例 1.5】 某二阶平稳时齐马尔可夫信源,设信源符号集为 $\{a_1, a_2\}$, 状态集为 $\{s_1 = a_1 a_1, s_2 = a_1 a_2 \text{ 或 } a_2 a_2, s_3 = a_2 a_1\}$, 各状态之间的转移情况如图 1-2 所示,求各状态的概率分布。

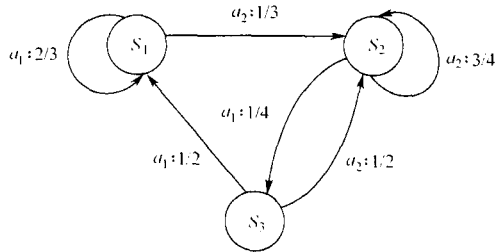


图 1-2 状态转移图

由图 1-2 可知,已知状态下发符号的概率分别为

$$\begin{aligned} p(a_1 | s_1) &= \frac{2}{3} & p(a_2 | s_1) &= \frac{1}{3} \\ p(a_1 | s_2) &= \frac{1}{4} & p(a_2 | s_2) &= \frac{3}{4} \\ p(a_1 | s_3) &= \frac{1}{2} & p(a_2 | s_3) &= \frac{1}{2} \end{aligned}$$